

Chapitre 21

Analyse factorielle multiple (A.F.M)

L'analyse factorielle multiple¹ (A.F.M.) consiste à opérer une analyse factorielle, non plus sur un ensemble de données, mais sur un **ensemble de groupes de données**.

Comment regrouper les variables en fonction d'une typologie ?

- par une analyse procustéenne [Gower, 1975];
- par une méthode STATIS [Lavit, 1988];
- par une analyse factorielle multiple.

Il existe plusieurs A.F.M. :

- l'A.F.M. standard;
- l'A.F.M. avec un modèle INDSCAL;
- l'A.F.M. avec une approche *Partial Least Squares* (P.L.S.);
- l'A.F.M. hiérarchique (A.F.M.H.);
- l'A.F.M. procustéenne (A.F.M.P.).

21.1 L'analyse factorielle multiple

L'A.F.M. fut inventée par B. Escofier-Cordier et J. Pagès [Escofier et Pagès, Brigitte Escofier-Cordier (1941-1994) Jérôme Pagès (né en 1946) 2016]. Elle étudie des **tableaux multiples**. Elle consiste à pondérer les tableaux, à **réaliser une analyse factorielle pondérée**. Pour ce, l'A.F.M. utilise trois espaces vectoriels :

1. \mathbb{R}^K pour les individus;
2. \mathbb{R}^I pour les variables;
3. \mathbb{R}^{I^2} pour les groupes de variables, constituant les tableaux multiples.

Les variables peuvent être quantitatives ou qualitatives.

1. En anglais : *Multiple Factor Analysis* (M.F.A.)

21.1.1 Cas des variables quantitatives

Les variables quantitatives sont toujours centrées réduites.

L'ensemble des individus est décrit par plusieurs groupes de variables. À chaque groupe de variables correspond un tableau. Tous les groupes de variables étant définis sur le même ensemble d'individus, tous les tableaux peuvent être juxtaposés en lignes et former ainsi un seul tableau croisant individus et variables. L'ensemble initial de plusieurs tableaux apparaît comme un unique tableau **structuré en sous-tableaux** (Tab. 21.1). On note :

- X le tableau complet ;
- I l'ensemble des individus ;
- K l'ensemble des variables ;
- J l'ensemble des sous-tableaux ;
- K_i l'ensemble des variables du groupe j ;
- X_j le tableau associé au groupe j .

Groupes		1	...	j					...	J
Variables		1	...	k	...	K_j
Individus	1	X_1	...	X_j					...	X_J
	...									
	i									
	...									
	I									

TABLE 21.1 – Tableau X des données en A.F.M.

Attention ! Les symboles I , J , K ou K_j désignent à la fois l'ensemble et son cardinal.

Une variable du groupe K_j est notée : v_k avec $k \in K_j$.

On suppose que les individus et les variables sont soumis de poids des individus p_i et des variables m_k .

$$\sum_{i=1}^I p_i = 1 \quad (21.1)$$

Les matrices diagonales des poids des individus et des variables sont notées respectivement D , M_j (pour le groupe K_j), et M (pour K). Il faut distinguer deux types de poids :

1. le poids de variables dans les analyses séparées des groupes de variables ;
 - Les variables numériques ont presque toujours le poids 1.

- Il est possible d'affecter des poids différents d'une variable à l'autre.
- 2. le poids des variables dans l'analyse d'ensemble. Les poids initiaux des variables sont modifiés, par exemple, en divisant les colonnes de chaque groupe par sa valeur propre associée (inertie du groupe) ou sa racine carrée. **Cette pondération a pour objectif d'équilibrer le rôle des groupes dans tous les aspects de l'analyse.**

Les objectifs de l'A.F.M.

1. Étudier les ressemblances entre individus du point de vue de l'ensemble des variables **et** les relations entre variables
2. Prendre en compte la structure en groupes
 - a. Étudier globalement les ressemblances et les différences entre groupes
 - b. Étudier les ressemblances et les différences entre groupes du point de vue individuel
 - c. Comparer les typologies issues des analyses séparées
 - d. Équilibrer l'influence de chaque groupe dans l'analyse

L'A.F.M. dans l'espace des individus \mathbb{R}^K

On recherche deux types de représentation.

1. Une représentation du nuage des individus caractérisés par l'ensemble des variables.
2. Une représentation superposée des J nuages d'individus caractérisés chacun par un seul groupe de variables.

Influence de la pondération des groupes sur les J nuages N_I^j À chaque groupe de variable j , correspond un nuage représentant les individus. Ce nuage est noté N_I^j . Il est situé dans un espace de dimension K_j , noté \mathbb{R}^{K_j} .

La pondération vise à équilibrer le rôle des groupes de variables. On choisit la valeur propre $\frac{1}{\lambda_1^j}$. Cela revient à diviser le poids initial de chaque variable du groupe j . Ce coefficient affecte toutes les variables du groupe j . De fait, il ne modifie pas la forme des nuages N_I^j . Toutefois, il normalise les nuages des individus. L'inertie maximale de tout nuage N_I^j dans une direction quelconque vaut 1. Dernier avantage, et pas des moindres, avec cette pondération, deux nuages homothétiques deviennent égaux.

Influence de la pondération des groupes sur le nuage N_I associé à toutes les variables Dans le nuage N_I , le carré de la distance entre deux points i et l , $d^2(i, l)$, est la somme des carré de leur distance dans les N_I^j . On note i^j le point représentant i dans le nuage N_I^j et $v_k(i)$ la valeur de la variable k pour i .

$$d^2(i, l) = \sum_{k \in K} m_k (v_k(i) - v_k(l))^2 \quad (21.2)$$

$$d^2(i, l) = \sum_{j \in J} \sum_{k \in K_j} m_k (v_k(i) - v_k(l))^2 \quad (21.3)$$

$$d^2(i, l) = \sum_{j \in J} d^2(i^j, l^j) \quad (21.4)$$

Dans la distance entre deux éléments du nuage N_I , l'influence des différents groupes n'est équilibrée que si les distances dans les différents nuages N_I^j sont du même ordre de grandeur. De fait, multiplier les poids initiaux des variables du groupe j par un coefficient α_j est un moyen d'équilibrer l'influence des groupes puisque la distance s'écrit alors :

$$d^2(i, l) = \sum_{j \in J} \alpha_j d^2(i^j, l^j) \quad (21.5)$$

On comprend ainsi qu'en choisissant la pondération $\alpha_j = \frac{1}{\lambda_1^j}$, **aucun groupe ne peut être prépondérant dans la première direction d'inertie du nuage moyen.** Par contre, cela a pour conséquence que le nombre de directions de N_I sur lesquelles le groupe j influe, croît avec la dimensionnalité de N_I^j .

Représentation des J nuages N_I^j dans \mathbb{R}^K et nuage moyen Pour représenter **simultanément** les J nuages N_I^j dans l'espace \mathbb{R}^K , il suffit que remarquer que \mathbb{R}^K peut se décomposer en **somme directe** de J sous-espaces orthogonaux, deux à deux isomorphes aux espaces \mathbb{R}^{K_j} .

$$\mathbb{R}^K = \oplus \mathbb{R}^{K_j} \quad (21.6)$$

Sur chacun de ces sous-espaces, la métrique induite par M est la métrique M_j . Il s'agit, de fait, d'un isomorphisme d'espaces euclidiens. Les coordonnées de ces points dans l'espace \mathbb{R}^K sont contenues dans un tableau de dimensions I et K , dans lequel X_j est complété par des zéros. Ce tableau est noté \tilde{X}_j .

$$\tilde{X}_j = \begin{bmatrix} 0 & \dots & K_j & \dots & 0 \end{bmatrix} \quad (21.7)$$

Les nuages N_I^j se situent dans des sous-espaces orthogonaux. Dit autrement, cette représentation simultanée est **artificielle** et **inutilisable directement**. Cela étant, elle sert de base à une véritable représentation simultanée obtenue par projection sur des sous-espaces de \mathbb{R}^K .

On note N_I^* le nuage des centres de gravité, i^* les centres de gravité, des J points i^j représentant le même individu i dans les N_I^j . Ce nuage se déduit de N_I par une **homothétie de rapport** $\frac{1}{J}$. Le nuage N_I^* est un **nuage moyen** pour les N_I^j .

Représentation du nuage moyen On souhaite projeter le nuage N_I , ou son homothétique N_I^* sur un sous-espace de petite dimension de sorte que la projection obtenue ressemble le plus possible à N_I .

Pour y parvenir, on réalise une **A.C.P. du tableau X**, c'est-à-dire une A.C.P. avec des variables pondérées afin d'**équilibrer le rôle des groupes**.

Représentation superposée des J nuages définis par chaque groupe de variables Les nuages N_I^j étant tous situés dans \mathbb{R}^K , il est possible d'en obtenir une représentation simultanée par projection sur un même sous-espace. Pour qu'une telle représentation permette de comparer la position d'un même individu dans les différents nuages, le choix du sous-espace cherche à satisfaire deux conditions.

1. Chaque nuage N_I^j doit être bien représenté. Pour y parvenir, il suffit de choisir une projection orthogonale. En faisant cela, **la qualité d'une représentation peut être mesurée par son inertie**. De fait, on cherche des projections des N_I^j **en maximisant** l'inertie de l'union des N_I^j , c'est-à-dire $N_I^J = \bigcup_j N_I^j$.
2. Les représentations des nuages N_I^j doivent se « ressembler entre elles ». Pour y parvenir, il faut que les points représentant le même individu (points homologues) soient le plus proche possible les uns des autres. Le nuage N_I^j a été partitionné en J nuages contenant chacun I points et notés N_I^j , et représentant chacun l'ensemble des individus vus au travers d'un groupe de variables. Cette partition ne convient pas pour répondre à la condition. On partitionne N_I^J en I nuages contenant chacun J points et notés N_i^J et représentant chacun le même individu i vu au travers de chaque groupe de variables. Le centre de gravité de N_I^J est i^* . Selon le théorème de König-Huygens, l'inertie totale de N_I^J se décompose en inertie intra, c'est-à-dire l'inertie des N_i^J autour des i^* , et en inertie inter, c'est-à-dire l'inertie de N_I^* . Pour que les points associées au même individu i soient proches entre eux, on cherche à **minimiser** l'inertie projetée de chaque N_i^J , donc l'inertie intra de N_I^J .

L'A.F.M. dans l'espace des variables \mathbb{R}^I

Pour satisfaire les deux conditions simultanément, le sous-espace cherché devrait être tel que, en projection, le nuage N_I^J ait une **inertie totale maximale** et une **inertie intra minimale**. Néanmoins, les deux propriétés sont en général **incompatibles**. Il est nécessaire d'établir un **compromis** entre les deux extrêmes en définissant un **critère** donnant *a priori* une **importance équivalente** aux deux propriétés. Pour ce, on utilise le théorème de König-Huygens :

$$\text{inertie inter} = \text{inertie totale} - \text{inertie intra} \quad (21.8)$$

Dit autrement, l'**inertie inter maximale** est un excellent compromis entre une inertie totale maximale et une inertie intra minimale.

Le sous-espace de \mathbb{R}^K sur lequel la projection de N_I^J a une inertie inter maximale est engendré par les premiers axes d'inertie, notés u_s , du nuage N_I^* des centres de gravité. De plus, ce nuage est homothétique au nuage N_I associé à l'ensemble de toutes les variables. Le sous-espace cherché s'obtient par une **A.C.P. du tableau X tout entier**.

Les coordonnées des points de N_I^j sont contenues dans le tableau \tilde{X}_j de dimension (I, K) . En introduisant ces tableaux supplémentaires de l'A.C.P. de X, on obtient la représentation simultanée des N_I^j . Ainsi, cette représentation simultanée suit des règles d'interprétation dérivée directement de l'A.C.P.

Le nuage N_I^j appartenant au sous-espace \mathbb{R}^{K_j} , est projeté sur un vecteur u_s de \mathbb{R}^K n'appartenant pas à \mathbb{R}^{K_j} . La projection de N_I^j sur u_s revient à réaliser **successivement** une projection sur un vecteur u_s^j , la projection de u_s sur \mathbb{R}^{K_j} , puis une projection sur u_s qui contracte le nuage en multipliant les coordonnées par $\cos(\theta_s^j)$ avec θ_s^j l'angle entre u_s et u_s^j .

N.B. Pourquoi ne pas conserver les projections sur les u_s^j pour une représentation simultanée? Dans \mathbb{R}^K , les axes u_s sont orthogonaux, ce qui n'est pas le cas des u_s^j . Dit autrement, on superposerait des nuages dans des espaces munis de métriques différentes.

La **qualité de représentation de chaque nuage** N_I^j se mesure par le rapport entre l'inertie projetée et l'inertie totale du nuage.

$$\text{qualité de représentation de } N_I^j \text{ sur } u_s = \cos^2(\theta_s^j) \times (\text{qualité de représentation sur } u_s^j) \quad (21.9)$$

La **ressemblance entre les représentations des différents nuages** N_I^j cherche à rendre petite l'inertie intra du nuage N_I^j pour que les points i^j représentant le même individu i soient proches entre eux. Néanmoins, cette valeur n'a de signification que comparée à l'inertie totale. Pour chaque axe, on calcule le rapport :

$$\frac{\text{inertie inter}}{\text{inertie totale}} \quad (21.10)$$

Ce rapport n'étant pas la quantité minimisée, ne décroît pas forcément avec l'ordre des axes, mais il constitue un **indicateur sur l'utilité globale** de la représentation superposée des nuages N_I^j . Si ce rapport est proche de 1, tous les nuages N_I^j ont suffisamment de caractères communs pour autoriser une étude fine de leurs différences.

L'A.F.M. dans l'espace des variables \mathbb{R}^I

L'espace \mathbb{R} est l'espace des fonctions numériques définies sur l'ensemble des individus. C'est dans cet espace que sont situées les variables initiales.

Les composantes principales issues des A.C.P. séparées de chacun peuvent être situées dans \mathbb{R}^I .

L'A.F.M. peut être située par rapport à l'**analyse multicanonique** définie par J. D. Carroll [Carroll et Chang, 1970].

J. Douglas Carroll
(1939-2011)

Influence de la pondération des groupes sur les nuages des variables Chaque groupe de variable K_j est représenté par un nuage N_K^j . La pondération des groupes en divisant le poids de chaque variable du groupe j par λ_1^j , rend égale à 1 l'inertie de la première composante principale de chaque nuage N_K^j . En A.F.M., la pondération des variables d'un groupe tient compte à la fois du nombre de variables **et** de leurs liaisons.

N.B. Une pondération qui ne tiendrait pas compte des liaisons entre les variables rendrait relativement faible l'inertie, dans chaque direction, d'un groupe composé de beaucoup de variables indépendantes. À l'opposé, une telle pondération rendrait relativement forte l'inertie dans une direction d'un groupe composé d'une seule variable.

Représentation des variables La représentation des variables est obtenue directement dans l'**A.C.P. du tableau complet X**. Elle est **duale** de l'image de N_I obtenue dans \mathbb{R}^K . Elle est utile pour :

1. aider à l'interprétation de la représentation du nuage des individus ;
2. obtenir une représentation optimale des corrélations entre variables.

Dans le cas d'un tableau multiple, on obtient une image simplifiée des corrélations inter et intra-groupe.

Les composantes principales rendent **maximales** l'inertie des projections de toutes les variables. L'inertie projetée de chaque nuage N_K^j peut être interprétée comme la **contribution d'un groupe**. La pondération des groupes par $\frac{1}{\lambda_1^j}$ équilibre leur influence en ce sens que la contribution d'un groupe à la construction d'un axe est bornée à 1. On retrouve deux idées essentielles.

1. Aucun regroupement ne peut déterminer le premier axe.
2. Un groupe influe sur d'autant plus d'axes qu'il est de grande dimensionnalité.

Représentation des composantes principales de chaque groupe Une étude systématique des corrélations entre les premières composantes de chaque groupe apporte des éléments intéressants pour la comparaison de ces groupes. Elle est réalisée par une A.C.P. des composantes principales de tous les groupes. Les composantes principales du tableau \mathbf{X}_j étant les projections du nuage d'individus sur une base orthonormée, les nuages d'individus définis dans l'A.C.P. de \mathbf{X}_j et dans celle du tableau des composantes de \mathbf{X}_j sont identiques, à condition de conserver les valeurs brutes de ces composantes. Une autre manière de respecter l'inertie λ_s^j de la composante rang s du groupe j consiste à normer cette composante et à lui affecter le poids λ_s^j .

Pour comparer les composantes principales des groupes, il suffit de les introduire en **éléments supplémentaires** dans l'analyse du tableau complet. On peut lui calculer la contribution d'une composante d'un groupe à la construction des axes.

Le groupe K_j est caractérisé par l'**opérateur** $\mathbf{W}_j\mathbf{D}$. La variable canonique associée à une variable z est son image par $\mathbf{W}_j\mathbf{D}$. L'opérateur extrait du groupe j une part d'inertie plus importantes que la projection $P_j(z)$.

$$\mathbf{W}_j\mathbf{D} = \mathbf{X}_j \cdot \mathbf{M} \cdot {}^t\mathbf{X}_j \cdot \mathbf{D} \quad (21.11)$$

et

$$\mathbf{W}_j\mathbf{D} = \mathbf{W}_j \cdot \mathbf{D} \cdot \mathbf{P}_j \quad (21.12)$$

Soit $\{l_r; r = 1, K_j\}$ une base de vecteurs propres de $\mathbf{W}_j\mathbf{D}$ triés par valeurs propres décroissantes : $\lambda_r \geq \lambda_{r+1}$. Si on exprime $P_j(z)$ dans cette base :

$$P_j(z) = \sum_{r=1}^{K_j} x_r l_r \quad (21.13)$$

alors :

$$\mathbf{W}_j\mathbf{D}(z) = \mathbf{W}_j\mathbf{D} \cdot \mathbf{P}_j(z) = \sum_{r=1}^{K_j} \lambda_r x_r l_r = \sum_{k \in K_j} r(z, v_k) \cdot v_k \quad (21.14)$$

L'application $\mathbf{W}_j\mathbf{D}$ renforce d'autant plus les coordonnées, ici de $P_j(z)$, dans la base des l_r qu'elles correspondent à un axe de faible rang. Ainsi, dans sa transformation à l'aide de $\mathbf{W}_j\mathbf{D}$, un vecteur est d'abord projeté sur E_j , puis est rapproché

des premières directions de $\mathbf{W}_j \mathbf{D}$. De fait, $\mathbf{W}_j \mathbf{D}(z_1)$ correspond à une direction de plus grande inertie que $P_j(z_s)$, sauf si $P_j(z_s)$ est colinéaire à un vecteur propre auquel cas $P_j(z)$ et $\mathbf{W}_j \mathbf{D}(z)$ ont des directions identiques.

Les variables générales permettent la représentation d'une structure moyenne des individus. Cette représentation coïncide avec celle du nuage moyen proposé dans \mathbb{R}^K . Les variables canoniques coïncident, à la norme près, avec les projections des N_I^j dans la représentation simultanée. Soit u_s l'axe d'inertie d'ordre s du nuage d'individus N_I associé au tableau \mathbf{X} dans \mathbb{R}^K . Il se déduit de la composante principale \mathbf{F}_s par la relation :

$$u_s = \frac{1}{\lambda_s} {}^t \mathbf{X} \cdot \mathbf{D} \cdot \mathbf{F}_s \quad (21.15)$$

dans laquelle λ_s est la valeur propre $\mathbf{W} \cdot \mathbf{D}$ associé à \mathbf{F}_s ($\mathbf{W} = \sum_{j=1}^J \mathbf{W}_j$)

La projection de N_I^j sur u_s s'écrit :

$$\mathbf{F}_s^j = \tilde{\mathbf{X}}_j \cdot \mathbf{M} \cdot u_s = \frac{1}{\lambda_s} \tilde{\mathbf{X}}_j \cdot \mathbf{M} \cdot {}^t \mathbf{X} \cdot \mathbf{D} \cdot \mathbf{F}_s = \frac{1}{\lambda_s} \mathbf{W}_j \cdot \mathbf{D} \cdot \mathbf{F}_s \quad (21.16)$$

La représentation superposée des N_I^j conduit à calculer un **rapport d'inertie**. L'A.F.M. permet de mettre en évidence les facteurs communs à certains groupes et les facteurs spécifiques d'un groupe. Lors de l'interprétation, on distingue :

1. le coefficient de corrélation entre \mathbf{F}_s et \mathbf{F}_s^j qui indique dans quelle mesure le facteur commun \mathbf{F}_s est effectivement présent dans le groupe K ;
2. la mesure de liaison \mathcal{L} qui indique l'importance relative dans le groupe K_j du facteur commun de rang s . Le g est l'inertie projetée de toutes les variables de K sur un axe factoriel.

L'A.F.M. dans l'espace des groupes de variables \mathbb{R}^{I^2}

Dans l'étude de plusieurs groupes de variables, l'un des objectifs est de comparer **globalement les groupes**. L'espace des groupes de variable \mathbb{R}^{I^2} est la base de cette représentation, qui peut être une aide à l'interprétation de l'A.C.P. du tableau complet \mathbf{X} . Il est à noter qu'elle possède sa propre optimalité.

Le nuage N_J des groupes de variables Pour étudier l'ensemble des groupes, on construit un nuage de points, noté N_J , dans un espace euclidien. L'opérateur $\mathbf{W}_j \cdot \mathbf{D}$ représente le groupe de variable K_j . Il permet, par sa diagonalisation, une restitution parfaite de la structure du nuage N_I^j des individus définis par le groupe K_j . En tant qu'ensemble de I^2 scalaires, la matrice $\mathbf{W}_j \cdot \mathbf{D}$ peut être considéré comme un élément d'un espace vectoriel de dimension I^2 , noté \mathbb{R}^{I^2} . Un groupe

j est représenté dans \mathbb{R}^{I^2} pour la matrice $\mathbf{W}_j \cdot \mathbf{D}$. Cet espace est muni du produit scalaire classique, qui s'écrit pour les éléments $\mathbf{W}_j \cdot \mathbf{D}$ et $\mathbf{W}_l \cdot \mathbf{D}$.

$$\langle \mathbf{W}_j \cdot \mathbf{D}, \mathbf{W}_l \cdot \mathbf{D} \rangle = \sum_{i=1}^I \sum_{i'=1}^I p_i p_{i'} \mathbf{W}_j(i, i') \mathbf{W}_l(i, i') = \text{Tr}(\mathbf{W}_j \cdot \mathbf{D}, \mathbf{W}_l \cdot \mathbf{D}) \quad (21.17)$$

L'influence de la pondération des groupes sur le nuage N_J Dans \mathbb{R}^{I^2} , la norme de $\mathbf{W}_j \cdot \mathbf{D}$ s'écrit :

$$\|\mathbf{W}_j \cdot \mathbf{D}\|^2 = \sum_{s=1}^J (\lambda_s^j)^2 \quad (21.18)$$

La pondération des variables du groupe j par $\frac{1}{\lambda_1^j}$ se traduit dans \mathbb{R}^{I^2} par une homothétie des vecteurs représentant les groupes. Après cette pondération, le norme du vecteur $\mathbf{W}_j \cdot \mathbf{D}$ représentant le groupe j n'est pas égale à 1, mais elle dépend de la structure du groupe. Cette norme est d'autant plus grande que cette structure est multidimensionnelle, c'est-à-dire qu'il existe de nombreux facteurs d'importance comparable à celle du premier d'entre eux. Elle constitue un **indicateur de dimensionalité d'un nuage**.

Stricto sensu, la dimensionalité d'un nuage est égale au nombre de directions orthogonales d'inertie non nulle, soit le nombre de valeurs propres non nulles.

L'indicateur de dimensionalité N_g^2 peut s'écrire en faisant apparaître explicitement la pondération de l'A.F.M.

$$N_g^2(K_j) = \|\mathbf{W}_j \cdot \mathbf{D}\|^2 = \sum_{s=1}^J \left(\frac{\lambda_s^j}{\lambda_1^j} \right)^2 \quad (21.19)$$

L'interprétation du produit scalaire entre deux groupes Le nuage N_J des groupes s'apparente davantage à un nuage de variables qu'à un nuage d'individus, car le produit scalaire entre les vecteurs représentant deux groupes s'interprète comme une mesure de liaison entre ces groupes.

Si les deux groupes comprennent chacun une seule variable, alors la pondération par $\frac{1}{\lambda_1^j}$ donne le poids 1 à une variable centrée réduite qui constitue à elle seule un groupe. À ce groupe d'une seule variable, correspond un élément de \mathbb{R}^{I^2} dit **élément de rang 1**, associé à une matrice symétrique de rang 1. Dans le cas général, l'écriture fait apparaître \mathbf{W}_j en tant que somme d'éléments de rang 1 :

$$\mathbf{W}_j = \sum_{k=1}^K \mathbf{v}_k \cdot \mathbf{m}_k \cdot {}^t \mathbf{v}_k \quad (21.20)$$

avec \mathbf{v}_k une variable, \mathbf{m}_k le poids de la variable, et j le groupe de la variable.

Soit \mathbf{z} et \mathbf{v} deux variables centrées et réduites de poids 1 constituant chacune un groupe, alors les éléments de \mathbb{R}^{I^2} associés à ces groupes ont chacun pour norme 1 et leur produit scalaire est le carré du coefficient de corrélation entre \mathbf{z} et \mathbf{v} muni du poids de l'individu i , p_i . L'inertie de la projection de \mathbf{v} sur \mathbf{z} vaut :

$$\langle \mathbf{z} \cdot \mathbf{z} \cdot \mathbf{D}, \mathbf{v} \cdot \mathbf{v} \cdot \mathbf{D} \rangle = \sum_{i=1}^I \sum_{i'=1}^I p_i p_{i'} z(i) z(i') v(i) v(i') \quad (21.21)$$

$$\langle \mathbf{z} \cdot \mathbf{z} \cdot \mathbf{D}, \mathbf{v} \cdot \mathbf{v} \cdot \mathbf{D} \rangle = \left(\sum_{i=1}^I p_i z(i) v(i) \right)^2 \quad (21.22)$$

$$\langle \mathbf{z} \cdot \mathbf{z} \cdot \mathbf{D}, \mathbf{v} \cdot \mathbf{v} \cdot \mathbf{D} \rangle = [r(\mathbf{v}, \mathbf{z})]^2 \quad (21.23)$$

Si on a à faire à un groupe d'une variable et à un groupe multidimensionnel, on note z la variable centrée réduite et de poids 1 du groupe K_1 réduit à un seul élément et \mathbf{v}_k les variables réduites et de poids \mathbf{m}_k du groupe K_2 , alors :

$$\langle \mathbf{W}_1 \cdot \mathbf{D}, \mathbf{W}_2 \cdot \mathbf{D} \rangle = \left\langle \mathbf{z} \cdot \mathbf{z} \cdot \mathbf{D}, \sum_{k=1}^K \mathbf{v}_k \cdot \mathbf{m}_k \cdot \mathbf{v}_k \cdot \mathbf{D} \right\rangle \quad (21.24)$$

$$\langle \mathbf{W}_1 \cdot \mathbf{D}, \mathbf{W}_2 \cdot \mathbf{D} \rangle = \sum_{k=1}^K \langle \mathbf{z} \cdot \mathbf{z} \cdot \mathbf{D}, \mathbf{v}_k \cdot \mathbf{v}_k \cdot \mathbf{D} \rangle \quad (21.25)$$

$$\langle \mathbf{W}_1 \cdot \mathbf{D}, \mathbf{W}_2 \cdot \mathbf{D} \rangle = \mathcal{L}\{\mathbf{z}, K_2\} \quad (21.26)$$

c'est-à-dire la mesure de liaison $\mathcal{L}\}$ entre une variable et un groupe de variables.

Si les deux groupes sont multidimensionnels, on note \mathbf{z}_l les variables réduites et de poids \mathbf{m}_k dun groupe K , alors :

$$\langle \mathbf{W}_1 \cdot \mathbf{D}, \mathbf{W}_l \cdot \mathbf{D} \rangle = \sum_{l=1}^K \mathbf{m}_l \sum_{k=1}^K \mathbf{m}_k \langle \mathbf{z}_l \cdot \mathbf{z}_l \cdot \mathbf{D}, \mathbf{v}_k \cdot \mathbf{v}_k \cdot \mathbf{D} \rangle \geq \sum_{k=1}^K \mathbf{m}_k \mathcal{L}\{\mathbf{v}_k, K_1\} \quad (21.27)$$

Cette quantité est nulle lorsque toutes les variables d'un groupe sont orthogonales à toutes les variables de l'autre groupe. De fait, toutes les variables du groupe j , sont non-corrélées à \mathbf{v}_k . Cette quantité vaut 1. \mathbf{v}_k est confondue avec la première composante principale de K_j . Cette quantité est d'autant plus grande que chacune des variables d'un groupe est davantage liée à l'ensemble des variables de l'autre groupe. Elle constitue un **indice de liaison général entre groupes de variables**.

On généralise la première définition de $\mathcal{L}\}$ à deux groupes quelconque de variables, ce qui fait apparaître explicitement la pondération de l'A.F.M.

$$\mathcal{L}\{K_1, K_2\} = \left\langle \frac{\mathbf{W}_1 \cdot \mathbf{D}}{\lambda_1^1}, \frac{\mathbf{W}_2 \cdot \mathbf{D}}{\lambda_1^2} \right\rangle \quad (21.28)$$

Le **coefficient** RV est l'indicateur classique de liaison entre deux groupes :

$$RV(K_1, K_2) = \left\langle \frac{W_1 \cdot D}{\|W_1 \cdot D\|}, \frac{W_2 \cdot D}{\|W_2 \cdot D\|} \right\rangle \quad (21.29)$$

Le coefficient RV s'interprète dans \mathbb{R}^{I^2} comme un cosinus. Sa valeur est toujours positive, et $RV \in [0, 1]$

L'étude du nuage des groupes de variables On peut représenter les groupes en tant qu'aide à l'interprétation de l'A.C.P. Dans l'espace \mathbb{R}^{I^2} , soit le groupe formé par la variable définie par l'axe z de \mathbb{R}^{I^2} , alors sa représentation dans \mathbb{R}^{I^2} est l'opérateur de rang 1 :

$$W_z \cdot D = z \cdot {}^t z \cdot D \quad (21.30)$$

L'inertie projetée des variables du roupe K_g sur z est égale au produit scalaire entre $W_j \cdot D$ et $z \cdot {}^t z \cdot D$, c'est-à-dire à la projection de $W_j \cdot D$ sur $z \cdot {}^t z \cdot D$. Si, dans \mathbb{R}^{I^2} , deux vecteurs sont orthogonaux, alors les vecteurs associés dans \mathbb{R}^{I^2} le sont également. Il en résulte que la représentation graphique des groupes de variables peut être interprétée comme une projection du nuage N_J sur une suite d'axes orthogonaux.

On peut représenter des groupes en tant qu'image optimale du nuage N_J . Le produit scalaire entre $W_j \cdot D$ et $W_l \cdot D$ est une mesure de liaison entre les groupes de variables j et l . Pour comparer globalement les groupes, on cherche à décrire les **proximités** entre les $W_j \cdot D$ en les projetant sur un espace de faible dimension de \mathbb{R}^{I^2} . Pour ce, deux conditions doivent être respectées.

1. Les angles entre les $W_j \cdot D$ doivent être bien représentés.
2. Il ne convient pas de centrer le nuage N_J .

En A.F.M., on impose aux axes du repère d'être des éléments symétriques de rang 1. Ces derniers sont de la forme $z_s \cdot {}^t z_s \cdot D$. Ils sont associés à des groupes d'une seule variable z_s . Ils s'interprètent à partir de z_s et des liaisons avec les variables initiales. On recherche un repère orthonormé dans \mathbb{R}^{I^2} dont chaque composant est de la forme $z \cdot {}^t z \cdot D$ et qui « ajuste » au mieux le nuage des $W_j \cdot D$. Ce repère est construit progressivement en cherchant un premier vecteur, puis un second orthogonal au premier, et ainsi de suite. Pour y parvenir, on utilise le **critère de maximisation de la somme des projections**.

$$\sum_{j=1}^J \langle W_j \cdot D, z \cdot {}^t z \cdot D \rangle \quad (21.31)$$

Cette somme est égale à l'inertie dans \mathbb{R}^{I^2} des variables de tous les groupes projetées sur z_s .

La représentation des groupes en A.F.M. est :

1. une aide à l'interprétation des autres graphiques ;
2. une image du nuage des groupes optimale en elle-même.

La coordonnée de $\mathbf{W}_j.\mathbf{D}$ sur l'axe factoriel $\mathbf{z}^t.\mathbf{z}.\mathbf{D}$ s'interprète comme :

1. l'inertie de la projection du nuage N_K^j , défini par le groupe j dans \mathbb{R}^I , sur la composante principale \mathbf{z}_s du tableau \mathbf{X} ; c'est la **contribution absolue** du groupe j à l'axe s ;
2. une mesure de liaison \mathcal{L} entre le groupe j et la composante \mathbf{z}_1 de l'A.F.M. ;
3. la projection du groupe j dans l'espace \mathbb{R}^{I^2} .

Du fait de la pondération des groupes, les coordonnées des $\mathbf{W}_j.\mathbf{D}$ sont comprises entre 0 et 1. Selon la répartition de l'inertie des nuages associés N_I^j et N_K^j , un groupe j peut avoir plusieurs coordonnées proches de 1.

En tant que projection sur un sous-espace, la représentation des groupes s'accompagne des aides à l'interprétation usuelles. Du fait de la contrainte sur les axes du repère (éléments symétriques de rang 1), la **qualité de représentation** des $\mathbf{W}_j.\mathbf{D}$ par ces axes, mesurée par $\frac{\text{inertie projetée}}{\text{inertie totale}}$, n'atteint en général pas 1.

Les relations de transition

Soient λ_1 la valeur propre de l'axe s , \mathbf{F}_s l'axe s et \mathbf{G}_s le groupe s , les formules de transition sont :

— pour les points moyens i :

$$\mathbf{F}_s(i) = \frac{1}{\lambda_s} \sum_{j=1}^J \left(\frac{1}{\lambda_1^j} \sum_{k=1}^{K_j} x_{ik} \mathbf{G}_s(k) \right) \quad (21.32)$$

— pour les points partiels i^j :

$$\mathbf{F}_s(i^j) = J \frac{1}{\lambda_s} \frac{1}{\lambda_1^j} \sum_{k=1}^{K_j} x_{ik} \mathbf{G}_s(k) \quad (21.33)$$

Le ratio d'inertie pour un axe s

Soit \mathbf{F}_s l'axe s , les inerties de l'axe s sont :

— l'inertie totale :

$$\sum_{i=1}^I \sum_{j=1}^J (\mathbf{F}_{ijs})^2 \quad (21.34)$$

— l'inertie inter individu :

$$\sum_{i=1}^I \sum_{j=1}^J (\mathbf{F}_{is})^2 \quad (21.35)$$

— l'inertie intra individu :

$$\sum_{i=1}^I \sum_{j=1}^J (\mathbf{F}_{ijs} - \mathbf{F}_{is})^2 \quad (21.36)$$

La contribution du groupe k à l'axe s

La contribution du groupe k à l'axe s vaut :

$$CTR_s(k) = \frac{\mathbf{F}_{ks}}{\sum_{k=1}^K \mathbf{F}_{ks}} \quad (21.37)$$

21.1.2 Cas de variables qualitatives et des tableaux mixtes

Il s'agit de :

1. équilibrer les groupes entre eux ;
2. rechercher des facteurs communs ;
3. comparer globalement les groupes ;
4. construire une représentation superposée ;
5. *etc.*

L'A.F.M. s'applique aux tableaux disjonctifs complets dans lesquels les variables sont structurées en groupes.

Il existe une équivalence entre l'A.C.M. et l'A.C.P. appliqué aux variables indicatrices pondérés de manière adéquate.

De fait, il est possible de traiter **simultanément** les variables quantitatives et qualitatives. Le tableau de données est alors dit « **mixte** ».

L'inertie de k par rapport à l'origine O vaut :

$$\frac{I - I_k}{I} = \sum_{i=1}^I \frac{1}{I} x_{ik} \left(\frac{I_k (I - I_k)}{I^2} \right)^{-1} = i \quad (21.38)$$

avec la modalité k , I_k le nombre d'individus possédant la modalité k , $x_{ik} = 0$ ou $x_{ik} = 1$, est J le nombre de variables actives.

21.1.3 Plan d'une A.F.M.

1. Définir la composition des groupes dans un jeu de données
2. Définir les groupes actifs et supplémentaires
3. Réduire ou non les variables
4. Réaliser l'A.F.M.
5. Choisir le nombre de dimension à interpréter
6. Interpréter le graphe simultané individus vs. variables
7. Étudier les groupes
8. Utiliser des analyses partielles

21.2 Le modèle INDSCAL

INDSCAL signifie *Analysis of Individual Differences in Multidimensional Scaling*. Le modèle INDSCAL fut inventé par J. D. Carroll et J. J. Chang [Carroll et Chang, 1970]. Selon ce modèle, les distances entre individus peuvent se décomposer suivant un certain nombre de « facteurs » communs à tous les groupes, le poids affecté à chaque facteur différant en fonction des groupes :

$$d_j^2(i, l) = \sum_{s=1}^S q_s^j [z_s(i) - z_s(l)]^2 \quad (21.39)$$

avec $z_s(i)$ la valeur du s -ième facteur pour l'individu i avec $s = 1, \dots, S$, q_s^j le poids affecté à z_s par le j -ième groupe, et $d_j(i, l)$ la distance entre les individus i et l induit par le j -ième groupe.

Dans ce modèle, tous les individus ont le **même** poids.

Le modèle a été utilisé par F. Husson et J. Pagès dans le cadre de l'A.F.M. [Husson et Pagès, 2006]. Soit I individus et J caractères, on évalue la dissimilarité entre les individus i et l . Elle est notée : $d_j(i, l)$.

Selon le modèle INDSCAL, la dissimilarité dérive d'une configuration des individus en R dimensions. Soit $z_r(i)$ la coordonnée de l'individu i sur l'axe de rang r par rapport à la configuration. Chaque caractère j accorde un poids spécifique q_r^j à la dimension r :

$$d_j^2(i, l) = \sum_{r=1}^R q_r^j (z_r(i) - z_r(l))^2 + e_j(i, l) = \hat{d}_j^2(i, l) + e_j(i, l) \quad (21.40)$$

avec $e_j(i, l)$ le résidu du modèle et \hat{d}_j^2 le carré de la distance reconstituée par le modèle.

Si $d_j(i, l)$ est une distance euclidienne, alors on peut associer à chaque couple de produit (i, l) un produit scalaire noté $\langle i, l \rangle$ et le modèle peut s'écrire :

$$\langle i, l \rangle_j = \sum_{r=1}^R q_r^j z_r(i) z_r(l) + \varepsilon_j(i, l) \quad (21.41)$$

Warren S. Torgerson (1924-1999) Le produit scalaire est obtenu par la formule de W. S. Torgerson [Torgerson, 1958].

$$\langle i, l \rangle_j = \frac{1}{2} (d_j^2(i, l) - d_j^2(i, \cdot) - d_j^2(\cdot, l) + d_j^2(\cdot, \cdot)) \quad (21.42)$$

avec $d_j^2(i, \cdot) = \frac{1}{I} \sum_{k=1}^I d^2(i, l)$ et $d_j^2(\cdot, \cdot) = \frac{1}{I} \sum_{k=1}^I d^2(i, \cdot)$

On peut écrire le modèle INDSCAL matriciellement :

$$S_j = \mathbf{X} \cdot \mathbf{W}_j \cdot {}^t\mathbf{X} + E_j = \hat{S}_j + E_j \quad (21.43)$$

avec S_j la matrice $(I \times I)$ des produits scalaires des caractères j , \mathbf{X} la matrice $(I \times R)$ des coordonnées des produits dans un espace à R dimension, \mathbf{W}_j une matrice diagonale $(R \times R)$ de poids (ou saliences) et \hat{S}_j la matrice $(I \times I)$ des produits scalaires reconstitués par le modèle.

Pour estimer le modèle INDSCAL, il faut recourir à des algorithmes, car il n'existe pas de solution algébrique. La méthode de J. D. Carroll et J. J. Chang minimise par un algorithme de type moindres carrés alternés [Carroll et Chang, 1970] :

$$\text{Strain} = \frac{1}{J} \sum_{j=1}^J \|\mathbf{S}_j - \mathbf{X} \cdot \mathbf{W}_j \cdot {}^t\mathbf{X}\|^2 \quad (21.44)$$

le carré de la norme d'une matrice \mathbf{A} de terme général a_{ij} étant $\|\mathbf{A}\|^2 = \sum_{i,j} a_{ij}^2$

Le critère Strain demeure le critère le plus intéressant, car il bénéficie d'une interprétation géométrique claire. Elle permet de décomposer ce critère en individus et caractères, et d'interpréter ce critère comme un rapport d'inertie projetée sur l'inertie totale.

$$\frac{\text{inertie projetée}}{\text{inertie totale}} = \frac{\sum_{j=1}^J \|\hat{S}_j\|^2}{\sum_{j=1}^J \|\mathbf{S}_j\|^2} \quad (21.45)$$

Cela permet de généraliser Strain :

$$\text{Strain} = \frac{\sum_{j=1}^J \|\mathbf{S}_j - \hat{S}_j\|^2}{\sum_{j=1}^J \|\mathbf{S}_j\|^2} \quad (21.46)$$

avec S_j non normé.

La qualité de représentation du caractère j vaut :

$$\frac{\|\hat{S}_j\|^2}{\|S_j\|^2} \quad (21.47)$$

21.3 L'analyse factorielle multiple et l'approche P.L.S.

L'approche *Partial Least Squares* (P.L.S.) est une méthode de régression inventée par H. Wold [Wold, 1975] [Wold, 1982] [Wold, 1985]. Il a été reprise Herman Wold dans l'A.F.M. par Jérôme Pagès et Michel Tenenhaus [Pagès et Tenenhaus, 2002]. (1908-1992) Lorsqu'il n'existe pas de relations de causalité entre les blocs, on peut adopter le point de vue de l'approche P.L.S. Pour y parvenir, H. Wold propose de constituer Michel Tenenhaus (né en 1944) un **bloc fictif**. L'approche P.L.S. permet de retrouver diverses méthodes.

On pose :

- X le tableau complet ;
- I l'ensemble des individus ;
- K l'ensemble des variables (tous groupes confondus) ;
- J l'ensemble des sous-tableaux ;
- K_j l'ensemble des variables du groupe j ;
- X_j le tableau associé au groupe j ;
- v_k la variable du groupe K_j avec $k \in K_j$.

L'ensemble des notations désigne à la fois un ensemble mathématique et son cardinal.

Les variables sont supposées centrées réduites. Les individus ont tous les mêmes poids $\frac{1}{I}$.

Le cardinal de k vaut :

$$K = U_j.K_j \quad (21.48)$$

Appliquée à l'A.F.M., l'approche P.L.S. permet de répondre à la question suivante. **Quelles sont les structures communes aux différents groupes ?**

La mise en évidence de plusieurs structures communes conduit à rechercher dans chaque groupe j une suite de combinaisons linéaires des variables de ce groupe $\{F_s^j; s = 1, S\}$ telles que, entre les groupes, les combinaisons ayant le même rang s $\{F_s^j; j = 1, J\}$ soient corrélées le plus possible.

On exprime que F_s^j est combinaison linéaire des variables v_k du groupe K_j en écrivant :

$$F_s^j = \sum_{k \in K_j} a_{k \cdot}^s v_k \quad (21.49)$$

Il existe plusieurs méthodes pour résoudre le problème.

21.3.1 Pondération des variables

En A.F.M., chaque variable v_k du groupe K_j est affectée du poids m_k

$$m_k = \frac{1}{\lambda_1^j} \quad (21.50)$$

avec λ_1^j la première valeur propre de l'A.C.P. séparée du groupe K_j . En faisant l'A.C.P. séparée du groupe K_j avec les poids m_k , la première valeur propre vaut 1. Avec cette pondération, l'inertie axiale maximale de la configuration des individus, ou celles des variables, vaut 1.

M_j est la matrice égale à l'identité à un coefficient près des poids des variables du groupe K_j .

M est la matrice diagonale des poids des variables tous groupes confondus.

21.3.2 Mesure de liaison entre une variable z et un groupe K_j

En A.F.M., la mesure de liaison entre une variable z et un groupe K_j est définie par :

$$\mathcal{L}\}(z, K_j) = \sum_{k \in K_j} m_k r^2(z, v_k) = \frac{1}{\lambda_1^j} \sum_{k \in K_j} r^2(z, v_k) \quad (21.51)$$

avec v_k une variable centrée réduite et r la corrélation.

N.B. La mesure de liaison est apparentée à la redondance R_d :

$$R_d(z, K_j) = \frac{1}{K_j} \sum_{k \in K_j} r^2(z, v_k) \quad (21.52)$$

Les deux mesures varient entre 0 et 1.

Les deux mesures vérifient la propriété :

$$\forall k \in K_j, \mathcal{L}\}(z, K_j) = 0 \Leftrightarrow R_d(z, K_j) = 0 \Leftrightarrow r(z, v_k) = 0 \quad (21.53)$$

La liaison entre z et K_j est nulle lorsque z est non corrélée à chaque variable du groupe K_j .

Pour les deux mesures, la valeur de 1 correspond à une liaison maximale, mais ce maximal n'a pas la même signification pour R_d et $\mathcal{L}\}$.

$$\forall k \in K_j, R_d(z, K_j) = 1 \Leftrightarrow r(z, v_k) = 1 \quad (21.54)$$

N.B. La situation dans laquelle toutes les variables du groupe K_j sont parfaitement corrélées entre elles, ne correspond pas à une situation concrète.

$$\mathcal{L}\}(z, K_j) = 1 \Leftrightarrow z \text{ est la première composante principale de } K_j \quad (21.55)$$

Les variables synthétiques sont liées le plus possible aux variables initiales.

21.3.3 Recherche des variables générales

En A.F.M., on cherche une suite de variables générales (ou auxiliaires) liées le plus possible avec l'ensemble des groupes.

La liaison entre une variable z et l'ensemble des groupes K_j est :

$$\text{liaison}(z, U_j, K_j) = \sum_{j=1}^J \mathcal{L}\}(z, K_j) \quad (21.56)$$

La variable générale de rang s est définie par :

$$\sum_{j=1}^J \mathcal{L}\}(z, K_j) \text{ maximal} \quad (21.57)$$

avec $\text{var}[z_s] = 1$ et $\forall t < s, r(z_s, z_t) = 0$.

En remarquant que :

$$\mathcal{L}\}(z_s, K_j) = \frac{1}{I^2} {}^t z_s \cdot X_j \cdot M_j \cdot {}^t X_j \cdot z_s \quad (21.58)$$

la quantité à maximiser s'écrit :

$$\sum_{j=1}^J (z_s, K_j) = \frac{1}{I^2} {}^t z_s \cdot X_j \cdot M_j \cdot {}^t X_j \cdot z_s \quad (21.59)$$

z_1 est la composante principale normée de rang s issue de l'A.C.P. avec des variables pondérées par M du tableau complet X .

En notant λ_s , la valeur propre de rang s de cette A.C.P. et F_s la composante principale associée, on a :

$$F_s = \sqrt{\lambda_s} z_s \quad (21.60)$$

L'intérêt de cette approche par rapport à celle de J. D. Carroll est double [Carroll et Chang, 1970].

1. Elle conduit à des variables générales plus liées aux variables initiales tout en assurant, du fait des poids m_k , une forme d'équilibre entre les groupes de variables.
2. Étant des composantes principales, les variables générales bénéficient de toutes les propriétés de l'A.C.P.

21.3.4 Recherche des variables canoniques

Suivant le principe de J. D. Carroll, on associe, à chaque variable générale z_s , une variable canonique F_s^j dans chaque groupe j . En A.F.M., cette variable est définie par :

$$F_s^j = \frac{1}{\sqrt{\lambda_s}} \frac{1}{I} X_j \cdot M_j \cdot {}^t X_j \cdot z_s \quad (21.61)$$

$$F_s^j = \frac{1}{\lambda_1} \frac{1}{I} X_j \cdot M_j \cdot {}^t X_j \cdot z_s \quad (21.62)$$

$$F_s^j = \frac{1}{\lambda_s \lambda_1^j} \frac{1}{I} X_j \cdot {}^t X_j \cdot F_s \quad (21.63)$$

À un coefficient près, on retrouve la première composante P.L.S. de la régression P.L.S de F_s en fonction de X_j .

Par rapport à la démarche de J. D. Carroll, l'intérêt de la démarche est double.

1. Elle conduit à des variables canoniques plus liées aux variables initiales.
2. Les F_s^j fournissent chacun une représentation des individus qui peut être superposée à celle de l'A.C.P. F_s grâce aux deux propriétés suivantes.

Propriété 1

$$\sum_{j=1}^J F_s^j = \sum_{j=1}^J \frac{1}{\lambda_s} \frac{1}{I} X_j \cdot M_j \cdot {}^t X_j \cdot F_s = \frac{1}{\lambda_s} \frac{1}{I} X \cdot M \cdot {}^t X \cdot F_s = F_s \quad (21.64)$$

Au coefficient $\frac{1}{J}$ près, la représentation de l'individu i par F_s est au barycentre des représentations de i par F_s^j . En pratique, on dilate les F_s^j selon le coefficient J pour que $F_s(i)$ soit un barycentre exact.

$$F_s(i) = \frac{1}{J} \sum_{j=1}^J J \cdot F_s^j(i) \quad (21.65)$$

En A.F.M., on dit que l'image globale d'un individu est au centre de gravité de ses images partielles. L'**image globale** fournit le point de vue de l'ensemble des groupes. L'**image partielle** fournit le point de vue de l'un des groupes.

Propriété 2

Soit $G_s(k)$ la coordonnée de la variable v_k sur l'axe de rang s c'est-à-dire le coefficient de corrélation entre v_k et F_s . On note $v_k(i)$ la valeur de v_k pour l'individu.

On applique à l'A.F.M. la relation de transition usuelle de l'A.C.P. :

$$F_s(i) = \frac{1}{\lambda_s} \sum_{j=1}^J \frac{1}{\lambda_1^j} \sum_{k \in K_j} v_k(i) G_s(k) \quad (21.66)$$

Cette relation peut être restreinte au groupe j :

$$F_s^j(i) = \frac{1}{\lambda_s} \frac{1}{\lambda_1^j} \sum_{k \in K_j} v_k(i) G_s(k) \quad (21.67)$$

Cette définition de F_s^j est équivalente à celle de données précédemment dans laquelle apparaît la quantité :

$$\frac{1}{I} X_j \frac{F_s}{\sqrt{\lambda_s}} \quad (21.68)$$

dont le k -ième terme n'est autre que $G_s(k)$.

La relation restreinte au groupe j permet une interprétation des F_s^j d'un point de vue analytique. Les coefficients $G_s(k)$ sont proportionnels à ceux de la combinaison linéaire exprimant F_s^s en fonction de v_k . Il existe également un point de vue géométrique. Un individu partiel i^j est du côté des variables pour lesquelles il a une forte valeur et à l'opposé de celles pour lesquelles il a une faible valeur.

Matriciellement, F_s et F_s^j sont des combinaisons linéaires des variables initiales :

$$F_s = X \cdot \varphi_s \quad (21.69)$$

et

$$F_s^j = X_j \cdot \varphi_s^j \quad (21.70)$$

φ_s^j est le fragment de φ_1 correspondant aux variables du groupe j .

21.3.5 Régression P.L.S. et A.F.M.

On note :

- C_s^j la composante P.L.S. de rang s pour le groupe j . Elle est analogue à la variable F_s^j pour l'A.F.M. ;
- M_j la matrice diagonale des poids des variables.

N.B. 1. Elle est en général confondue avec l'identité dans les présentations des méthodes P.L.S.

N.B. 2. En A.F.M., elle vaut la matrice identité au coefficient $\frac{1}{\lambda_1^j}$ près.

Cas général

La première composante de la régression P.L.S. vaut :

$$C_1^x \propto \frac{1}{I} X \cdot X \cdot Y \quad (21.71)$$

En A.F.M., la relation devient :

$$F_1^x = \frac{1}{\lambda_1} \frac{1}{\lambda_1^x} \frac{1}{I} X \cdot X \cdot 4F_1 \quad (21.72)$$

Attention ! Les deux formules ne sont pas identiques.

On peut calculer :

- $r(., Y)$ la corrélation entre la variable étudiée et la variable Y à prédire ;
- $V_r = \sum_{k=1} r(., v_k)^2$ la variance de l'ensemble des variables X expliquée par la variable étudiée ;
- $\mathcal{L}\}(., X)$ la mesure de liaison entre X et la variable étudiée.

La deuxième composante de la régression P.L.S. vaut :

$$C_1^x \propto \frac{1}{I} X \cdot {}^t X \cdot C_1^y \quad (21.73)$$

$$C_1^y \propto \frac{1}{I} Y \cdot {}^t Y \cdot C_1^x \quad (21.74)$$

De même,

$$F_1^x = \frac{1}{\lambda_1} \frac{1}{\lambda_1^x} \frac{1}{I} X \cdot {}^t X \cdot F_1 \quad (21.75)$$

$$F_1^y = \frac{1}{\lambda_1} \frac{1}{\lambda_1^y} \frac{1}{I} Y \cdot {}^t Y \cdot F_1 \quad (21.76)$$

Cas de J groupes de variables

Dans l'approche P.L.S., un modèle relie ces groupes de la façon suivante :

1. Chaque groupe est représenté par une variable latente.
2. Les variables latentes sont reliées entre elles par un ensemble de relations linéaires.

L'algorithme recherche l'ensemble des variables latentes de façon telle que :

1. chaque variable latente « représente » bien le groupe en ce sens qu'elle est corrélée aux variables de ce groupe ;
2. les variables latentes reliées par les équations du modèle sont corrélées entre elles.

21.4. L'ANALYSE FACTORIELLE MULTIPLE HIÉRARCHIQUE (A.F.M.H.)²³

Représentation des variables Les composantes P.L.S. successives d'un même groupe sont orthogonales. Il est possible de représenter l'ensemble des variables, tous groupes confondus, en projection sur un couple de composantes d'un même groupe de la façon usuelle.

- La coordonnée de la variable v_k sur la composante C_s^j vaut $r(v_k, C_s^j)$.
- L'ensemble des points représentant les variables s'interprète comme une projection du nuage des variables définis à partir de l'ensemble des coordonnées.

Représentation des individus Les représentations des individus dans le cadre de l'approche P.L.S. consistent à croiser des variables latentes de même rang pour deux groupes différents. Ces variables ne sont pas orthogonales. Elles sont corrélées par construction. Il n'existe pas d'autres représentations.

21.4 L'analyse factorielle multiple hiérarchique (A.F.M.H.)

L'analyse factorielle multiple hiérarchique² a été inventé par S. Le Dien (ou Lè) et J. Pagès [Christaller, 1950].

Sébastien Le Dien
(ou Lè) (né en
1943)

Il existe une difficulté lorsque l'on veut tenir compte d'une hiérarchie comprenant la notion de groupes et de sous-groupes de variables. Par exemple, des questionnaires sont souvent structurés en thèmes et en sous-thèmes. Dans ce cas, il ne s'agit pas de partitionner les variables, mais d'obtenir une **partition emboîtées** :

1. en équilibrant les thèmes et les sous-thèmes ;
2. en fournissant des sorties (graphiques ou indicateurs numériques) permettant de prendre en compte dans l'interprétation la structure *a priori* sur les variables.

On note :

- X le tableau de données individus vs. variables ;
- I l'ensemble des individus ;
- D la matrice diagonale des poids des individus ;
- p_i le poids de l'individu i ;

$$\sum_{i \in I} p_i = 1 \quad (21.77)$$

- K l'ensemble des variables ;

2. En anglais : *Hierarchical Multiple Factor Analysis* (H.M.F.A.)

— J le nombre de groupes partitionnant les K variables.

En A.F.M.H., on ne considère plus une seule partition, mais **plusieurs partitions emboîtées**. Soient deux partitions $\mathcal{P}_1 = \{A_1, \dots, A_{J_1}\}$ et $\mathcal{P}_2 = \{B_1, \dots, B_{J_2}\}$ sur un même ensemble de variable K . \mathcal{P}_1 est emboîtée dans \mathcal{P}_2 , c'est-à-dire $\mathcal{P}_1 \subset \mathcal{P}_2$. Si tout élément \mathcal{P}_1 est inclus dans un seul élément de \mathcal{P}_2 .

$$\forall j_1 = 1, \dots, J_1, \exists! j_2 / A_{j_1} \subset B_{j_2} \quad (21.78)$$

Plus généralement, soient P partitions sur un même ensemble de variables K , une suite de partitions emboîtées à P éléments est un ensemble de partitions $\{\mathcal{P}_1, \dots, \mathcal{P}_P\}$ noté S_P , tel que :

$$\mathcal{P}_1 \subset \dots \subset \mathcal{P}_P \quad (21.79)$$

Cette suite S_P , en tant que hiérarchie sur les variables K , peut être représentée sous la forme d'un arbre hiérarchique dont les éléments terminaux correspondent chacun à une partition.

$$\text{indice de niveau} = \text{niveau de partition} \quad (21.80)$$

Les groupes et variables sont repérés par leur numéro de nœud, le dernier nœud correspondant à l'ensemble des variables K . Il peut être utile de les repérer par leur numéro de groupe associé à leur partition.

Par un niveau de partition donné, l'appartenance d'une variable à une classe est définie à l'aide de la fonction f_p . Par définition, f_p est l'application qui, à la variable k , associe l'indice du groupe auquel elle appartient au niveau de partition p .

On note J_P l'ensemble des groupes de la partition \mathcal{P}_{p_p} , K_j^p l'ensemble des variables du groupe j au niveau de partition p et X_j^p le tableau associé à K_j^p .

$$\forall p = 1, \dots, P, K = \bigcup_{j=1}^{J_p} K_j^p \quad (21.81)$$

N.B. Les symboles I , J , K , J_p et K_j^p désignent à la fois l'ensemble et son cardinal.

À toute suite de partitions S_p , on associe généralement deux partitions supplémentaires :

1. la partition la plus fine (niveau d'agrégation 0) dans laquelle chaque variable est un groupe ;
2. la partition la plus grossière (niveau d'agrégation $P+1$) dans laquelle toutes les variables appartiennent à un même groupe.

21.4. L'ANALYSE FACTORIELLE MULTIPLE HIÉRARCHIQUE (A.F.M.H.) 25

L'A.F.M.H. est une analyse du tableau de données X structuré selon la hiérarchie S_p qui peut alors se décomposer de la façon suivante.

Étape 1. On réalise une analyse factorielle de chacun des groupes de variables associés au premier niveau de partition. On stocke les facteurs obtenus.

Étape 2. À chaque nœud de l'indice de niveau 2, on réalise une A.F.M. dans laquelle les groupes de variables sont ceux qui constituent le nœud. En pratique, à ce niveau, les variables d'un groupe sont remplacées par les facteurs issus de son A.C.P. À leur tour, les facteurs de ces A.F.M. sont stockés : ce sont eux qui représenteront les groupes de variables associés aux nœuds dans les A.F.M. de l'étape suivante.

Étape p . On réitère l'étape $p-1$ jusqu'à l'indice de niveau le plus élevé. À chaque étape, les variables sont remplacées par les facteurs issus des A.F.M. réalisées à l'étape précédente.

Lors de la dernière étape P , les principaux facteurs de variabilité de l'A.F.M.H. sont obtenues à l'issue de l'A.F.M. des groupes de facteurs principaux associés aux nœuds de l'indice de niveau P . Cette étape, qui correspond à une A.C.P. pondérée du tableau de données X , constitue l'analyse globale de l'A.F.M.H.

On utilise pour l'analyse globale de l'A.F.M.H. les notations usuelles de l'A.C.P. On note u_s l'axe principal de rang s du nuage des individus. F_s est le facteur associé, c'est-à-dire le vecteur des coordonnées des individus sur u_s . On note v_s l'axe principal de rang s du nuage des variables. G_s est le vecteur des coordonnées des variables sur v_s . On note λ_s la s -ième valeur propre.

Le calcul des principaux facteurs de variabilité de l'A.F.M.H. passe par la réalisation d'A.C.P. pour les nœuds correspondant à \mathcal{P}_1 et d'A.F.M. séparées pour les autres nœuds. Les valeurs propres de ces analyses permettent de se faire une idée précise des structures d'inertie par chacun des groupes de variables.

21.4.1 Équivalence avec une A.C.P.

L'A.F.M. étant une A.C.P. pondérée, l'analyse globale de l'A.F.M.H. peut être présentée comme une A.C.P. dans laquelle les variables sont pondérées de telle sorte que le rôle de chaque groupe soit équilibré au sein de sa partition, et ce à tous les niveaux de partition.

Dans l'algorithme présenté, le fait de travailler à partir des facteurs principaux non normés assure l'équilibre entre les groupes de variables. Il faut expliciter le calcul des poids des variables de base à chaque étape.

L'A.F.M. pondère les variables d'un même groupe j par $\frac{1}{\lambda_1^j}$ dans laquelle λ_1^j désigne la première valeur de l'analyse factorielle du groupe j .

Soit $m_0(k)$ le poids initial accordé à la variable k . En général, toutes les variables ont un **poids initial** égal à 1.

La détermination des poids des variables en A.F.M.H. se déroule en P étapes.

Étape 1. $\forall j \in \{1, \dots, J_1\}$, on appelle $\lambda_1^{j(1)}$ la première valeur propre issue de l'A.C.P. du groupe j de \mathcal{P}_1 . Cette valeur propre peut également être notée $\lambda_1^{f_1(x)}$ si la variable k appartient au groupe j de \mathcal{P} . Pour tout $k \in \{1, \dots, K\}$, on calcule :

$$m_1(k) = m_0(k) \frac{1}{\lambda_1^{f_1(k)}} \quad (21.82)$$

$$\dots \quad (21.83)$$

Étape p . $\forall j \in \{1, \dots, J_p\}$, on appelle $\lambda_1^{j(p)}$ la première valeur propre issue de l'A.C.P. pondérée du groupe j de \mathcal{P}_p par l'application m_{p-1} . Cette valeur propre peut également être notée $\lambda_p^{f_p(k)}$, on calcule :

$$m_p(k) = m_{p-1}(k) \frac{1}{\lambda_1^{f_p(k)}} = \prod_{h=1}^p \frac{1}{\lambda_1^{f_h(k)}} \quad (21.84)$$

$$\dots \quad (21.85)$$

Étape P . Le poids $m(k)$ de la variable k dans l'A.F.M.H. est donnée par la formule suivante :

$$m(k) = m_P(k) = m_0(k) \prod_{h=1}^p \frac{1}{\lambda_1^{f_h(k)}} \quad (21.86)$$

À un niveau de partition p donné et pour un groupe de variables j donnée ($j = 1, \dots, J_p$), le rôle de chacun des sous-groupes constituant le groupe j au niveau de partition $p - 1$ est équilibré. Aucun d'eux ne peut être prépondérant dans la première direction d'inertie du nuage décrit par les variables du groupe j . Le rôle de chacun des groupes est ainsi équilibré par nœud et ce quel que soit le niveau de partition.

À la fonction de pondération m , on associe $M \in \mathcal{M}(K \times K)$ avec $\mathcal{M}(K \times K)$ désignant l'ensemble des matrices de dimension $K \times K$, la matrice diagonale des poids de variables de l'A.F.M.H., définie par :

$$\forall k = 1, \dots, K, M(k, k) = m(k) \quad (21.87)$$

À la fonction m_P , on associe $\mathcal{M}_j^p \in \mathcal{M}(K_j^p \times K_j^p)$, la matrice diagonale des poids des variables du groupe j au niveau de partition p , définie par :

$$\text{for all } k = 1, \dots, K_j^p, \mathcal{M}_j^p(k, k) = m_p(k) \quad (21.88)$$

21.4.2 Équivalence avec une A.F.M.

L'A.F.M.H. est également une A.F.M. des groupes de variables associés aux nœuds de l'indice de niveau P . en A.F.M.H., l'interprétation des résultats se fait de manière « descendante », en partant du niveau le plus synthétique.

21.4.3 A.F.M.H., vue comme une analyse factorielle

Le cœur de l'A.F.M.H. repose sur une analyse factorielle pondérée du tableau de données X . Cette analyse globale de l'A.F.M.H. est une A.C.P. pondérée lorsque toutes les variables sont quantitatives. L'A.F.M.H. fournit les résultats classiques de l'A.C.P., à savoir des représentations graphiques du nuage des individus N_I et du nuage des variables N_K .

On note :

- u_s (resp. v_s) le vecteur propre de l'axe s ;
- F_s (resp. G_s) l'axe factoriel s ;
- λ_s la valeur propre de l'axe s .

21.4.4 Sorties spécifiques de l'A.F.M.H.

Considérer une suite de partitions emboîtées sur les variables engendre des sorties supplémentaires.

Représentation superposée des nuages d'individus partiels

Par définition, un individu partiel associé au groupe j de \mathcal{P}_p est un individu décrit par les seules variables de K_j^p . dans l'A.F.M., chaque individu possède autant de représentations partielles qu'il y a de groupes de variables. La représentation superposée de ces individus et leurs représentations partielles est telle que chaque individu est au centre de gravité de ses représentations partielles.

En A.F.M.H., le problème de la représentation graphique se complique, puisque l'on considère non plus une seule, mais plusieurs partitions. Leur prise en compte suggère autant de représentations partielles qu'il existe de groupes de variables distincts toutes partitions confondues ou, de façon équivalente, qu'il y a de nœuds dans la suite de partitions S_P représentée sous forme d'arbre hiérarchique.

Le principe de la représentation superposée des P nuages d'individus partiels généralise celui de l'A.F.M. Pour représenter l'ensemble des individus partiels, on les plonge dans un même espace \mathbb{R}^K , c'est-à-dire l'espace des individus décrits par l'ensemble des variables. Pour ce faire, l'espace \mathbb{R}^K est décomposé successivement en sous-espaces orthogonaux de plus en plus petits, suivant la décompo-

sition induite par la suite S_P , parcourue de façon permanente, c'est-à-dire de \mathcal{P}_P à \mathcal{P}_1 .

Étape 1. Partant du niveau de partition P plus général, on décompose \mathbb{R}^K en J_P sous-espaces orthogonaux deux à deux et isomorphes aux espaces $\mathbb{R}^{K_j^P}$:

$$\mathbb{R}^K = \bigoplus_{j=1, \dots, J_P} \mathbb{R}^{K_j^P} \quad (21.89)$$

Étape 2. Le j -ième groupe de variables au niveau de partition P se décompose en sous-groupes de \mathcal{P}_{P-1} , pour $p = P$, $\mathbb{R}^{K_j^P}$ peut se décomposer en :

$$\mathbb{R}^{K_j^P} = \bigoplus_{j' \in \mathcal{D}_j^P} \mathbb{R}^{K_j^{P-1}} \quad (21.90)$$

avec \mathcal{D}_j^P désignant à la fois l'ensemble des indices des sous-groupes de \mathcal{P}_{P-1} issus de la décomposition du groupe j de \mathcal{P}_P et son cardinal. On en déduit une décomposition de \mathbb{R}^K en somme directe de J_{P-1} sous-espaces orthogonaux deux à deux et isomorphes aux espaces $\mathbb{R}^{K_j^{P-1}}$

$$\mathbb{R}^K = \bigoplus_{j=1, \dots, J_P} \bigoplus_{j' \in \mathcal{D}_j^P} \quad (21.91)$$

et

$$\mathbb{R}^{K_j^{P-1}} = \bigoplus_{j=1, \dots, J_{P-1}} \bigoplus_{j' \in \mathcal{D}_j^{P-1}} \quad (21.92)$$

Étape P . $\mathbb{R}^{K_j^2}$ se décompose en :

$$\mathbb{R}^{K_j^2} = \bigoplus_{j' \in \mathcal{D}_j^2} \mathbb{R}^{K_{j'}^1} \quad (21.93)$$

$$\Rightarrow \mathbb{R}^K = \bigoplus_{j=1, \dots, J_P} \bigoplus_{j_1 \in \mathcal{D}_j^P} \dots \bigoplus_{j_{P-1} \in \mathcal{D}_{j_{P-2}}^2} \mathbb{R}^{K_{j_{P-1}}^1} = \bigoplus_{j=1, \dots, J_1} \mathbb{R}^{K_j^1} \quad (21.94)$$

Dans \mathbb{R}^K , les coordonnées du nuage des individus partiels associés au j -ième groupe de \mathcal{P}_P sont contenues dans la matrice $\mathcal{M}(I, K)$ dans laquelle X_j^P est complétée par des 0. Cette représentation simultanée est artificielle et inutilisable directement, mais elle sert de base à une représentation simultanée obtenue par projection sur les axe principaux u_s issus de l'analyse globale de N_I par l'A.F.M.H.

21.4. L'ANALYSE FACTORIELLE MULTIPLE HIÉRARCHIQUE (A.F.M.H.) 29

1. Une première solution consiste à représenter les individus partiels de telle sorte que, quel que soit le niveau de partition considéré, chaque individu soit au centre de gravité de ses représentations partielles. On bénéficie alors de la propriété barycentrique de l'A.F.M., à chaque niveau de partition. $\forall i = 1, \dots, I, \forall p = 1, \dots, P$, i est l'isobarycentre de $\{i_j^p, j = 1, \dots, J_p\}$ avec i_j^p la représentation partielle de l'individu i associée au j -ième groupe de variable de \mathcal{P}_p .
 2. Une second solution consiste à représenter les individus partiels de telle sorte que :
 - chaque individu est au centre de gravité de ses représentations partielles au niveau de partition P ;
 - chaque individu partiel associé à un groupe j de \mathcal{P}_p ($1 \leq p \leq P$) est au centre de gravité de « ses » représentations partielles, induites par la décomposition de j en sous-groupes de \mathcal{P}_{p-1} .
- $\forall i = 1, \dots, I$, on déduit que :
- i est l'isobarycentre de $\{i_j^P, j = 1, \dots, J_P\}$;
 - $\forall p = 2, \dots, P, \forall j = 1, \dots, J_P, i_j^p$ est l'isobarycentre de $\{i_{j'}^{p-1}, j' \in \mathcal{D}_j^p\}$

Aides à l'interprétation à la représentation superposée

Pour calculer les indicateurs, on utilise une **fonction de pondération** qui traduit la notion d'équilibre entre les sous-groupes de variables « issus » d'un même nœud. Cette pondération est telle que, quel que soit $i \in I$, les individus partiels « issus » d'un même nœud, c'est-à-dire associés aux sous-groupes de variables d'un même nœud, ont un **poids identique**. La somme de leur poids égale au poids de l'individu décrit par le groupe de variables associés au nœud dont ils sont issus.

Formellement, si p_j^i est le poids de l'individu partiel i associé au groupe de variables K_j^p , alors :

$$\forall i \in I, \forall j \in \mathcal{D}_{j'}^{p+1}, p(i_j^p) = \frac{p_{j'}^{p+1}}{\mathcal{D}_{j'}^{p+1}} \quad (21.95)$$

avec :

$$\forall i \in I, \forall p = 1, \dots, P, \sum_{j \in \mathcal{P}_p} p(i_j^p) = p_i \quad (21.96)$$

Par construction, $\forall i \in I, \forall p = 1, \dots, P$,

$$\sum_{j \in \mathcal{P}_p} p_j^i = \sum_{j' \in \mathcal{P}_{p+1}} \sum_{j \in \mathcal{D}_{j'}^{p+1}} p(i_j^p) = \sum_{j' \in \mathcal{P}_{p+1}} p(i_{j'}^{p+1}) \quad (21.97)$$

cette somme se réduisant pour \mathcal{P}_{P+1} à un terme unique qui est le poids p_i de l'individu i . La partition \mathcal{P}_{P+1} correspond au cas où les variables appartiennent toutes à un même groupe. Dans le cas particulier du nuage de points partiels $N_I^{J_P}$, i_1^{P+1} désigne l'individu i et son coefficient de pondération p_1^{P+1} est alors p_i .

De cette pondération, découlent deux propriétés.

1. Quel que soit le niveau de partition p , l'individu i est au centre de gravité de ses représentations partielles associées aux groupes de \mathcal{P}_p ainsi pondérés.
2. Quel que soit le niveau de partition p , les nuages des individus partiels, $N_I^{J_P} = \{i_j^p, i \in I, j \in J_p\}$, sont centrés.

On peut maintenant décomposer l'inertie. À tous les niveaux de partition p , l'A.F.M.H. calcule l'inertie du nuage des points partiels $N_I^{J_P}$, notée $I_{\text{tot}}(p)$. Un tel nuage de points peut se décomposer en $I \times J_{P+1}$ sous-groupes, de la façon suivante.

1. Dans un premier temps, on regroupe tous les points associés à un même individu i .
2. Dans un second temps, on décompose chacun des I groupes en J_{P+1} sous-ensembles de la forme $\{i_{j'}^p, j' \in \mathcal{D}_{j'}^{p+1}\}$, $i \in I, j' \in \mathcal{P}_{p+1}$.

Quel que soit $p = 1, \dots, P$, on note $\mathcal{P}(N_I^{J_P})$ la partition du nuage des individus partiels $N_I^{J_P}$. Par définition

$$I_{\text{tot}}(p) = \sum_{i \in I} \sum_{j \in \mathcal{P}_p} p(i_j^p) d^2(O, i_j^p) \quad (21.98)$$

d'après le théorème de König-Huygens appliqué au nuage de points $N_I^{J_P}$ partitionné selon $\mathcal{P}(N_I^{J_P})$, I_{tot} se décompose tel que :

$$I_{\text{tot}}(p) = I_{\text{inter}}(p) + I_{\text{intra}}(p) \quad (21.99)$$

$$\Leftrightarrow I_{\text{tot}}(p) = \sum_{i \in I} \sum_{j \in \mathcal{P}_{p+1}} \left(\sum_{j' \in \mathcal{D}_{j'}^{p+1}} p(i_j^p) \right) d^2(O, i_j^p) + \sum_{i \in I} \sum_{j' \in \mathcal{P}_{p+1}} \sum_{j \in \mathcal{D}_j^{p+1}} p(i_j^p) d^2(i_{j'}^p, i_j^{p+1}) \quad (21.100)$$

$$\Leftrightarrow I_{\text{tot}}(p) = \sum_{i \in I} \sum_{j' \in \mathcal{P}_{p+1}} p(i_{j'}^{p+1}) d^2(O, i_{j'}^{p+1}) + \sum_{i \in I} \sum_{j' \in \mathcal{P}_{p+1}} \sum_{j \in \mathcal{D}_j^{p+1}} p(i_j^p) d^2(i_{j'}^p, i_j^{p+1}) \quad (21.101)$$

$$\Leftrightarrow I_{\text{tot}}(p) = \Leftrightarrow I_{\text{tot}}(p+1) + \Leftrightarrow I_{\text{intra}}(p) \quad (21.102)$$

avec $I_{\text{intra}}(p)$ l'inertie intra-classe du nuage des points partiels $N_I^{J_P}$, au niveau de partition p , pondérés par $p(i_j^p)$, regroupés suivant la partition $\mathcal{P}(N_I^{J_P})$.

21.4. L'ANALYSE FACTORIELLE MULTIPLE HIÉRARCHIQUE (A.F.M.H.) 31

Par récurrence, on obtient la formule de décomposition suivante :

$$I_{\text{tot}}(1) = I_{\text{tot}}(N_I) + \sum_{p=1}^P I_{\text{intra}}(p) \quad (21.103)$$

avec $I_{\text{tot}}(N_I) = I_{\text{tot}}(P+1) = \sum_{i \in I} p_i d^2(O, i)$ l'inertie du nuage des individus N_I calculées par rapport à son centre de gravité.

Soit $F_s(i_j^p)$ la coordonnée sur u_s de la représentation partielle de l'individu i associée à K_j^p . une **analyse de variance** des $F_s(i_j^1)$, pondérée par $p(i_j^1)$, permet de retrouver la décomposition de l'inertie de $N_I^{J_1}$. Pour cela, on considère le modèle d'analyse de variance hiérarchique suivant :

$$F_s(i_j^1) = i + \mathcal{P}_P[i] + \mathcal{P}_{P-1}[\mathcal{P}_P[i]] + \dots + \mathcal{P}_1[\dots \mathcal{P}_P[i]] \quad (21.104)$$

avec i le facteur individu, \mathcal{P}_p le facteur de partition associé au niveau de la partition p , et $B[A]$ l'effet du facteur B hiérarchisé dans A .

Les différentes sommes des carrés des écarts associés aux facteurs introduits dans le modèle correspondant chacune à un élément de la décomposition de $I_{\text{tot}/u_s}(1)$ (Tab. 21.2).

Inertie	Facteurs du modèle de l'analyse de variance
$I_{\text{tot}/u_s}(1)$	$F_s[i_j^1]$
$I_{\text{tot}/u_s}(N_I)$	i
$I_{\text{intra}/u_s}(P)$	$\mathcal{P}_P[i]$
...	...
$I_{\text{intra}/u_s}(P)$	$\mathcal{P}_P[\dots \mathcal{P}_P[i]]$
...	...
$I_{\text{intra}/u_s}(1)$	$\mathcal{P}_1[\dots \mathcal{P}_P[i]]$

TABLE 21.2 – Synthèse des inerties utilisées en A.F.M.H.

Variables canoniques

Soit $F_s^{j(p)}$ le vecteur des coordonnées des individus partiels associés au groupe j de \mathcal{P}_p sur l'axe de rang s , alors ce vecteur définit une nouvelle fonction sur les individus, obtenue en projetant le nuage partiel associé au groupe j de \mathcal{P}_p sur u_s , appelée **variable canonique**.

Les corrélations entre variables canoniques et variables générales (c'est-à-dire les facteurs de l'analyse globale) sont calculées à tous les niveaux de partition.

Quel que soit le niveau de partition, ces coefficients de corrélation se commentent comme ceux d'une A.F.M. Le coefficient de corrélation linéaire entre la variable canonique $F_s^{j(p)}$ et le facteur F_s de l'A.F.M.H. montre dans quelle mesure ce dernier peut être considéré comme une structure du groupe j de \mathcal{P}_p . En l'absence d'indicateurs statistiques qui permettraient de sélectionner les corrélations significatives, les coefficients relatifs aux axes de rang élevé servent empiriquement de base de comparaison.

Groupes de variables

En A.F.M.H., il existe une représentation de l'ensemble des groupes, toutes partitions confondues. Le principe de cette représentation repose sur celui de l'A.F.M. : les groupes sont représentés dans l'espace \mathbb{R}^{I^2} par leur matrice des produits scalaires ; leur coordonnées sur les vecteurs de \mathbb{R}^{I^2} induits par les composantes principales de N_K dans \mathbb{R}^I , sont obtenues par projection.

La pondération des variables dans le calcul des produits scalaires entre individus décrits par un même groupe peut être envisagée de deux manières.

1. La première méthode consiste à utiliser les poids « finaux » associés à la fonction de pondération m . Dans ce cas, la projection du j -ième groupe de \mathcal{P}_p sur le vecteur induit par la composante principale v_s s'interprète comme la contribution du groupe à l'inertie de la composante.
2. La seconde méthode consiste à utiliser les poids « instantanés » associés aux fonctions m_p , calculés à chaque étape de l'algorithme. Dans ce cas, la coordonnée du groupe j de \mathcal{P}_p sur le vecteur induit par la composante principale v_s s'écrit :

$$\langle W_j^p \cdot D, v_s \cdot v'_s \cdot D \rangle = \sum_{k \in K_j^p} m_p(k) r^2(k, v_s) \quad (21.105)$$

avec $W_j^p = X_j^p \cdot M_j^p \cdot {}^t X_j^p$ la matrice des produits scalaires associée aux variables du groupe K_j^p pondérées par m_p et $r^2(k, v_s)$. Le second membre de la formule correspond à la mesure de liaison \mathcal{L} classiquement utilisée en A.F.M. pour traduire la liaison entre une variable, ici v_s , et un groupe de variables, ici K_j^p , la fonction de pondération m_p garantissant que la plus grande valeur propre issue de l'analyse factorielle des variables de K_j^p est égale à 1.

Cette seconde solution est choisie pour l'A.F.M.H. Il en résulte la relation suivante entre la matrice des produits scalaires associées aux variables du groupe K_j^p , $p \geq 2$, et celles associées aux sous-groupes j' de \mathcal{P}_{p-1} qui constituent ce

dernier :

$$W_j^p = \frac{1}{\lambda_1^{j(p)}} \sum_{j' \in \mathcal{D}_j^p} W_{j'}^{p-1} \quad (21.106)$$

On en déduit la relation entre les coordonnées d'un groupe j de \mathcal{P}_p , avec $p \geq 2$, et celles des sous-groupes j' de \mathcal{P}_{p-1} qui le constituent :

$$\langle W_j^p \cdot D, v_s \cdot {}^t v_s \cdot D \rangle = \frac{1}{\lambda_1^{j(p)}} \sum_{j' \in \mathcal{D}_j^p} \langle W_{j'}^{p-1} \cdot D, v_s \cdot {}^t v_s \cdot D \rangle \quad (21.107)$$

Géométriquement, dans l'espace des groupes \mathbb{R}^{I^2} , le groupe j de \mathcal{P}_p est à l'intérieur du cône de sommet O et dont la base est engendrée par les points j' de \mathcal{P}_{p-1} appartenant à \mathcal{D}_j^p .

En complément au graphique, on calcule, pour chaque axe, le rapport $\frac{\text{inertie projetée}}{\text{inertie totale}}$, indicateur classique de qualité de représentation, qui, dans le cas d'un groupe, s'interprète comme un cosinus carré.

Facteurs des analyses séparées

Les facteurs normés issus des analyses factorielles réalisées à chaque nœud sont projetés comme des variables illustratives sur les facteurs issus de l'A.F.M.H. Leurs coordonnées s'interprètent comme des coefficients de corrélation.

Parmi les aides à l'interprétation, il existe :

- la qualité de représentation des axes partiels actifs qui permet de comparer la représentation optimale d'un groupe issue de son analyse séparée à celle obtenue par l'A.F.M.H. ;
- les contributions des facteurs des analyses séparées à la formation des facteurs de l'A.F.M.H.

21.5 L'analyse factorielle multiple procustéenne (A.F.M.P.)

L'objectif de l'analyse factorielle multiple procustéenne³ (P.M.F.A.) est de compléter l'A.F.M. par un **ajustement procustéen** de chacun des nuages initiaux sur le nuage moyen de l'A.F.M. [Morand et Pagès, 2007].

21.5.1 L'analyse procustéenne

L'**analyse procustéenne** est une technique pour comparer des formes. Elle eut pour précurseur C. I. Mosier [Mosier, 1939]. Elle est utilisée pour déformer un Charles I. Mosier (?-?)

3. En anglais : *Procustes Mutliple Factor Analysis*

objet afin de le rendre autant que faire se peut semblable à une référence (arbitraire ou non), ne laissant apparaître entre l'objet et la référence que les différences que les transformations autorisées (rotation, translation et mise à l'échelle) n'ont pu gommer [Hurley et Cattell, 1962]. La déformation supprime les différences qui ne sont pas dues à la forme intrinsèque de l'objet. Les différences restantes sont considérées objectives. Elles permettent d'évaluer le degré de ressemblance entre l'objet et la référence.

Remarque. Procuste est un brigand de la mythologie grecque. Il forçait ses victimes à s'allonger sur un lit et modifiait leur taille par la force pour qu'elles correspondissent à la taille du lit.

Dans l'analyse procustéenne généralisée⁴ (A.P.G.), on compare l'ensemble de nuages de points entre eux [Gower, 1975] [Ten Berge, 1977]. Pour y parvenir, on construit à la fois une représentation superposée des configurations partielles et une représentation de référence qui est, en pratique, le barycentre des configurations partielles. La représentation superposée met en évidence les traits communs aux différentes configurations, et ce sans déformer les configurations partielles initiales.

21.5.2 L'analyse factorielle multiple procustéenne

L'A.F.M.P. allie :

- une représentation de référence avec les propriétés intéressantes de l'A.F.M. ;
- une représentation superposée des configurations partielles non déformées.

On note :

- $\{i; i = 1, \dots, I\}$ l'ensemble des individus ;
- X le tableau complet ;
- J le nombre de sous-tableaux ;
- K_j le nombre de variables du groupe j ;
- X_j le tableau associé au groupe j ;
- i^j l'individu partiel correspondant à la i -ième ligne de X_j .

L'A.P.G. nécessite des sous-tableaux de même dimension. On suppose que cette hypothèse est vérifiée :

$$\forall i = 1, \dots, J, K_j = K \quad (21.108)$$

Il est toujours possible de se ramener à ce cas en prenant $K = \max \{K_j; j = 1, \dots, J\}$ et en complétant les configurations de dimension inférieure par des colonnes de 0.

4. En anglais : *Generalized Procustes Analysis* (G.P.A.)

Les variables sont supposées centrées. On note le produit scalaire usuel entre deux matrices A et B de même dimension : $\langle A, B \rangle = \text{Tr}(A.^t B)$ et $\|A\| = \sqrt{\text{Tr}(A.^t A)}$ la norme associée.

Pour le cas particulier du *mapping*, une configuration partielle est une **nappe**. Chaque sous-tableau X_j contient alors les $K_j = 2$ coordonnées des produits sur la nappe du groupe j .

On recherche une **représentation de référence** à partir du tableau X . Elle est utilisée pour ajuster les configurations partielles. Elle est constituée par les S premières composantes de la représentation moyenne de l'A.F.M. En pratique, on prend généralement $S = K$. Le tableau obtenu des K premières composantes principales de dimension (I, K) , est noté F_K .

N.B. Une représentation de référence est une représentation de compromis, de consensus, une représentation moyenne.

La configuration de l'A.F.M. est choisie comme une représentation de référence, car elle présente des propriétés intéressantes une interprétation directe à partir des variables initiales. En revanche, en A.P.G., la représentation moyenne n'est pas primordiale.

Dans le cas de l'A.P.G., dès que $J > 2$, on utilise un algorithme pour trouver la représentation de référence. Toutefois, si l'algorithme converge, ce n'est pas nécessairement vers un optimum globale. En A.F.M.P., la configuration de référence étant obtenue analytiquement, le problème ne se pose pas.

21.5.3 La représentation superposée

On considère pour chaque groupe j , les tableaux X_j pondérés comme en A.F.M., c'est-à-dire les tableaux $\frac{1}{\sqrt{\lambda_1^j}} X_j$ avec λ_1^j la première valeur propre de l'A.C.P. du tableau X_j . On souhaite les faire coïncider le mieux possible avec la représentation de référence F_K .

On effectue la **rotation procustéenne** H_j qui minimise le critère :

$$\text{Tr} \left(\frac{1}{\sqrt{\lambda_1^j}} X_j \cdot H_j - F_K \right)^t \left(\frac{1}{\sqrt{\lambda_1^j}} X_j \cdot H_j - F_K \right) \quad (21.109)$$

On obtient les **tableaux procustéanisés** par les relations :

$$\frac{1}{\sqrt{\lambda_1^j}} X_j \cdot H_j \quad (21.110)$$

avec $H_j = {}^t V_j \cdot U_j$ dans laquelle les matrices orthogonales V_j et U_j sont obtenues par la décomposition en valeurs singulières de $\frac{1}{\sqrt{\lambda_1^j}} X_j \cdot F_K = V_j \cdot S_j \cdot {}^t U_j$, S_j étant la matrice diagonale des valeurs singulières.

Lors de l'ajustement de chaque sous-configuration sur la configuration de référence, il est possible de choisir d'associer à chaque transformation orthogonale, H_j , une homothétie. Une fois obtenue la transformation orthogonale H_j , le rapport d'inertie⁵ ρ_j est calculé pour minimiser l'écart :

$$\text{Tr} \left(\rho_j \frac{1}{\sqrt{\lambda_1^j}} \cdot X_j \cdot H_j - F_K \right)^t \left(\rho_j \frac{1}{\sqrt{\lambda_1^j}} \cdot X_j \cdot H_j - F_K \right) \quad (21.111)$$

La solution est :

$$\rho_j = \frac{\text{Tr} \left(\frac{1}{\sqrt{\lambda_1^j}} \cdot X_j \cdot H_j - {}^t F_K \right)}{\text{Tr} \left(\frac{1}{\sqrt{\lambda_1^j}} \cdot X_j \cdot {}^t X_j \right)} \quad (21.112)$$

H_j est une matrice orthogonale, la matrice $H_j \cdot {}^t H_j$ est la matrice identité. Ainsi, l'expression précédente peut se réécrire sous la forme suivante :

$$\rho_j = \frac{\|F_K\|}{\left\| \frac{1}{\sqrt{\lambda_1^j}} \cdot X_j \cdot H_j \right\|} \frac{\text{Tr} \left(\frac{1}{\sqrt{\lambda_1^j}} \cdot X_j \cdot H_j - {}^t F_K \right)}{\sqrt{\text{Tr} \left(\frac{1}{\sqrt{\lambda_1^j}} \cdot X_j \cdot {}^t X_j \right) \text{Tr} (F_K \cdot {}^t F_K)}} \frac{\|F_K\|}{\left\| \frac{1}{\sqrt{\lambda_1^j}} \cdot X_j \cdot H_j \right\|} \beta_j \quad (21.113)$$

Ce rapport d'homothétie contient deux termes.

1. $\frac{\|F_K\|}{\left\| \frac{1}{\sqrt{\lambda_1^j}} \cdot X_j \cdot H_j \right\|}$ correspond à une standardisation tenant compte d'une possible homothétie entre les deux configurations.
2. β_j est le **coefficient de similarité de Procruste** entre les deux configurations X_j et F_K . Il indique le niveau de conformisme de la configuration partielle à la représentation de référence.

N.B. 1. K_j n'est pas constante.

N.B. 2. Dans le cas du *mapping*, la standardisation permet de tenir compte de l'utilisation plus ou moins importante de la surface de la nappe par les différents groupes.

Par rapport à l'A.F.M., la représentation superposée obtenue en A.F.M.P. est telle que les nuages partiels n'ont subi aucune autre déformation que les translations orthogonales et les homothéties. Dans le cas bidimensionnel, la représentation plane des nuages partiels n'est absolument pas déformée. Ainsi, chaque

5. En anglais : *scaling factor*

groupe peut visualiser sa propre configuration telle qu'il l'avait fournie. En contrepartie, il n'existe plus de relations de transition partielle. De fait, il n'est pas possible que la coordonnée sur l'axe principal de rang s de l'individu i vu par le groupe j s'exprime comme combinaison linéaire des coordonnées des seules variables du groupe j sur ce même axe.

Le critère global de ressemblance entre toutes les configurations transformées de l'A.P.G. vaut :

$$Sr = \text{Tr} \left(\sum_{j=1}^J \left(s_j \frac{1}{\sqrt{\lambda_1^j}} X_j R_j - \frac{1}{J} \sum_{i=1}^J s_i \frac{1}{\sqrt{\lambda_1^j}} X_i R_i \right) \right) \cdot \left(s_j \frac{1}{\sqrt{\lambda_1^j}} X_j R_j - \frac{1}{J} \sum_{i=1}^J s_i \frac{1}{\sqrt{\lambda_1^j}} X_i R_i \right) \quad (21.114)$$

avec R_j et s_j les transformations orthogonales et les homothéties qui minimisent ce critère.

Soit le modèle d'A.P.G. suivant :

$$s_j X_j R_j = A + E_j \quad (21.115)$$

avec A la configuration compromis, s_j le coefficient d'homothétie, R_j la matrice ($K \times K$) orthogonale et E_j la matrice ($I \times K$) des résidus. En prenant $A = \frac{1}{J} \sum_{i=1}^J s_i \frac{1}{\sqrt{\lambda_1^j}} X_i R_i$, Sr peut s'interpréter comme la somme des carrés des résidus du modèle A.P.G.

Dans le cas de l'A.F.M.P., on peut calculer le critère Sr en prenant $R_j = H_j$ avec H_j la matrice de la transformation orthogonale qui permet d'ajuster au mieux X_j sur F_s , et $s_j = \rho_j$ le rapport d'homothétie qui permet d'ajuster au mieux, après la transformation orthogonale, X_j sur F_s . Dans ce cas, la représentation de référence n'est pas la moyenne des représentations partielles superposées. Dans le cas de l'A.F.M.P., par construction, on minimise la somme des carrés des distances entre chaque représentation partielle et la représentation de référence. Lorsque la représentation de référence est égale à la moyenne des représentations partielles transformées, on minimise le même critère que dans le cas de l'A.P.G. La représentation superposée obtenue en A.F.M.P. n'est pas optimale au sens du critère utilisé dans l'A.P.G., mais cela influence peu les résultats finaux.

Si la représentation obtenue est trop petite, il est possible d'appliquer un rapport de dilatation

$$\hat{X}_j = \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_1^j}} X_j H_j \quad (21.116)$$

21.5.4 Les aides à l'interprétation

On note :

- \hat{X}_j les configurations partielles transformées ;

$$\hat{X}_j = p_j \frac{1}{\sqrt{\lambda_1^j}} X_j \cdot T_j \quad (21.117)$$

avec p_j le coefficient d'homothétie. $p_j = s_j$ en A.P.G. et $p_j = \sqrt{\lambda_1}$ en A.F.M.P.

- T les transformations orthogonales. $T = R_j$ en A.P.G. et $T = H_j$ en A.F.M.P. ;
- $Y = \frac{1}{J} \sum_{j=1}^J \hat{X}_j$ la moyenne des configurations partielles.

Représentation de référence et moyenne des configurations partielles

On minimise l'écart des représentations superposées à la représentation de référence.

On note :

- l'inertie inter-individu :

$$Sr(Y) = \sum_{j=1}^J \text{Tr} \left(\hat{X}_j - Y \right)^t \left(\hat{X}_j - Y \right) \quad (21.118)$$

avec Y la somme des carrés des écarts à la moyenne et F_K la configuration de référence.

- l'inertie des individus par rapport à la représentation de référence :

$$Sr(F_K) = \sum_{j=1}^J \text{Tr} \left(\hat{X}_j - F_K \right)^t \left(\hat{X}_j - F_K \right) \quad (21.119)$$

- le coefficient de distorsion :

$$\text{dist}(F_K, Y) = \text{Tr} \left(\hat{X}_j - F_K \right)^t \left(\hat{X}_j - F_K \right) \quad (21.120)$$

On établit l'égalité :

$$Sr(F_K) = Sr(Y) + J \times \text{dist}(F_K, Y) \quad (21.121)$$

Plus Y est proche de F_K , plus les critères minimisés en A.F.M.P. $Sr(F_K)$ et en A.P.G. $Sr(Y)$ sont proches. Par conséquent, il est d'évaluer par un indice la ressemblance entre Y et F_K . On utilise le coefficient de distorsion normalisé :

$$\frac{\text{Tr} (Y - F_K)^t (Y - F_K)}{(Y \cdot^t Y)^t (F_K \cdot^t F_K)} \quad (21.122)$$

Si cet indicateur est faible, alors le coefficient de distorsion est faible. Cet indice s'interprète comme un indicateur d'homogénéité globale entre les configurations partielles.

Comparaison de deux configurations

On peut comparer de deux configurations :

1. soit deux configurations partielles ;
2. soit une configuration partielle et la configuration de référence.

Les outils sont :

— le coefficient $RV \in [0, 1]$

$$RV(X_j, F_K) = \frac{\text{Tr}(X_j \cdot {}^t X_j \cdot F_K \cdot {}^t F_K)}{\sqrt{\text{Tr}(X_j \cdot {}^t X_j \cdot X_j \cdot {}^t X_j) \text{Tr}(F_K \cdot {}^t F_K \cdot F_K \cdot {}^t F_K)}} \quad (21.123)$$

- $RV = 1$ si X_j peut se déduire de F_K par une homothétie ou une transformation orthogonale.
- $RV = 0$ si toutes les colonnes de X_j sont orthogonales à toutes les colonnes de F_K .

— le coefficient RV standardisé

$$RV_{\text{std}} = \frac{RV(X_j, F_K) - \mathbb{E}(RV(X_j, F_K))}{\sigma(RV(X_j, F_K))} \quad (21.124)$$

avec \mathbb{E} et σ l'espérance et l'écart type du coefficient RV .

— l'indice de similarité de Procuste entre deux configurations A et B :

$$S(A, B) = \frac{\text{Tr}({}^t A \cdot B \cdot H)}{\sqrt{\text{Tr}(A \cdot {}^t A)} \sqrt{\text{Tr}(B \cdot {}^t B)}} \quad (21.125)$$

dans laquelle H est la matrice de transformation orthogonale qui ajuste au mieux B à A , est classiquement employé en A.P.G. pour comparer les sous-configurations deux à deux. Pour chaque sous-configuration en A.F.M.H., on peut calculer un indice de similarité entre X_j et F_K :

$$S(X_j, F_K) = \frac{\text{Tr}(X_j \cdot H_j \cdot {}^t F_K)}{\sqrt{\text{Tr}(X_j \cdot {}^t X_j)} \sqrt{\text{Tr}(F_K \cdot {}^t F_K)}} = \frac{\text{Tr}(\hat{X}_j \cdot {}^t F_K)}{\sqrt{\text{Tr}(\hat{X}_j \cdot {}^t \hat{X}_j)} \sqrt{\text{Tr}(F_K \cdot {}^t F_K)}} \quad (21.126)$$

avec $S(X_j, F_K) \in [0, 1]$.

- $S = 1$ si X_j est identique à F_K .
- $S = 0$ si X_j et F_K sont dans deux sous-espaces orthogonaux.

Décomposition de l'inertie

En A.P.G., on décompose Sr suivant les individus et suivant les sous-configurations.

$$Sr = \sum_{j=1}^J \text{Tr} \left[{}^t (\hat{X}_j - Y) \cdot (\hat{X}_j - Y) \right] \quad (21.127)$$

On sait que :

$$J \times \sum_{j=1}^J \text{Tr} [(X_j - Y) {}^t (X_j - Y)] = \sum_{u < v}^J \text{Tr} [(X_u - X_v) {}^t (X_u - X_v)] \quad (21.128)$$

De cette égalité, on interprète Sr comme un indicateur d'homogénéité globale entre les configurations partielles. Cet indicateur est valable aussi bien pour la représentation superposée de l'A.P.G. que pour celle de l'A.F.M.P.

De plus, chaque terme de la somme $\text{Tr} \left[{}^t (\hat{X}_j - Y) \cdot (\hat{X}_j - Y) \right]$ peut s'interpréter comme une proximité de la j -ième configuration partielle à la représentation moyenne. Plus une sous-configuration contribue à la somme des carrés des résidus, moins elle est en accord avec la représentation moyenne.

Bibliographie

- [Carroll et Chang, 1970] CARROLL, J. D. et CHANG, J. J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckert-young" decomposition. Psychometrika, 35:283–319.
- [Christaller, 1950] CHRISTALLER, W. (1950). Das grundgerüst der räumlichen ordnung in europa. die systeme der europäischen zentralen orte. Frankfurter Geographische Hefte, 24(1):10–97.
- [Escofier et Pagès, 2016] ESCOFIER, B. et PAGÈS, J. (2016). Analyses factorielles simples et multiples. Cours et études de cas. Sciences sup. Dunod, Paris.
- [Gower, 1975] GOWER, J. C. (1975). Generalized procrustes analysis. Psychometrika, 40(1):39–51.
- [Hurley et Cattell, 1962] HURLEY, J. et CATTELL, R. B. (1962). The procustes program : Producing direct rotation to test a hypothesized factor structure. Behavioral Science, 7(2):258–262.
- [Husson et Pagès, 2006] HUSSON, F. et PAGÈS, J. (2006). Aspects méthodologiques du modèle indscal. Revue de statistique appliquée, 54(2):83–100.
- [Lavit, 1988] LAVIT, C. (1988). Analyse conjointe de tableaux quantitatifs. Méthode + Programmes. Masson, Paris.
- [Morand et Pagès, 2007] MORAND, E. et PAGÈS, J. (2007). Analyse factorielle multiple procustéenne. Journal de la société française de statistique, 148(2):65–97.
- [Mosier, 1939] MOSIER, C. I. (1939). Determining a simple structure when loadings for certain tests are known. Psychometrika, 4:149–162.
- [Pagès et Tenenhaus, 2002] PAGÈS, J. et TENENHAUS, M. (2002). Analyse factorielle multiple et approche p.l.s. Revue de statistique appliquée, 50(1):5–33.
- [Ten Berge, 1977] TEN BERGE, J. M. F. (1977). Orthogonal procustes rotation of two or more matrices. Psychometrika, 42(2):267–276.
- [Torgerson, 1958] TORGERSON, W. S. (1958). Theory and Methods of Scaling. John Wiley & Sons, New York.

- [Wold, 1975] WOLD, H. (1975). Modeling in complex situations with soft information. In Third World Congress of Econometric Society, August 21-26, Toronto, Toronto.
- [Wold, 1982] WOLD, H. (1982). Soft Mdeling : the Basic Design and Some Extensions, pages 1–54. North-Holland Publishing Company, Amsterdam.
- [Wold, 1985] WOLD, H. (1985). Partial Least Squares, pages 581–591. John Wiley & Sons, New York.