

# REVUE DE STATISTIQUE APPLIQUÉE

J. PAGÈS

## **Éléments de comparaison entre l'analyse factorielle multiple et la méthode STATIS**

*Revue de statistique appliquée*, tome 44, n° 4 (1996), p. 81-95.

[http://www.numdam.org/item?id=RSA\\_1996\\_\\_44\\_4\\_81\\_0](http://www.numdam.org/item?id=RSA_1996__44_4_81_0)

© Société française de statistique, 1996, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/legal.php>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

## ÉLÉMENTS DE COMPARAISON ENTRE L'ANALYSE FACTORIELLE MULTIPLE ET LA MÉTHODE STATIS

J. Pagès

ENSAR, 65 rue de Saint-Brieuc F-35042 Rennes Cedex

### RÉSUMÉ

L'analyse factorielle multiple (AFM) et la méthode STATIS traitent des tableaux dans lesquels un ensemble d'individus est décrit par plusieurs groupes de variables.

Leurs objectifs sont très proches à bien des égards. Nous mettons en parallèle quelques caractéristiques techniques de ces deux méthodes et en déduisons quelques indications quant au choix entre l'une ou l'autre dans une optique d'application.

**Mots-clés :** *Analyse factorielle, analyse factorielle multiple, STATIS.*

### ABSTRACT

Multiple Factor Analysis (MFA) and STATIS deal with data in which a set of individuals is described by several sets of variables. Their aim are similar.

We compare some technical characteristics of these two methods in order to point out indications about their utilisation.

**Keywords :** *Factor analysis, multiple factor analysis, STATIS.*

### 1. Notations

Nous nous limitons aux tableaux individus  $\times$  variables quantitatives, seuls cas analysables – dans les programmes disponibles actuellement – par les deux méthodes (l'AFM s'applique également aux variables qualitatives). L'ensemble des variables est structuré en groupes.

Soient :  $I$  l'ensemble des individus;  $K$  l'ensemble des variables (supposées ici centrées-réduites) tous groupes confondus;  $K_j$  l'ensemble des variables du  $j^{\text{e}}$  groupe;  $X$  le tableau complet de terme général  $x_{ik}$ , valeur de l'individu  $i$  pour la variable  $k$ ;  $X_j$  le tableau (ou sous-tableau de  $X$ ) associé au groupe  $j$ ;  $J$  l'ensemble des groupes de variables.

$I$ ,  $K$ ,  $K_j$ , et  $J$  représentent à la fois un ensemble et son cardinal.

L'individu  $i$  correspond à une ligne du tableau  $X$ ; à cet individu, dit «moyen» ou «compromis», sont associés  $J$  individus dits «partiels», notés  $i^j$ , correspondant aux  $i^{\text{e}}$  lignes des tableaux  $X_j$ .

A  $I$  (resp.  $K$ ) est associé le nuage  $N_I$  (resp.  $N_K$ ) dans l'espace  $R^K$  (resp.  $R^I$ ). A  $I$  sont aussi associés les nuages  $N_I^j$ , dits nuages partiels, dans  $R^{K_j}$ . Les nuages  $N_I^j$  peuvent être représentés dans  $R^K$  puisque  $R^K = \oplus R^{K_j}$ . Enfin, les variables du groupe  $j$  induisent dans  $R^I$  le nuage  $N_K^j$ ;  $N_K = \cup N_K^j$ .

Le  $s^{\text{e}}$  facteur sur  $I$  de l'analyse de  $X$  (par l'AFM ou par STATIS) est noté  $F_s$ ; il est associé à la valeur propre  $\lambda_s$ ; la  $s^{\text{e}}$  valeur propre de l'ACP séparée de  $X_j$  est notée  $\lambda_s^j$ .

Les poids initiaux des variables du groupe  $j$  sont rassemblés dans la matrice diagonale  $M_j$ ; les poids des individus sont rassemblés dans la matrice diagonale  $D$ ; on note  $W_j = X_j M_j X_j'$ . A l'ensemble des  $W_j D$  est associé le nuage  $N_J$ , dit des groupes de variables, dans  $R^{I^2}$ .

## 2. Principes communs aux deux méthodes

Le cœur de chacune de ces deux méthodes repose sur une ACP pondérée du tableau complet  $X$ ; les poids affectés aux variables diffèrent d'une méthode à l'autre.

Par rapport à une simple ACP, la prise en compte de la structure en groupes de variables intervient dans :

- la surpondération des variables; en particulier cette surpondération a pour objet d'équilibrer l'influence des groupes dans l'analyse globale de  $X$ ;
- des représentations spécifiques; celles des groupes (projections de  $N_J$ ) et celles des nuages partiels (projections des  $N_I^j$ );
- des aides à l'interprétation spécifiques (e.g. indices de liaison entre groupes de variables).

## 3. Pondération des variables

### 3.1. En AFM

Elle s'interprète clairement dans  $R^I$  et  $R^K$ . Chaque variable du groupe  $j$  est surpondérée par  $1/\lambda_1^j$ , inverse de la première valeur propre de l'ACP séparée de  $X_j$ . Géométriquement, cela revient à rendre égale à 1 l'inertie axiale maximum des nuages  $N_I^j$  et  $N_K^j$  pour chaque groupe  $j$ .

### 3.2. Dans STATIS

Elle s'interprète clairement dans  $R^{I^2}$  muni du produit scalaire «usuel» –  $\langle W_j D, W_l D \rangle = \text{tr}(W_j D W_l D)$  – et s'effectue en deux temps.

a) Dans les applications courantes, cadre adopté ici, les  $W_j D$  sont normés; cela revient à surpondérer chaque variable du groupe  $j$  par le coefficient  $\beta_j$  tel que :

$$\beta_j = \frac{1}{\|W_j D\|} = \frac{1}{\sqrt{\sum_s (\lambda_s^j)^2}}$$

b) Dans  $R^{I^2}$ , on réalise l'analyse factorielle de  $N_J$ , dont le premier axe fournit la combinaison linéaire des  $W_j D$  (notée  $WD = \sum \alpha_j W_j D$  et appelée compromis) en un certain sens (cf. §6.2) la plus liée à l'ensemble des  $W_j D$ .

En outre,  $WD$  étant un vecteur unitaire associé au premier axe factoriel des  $W_j D$ , on a :

$$\alpha_j = \frac{\langle W_j D, WD \rangle}{\nu_1} \text{ (avec } \nu_1 \text{ première valeur propre de l'analyse factorielle de } N_J)$$

c) Finalement, STATIS analyse le compromis  $WD$ , ce qui revient à réaliser l'ACP de  $X$  en surpondérant les variables du groupe  $j$  par  $\alpha_j \beta_j$ .

### 3.3. Comparaison

La surpondération  $\beta_j$  de STATIS, même si elle est techniquement différente de celle de l'AFM, est analogue dans son objet : «normaliser» les groupes pour équilibrer leur influence au sein d'une analyse globale.

La surpondération  $\alpha_j$  de STATIS est, dans son objet, spécifique de cette méthode. On accorde ainsi au groupe  $j$  d'autant plus d'importance *a priori* qu'il est lié au compromis, c'est-à-dire, en un certain sens, à l'ensemble des groupes. Remarquons que, concrètement, les poids sont affectés aux variables : toutes les variables d'un groupe (globalement) lié au compromis voient leur poids initial amplifié, même celles qui ne participent pas à cette liaison.

### 3.4. Illustration

Deux jeux de données sont utiles puisque la surpondération s'effectue en 2 temps dans STATIS.

#### 3.4.1. Normalisation

La normalisation dans STATIS (coefficient  $\beta_j$ ) exerce une influence intermédiaire entre la surpondération de l'AFM et celle qui consiste à égaliser l'inertie totale des nuages  $N_K^j$  (ou des  $N_I^j$ ). Le tableau 1 illustre sur un exemple très simple le fait que, par rapport à ceux de l'AFM, les premiers axes de STATIS sont *a priori* plus influencés par les groupes unidimensionnels.

TABLEAU 1  
Exemple d'influence sur les valeurs propres (notées  $\lambda_i$ ) de chaque groupe,  
de trois types de normalisation des données;  
les quantités normalisées apparaissent en gras.

|                      |          | $\lambda_1$  | $\lambda_2$  | $\sum \lambda_i$ | $\sum (\lambda_i)^2$ | $\sqrt{\sum (\lambda_i)^2}$ |
|----------------------|----------|--------------|--------------|------------------|----------------------|-----------------------------|
| données<br>initiales | groupe 1 | 2            | 0            | 2                | 4                    | 2                           |
|                      | groupe 2 | 1            | 1            | 2                | 2                    | $\sqrt{2}$                  |
| AFM                  | groupe 1 | 1            | 0            | 1                | 1                    | 1                           |
|                      | groupe 2 | 1            | 1            | 2                | 2                    | $\sqrt{2}$                  |
| STATIS               | groupe 1 | 1            | 0            | 1                | 1                    | 1                           |
|                      | groupe 2 | $1/\sqrt{2}$ | $1/\sqrt{2}$ | $\sqrt{2}$       | <b>1</b>             | 1                           |
| iner. tot.<br>égales | groupe 1 | 1            | 0            | <b>1</b>         | 1                    | 1                           |
|                      | groupe 2 | ,5           | ,5           | <b>1</b>         | ,5                   | $\sqrt{2}$                  |

- Données initiales : les deux groupes comprennent chacun 2 variables centrées réduites de poids 1. Le groupe 1 est unidimensionnel, le groupe 2 aussi multidimensionnel que possible (pour un groupe comportant 2 variables).
- AFM : aucun groupe ne peut induire à lui seul un premier axe mais le groupe 2 contribue deux fois plus que l'autre à l'ensemble de l'analyse.
- Egalisation des inerties totales des groupes : le groupe 1 infléchit de façon prépondérante le premier axe.
- STATIS : la normalisation a un effet intermédiaire entre les deux situations précédentes.

### 3.4.2. Surpondération $\alpha_j$ de STATIS

#### a) Données choisies

Nous utilisons un jeu de données un peu caricatural, comportant 4 groupes construits à partir de 6 variables centrées, réduites et deux à deux non corrélées. Chaque variable est présente, avec un poids donné, dans un, deux ou trois groupes. Le tableau 2 rassemble les poids initiaux des différentes variables dans chaque groupe.

A l'intérieur d'un groupe, les variables initiales étant orthogonales deux à deux, chaque composante principale du groupe est confondue avec une variable initiale, la valeur propre associée étant égale au poids de la variable.

Pour chaque groupe, le plus grand poids affecté à une variable (ici égal à la première valeur propre) vaut 1 : la surpondération de l'AFM ne joue pas. En outre, d'un groupe à l'autre, la somme des carrés des poids des variables (ici égale à la somme des carrés des valeurs propres) est pratiquement constante : les  $W_j D$  ont presque la même norme. Au niveau des poids des variables, les deux méthodes ne diffèrent donc que par l'impact du coefficient  $\alpha_j$  de STATIS. Ces coefficients et les poids de variables qui en résultent se trouvent dans le tableau 3.

TABLEAU 2

Données choisies : poids des variables dans les données initiales.

| groupe   | V1 | V2     | V3     | V4     | V5     | V6     | $\Sigma$ | $\Sigma()^2$ |
|----------|----|--------|--------|--------|--------|--------|----------|--------------|
| 1        | 1  | 0,9409 |        | 0,3025 |        |        | 2,2434   | 1,9768       |
| 2        | 1  | 0,9409 |        | 0,3025 |        |        | 2,2434   | 1,9768       |
| 3        | 1  |        | 0,4489 |        | 0,8836 |        | 2,3325   | 1,9823       |
| 4        |    | 1      | 0,4489 |        |        | 0,8836 | 2,3325   | 1,9823       |
| $\Sigma$ | 3  | 2,8818 | 0,8978 | 0,6050 | 0,8836 | 0,8836 | 9,1518   | 7,9182       |

Exemples : le groupe 3 possède les variables 1, 3 et 5 (affectées respectivement des poids 1; 0,4489 et 0,8836); la variable V4 est présente dans les groupes 1 et 2 avec, dans les deux cas, le poids 0,3025.  
**Remarque** : les groupes 1 et 2 sont identiques.

La somme des poids de la colonne  $k$  donne l'inertie de la direction de la variable  $k$  dans le nuage  $N_K$ ; ici, les variables initiales étant orthogonales et de poids maximum égal à 1, ces poids coïncident avec les valeurs propres de l'AFM.

TABLEAU 3

Données choisies : poids des variables dans le compromis de STATIS.

| groupe   | V1     | V2     | V3     | V4     | V5     | V6     | $\Sigma$ | $\alpha_j$ |
|----------|--------|--------|--------|--------|--------|--------|----------|------------|
| 1        | 0,2610 | 0,2456 |        | 0,0790 |        |        | 0,5856   | 0,3670     |
| 2        | 0,2610 | 0,2456 |        | 0,0790 |        |        | 0,5856   | 0,3670     |
| 3        | 0,1718 |        | 0,0730 |        | 0,1518 |        | 0,4006   | 0,2418     |
| 4        |        | 0,1627 | 0,0730 |        |        | 0,1438 | 0,3796   | 0,2291     |
| $\Sigma$ | 0,6938 | 0,6539 | 0,1501 | 0,1580 | 0,1518 | 0,1438 | 1,9514   | 1,2049     |

Les poids découlent du tableau 2 et des  $\alpha_j$  (poids du groupe  $j$  dans STATIS) donnés par le programme.  
 Exemple de calcul (groupe 1, V2) :  $,2456 = ,9409 \times ,3670 / \sqrt{1,9768}$

La somme des poids de la ligne  $j$  donne l'inertie totale du groupe  $j$ . La somme des poids de la colonne  $k$  donne l'inertie de la direction de la variable  $k$  dans le nuage  $N_K$ ; les variables initiales étant orthogonales, ces poids coïncident avec les valeurs propres de l'analyse du compromis.

### b) Interprétation des coefficients $\alpha_j$

Du fait de la simplicité de l'exemple, il est aisé de relier les coefficients  $\alpha_j$  aux données.

- $\alpha_1$  et  $\alpha_2$  sont sensiblement supérieurs à  $\alpha_3$  et  $\alpha_4$  : les groupes 1 et 2 sont chacun plus liés que les deux autres groupes à la structure commune matérialisée par le compromis (étant identiques, ils pèsent fortement sur le compromis); par rapport à la situation initiale et aux deux autres groupes, toutes leurs variables sont «revalorisées» (comparer, par exemple, les poids des variables V1 et V2 dans les 4 groupes pour les deux méthodes).

- $\alpha_3$  est légèrement supérieur à  $\alpha_4$  : le groupe 3 est légèrement plus lié aux autres que ne l'est le groupe 4; par rapport à situation initiale et aux variables du

groupe 4, celles du groupe 3 sont «revalorisées» (comparer, par exemple, le poids de V1 dans le groupe 3 et celui de V2 dans le groupe 4).

*c) Conséquences pour la représentation du nuage («moyen» ou «compromis»)  $N_I$*

Les coefficients de corrélation entre les variables initiales valant 0 ou 1, les facteurs des deux analyses sont identiques; seules diffèrent les inerties qui leur sont associées (lues sur la dernière ligne, du tableau 2 pour l'AFM et du tableau 3 pour STATIS), et par conséquent éventuellement leur rang. L'essentiel de la comparaison entre les rangs des facteurs tient en trois points.

- Les deux premiers facteurs (liés respectivement à V1 et à V2) sont identiques entre les deux analyses.

- La direction liée à V4 correspond au sixième facteur de l'AFM et au troisième de STATIS. Ceci illustre comment STATIS renforce une direction peu importante de groupes qui par ailleurs sont très liés au compromis.

- Autre illustration de ce phénomène : d'une analyse à l'autre, l'importance relative des axes correspondant aux variables V3 et V5 est inversée (facteurs 3 et 4 de l'AFM; facteurs 5 et 4 dans STATIS); cette inversion est due au plus fort coefficient du groupe 3 (lui-même dû au fait que, dans les groupes 1 et 2, la dimension partagée avec le groupe 4 (V2) a une inertie plus faible que celle qui est partagée avec le groupe 3 (V1)).

#### **4. Représentations superposées des nuages partiels des groupes actifs (trajectoires)**

##### **4.1. En AFM**

On projette les nuages  $N_I^j$  sur les axes principaux de  $N_I$ ; on a la relation :

$$F_s(i) = \frac{1}{J} \sum_j F_s(i^j)$$

qui exprime qu'un individu se  $i$  trouve à l'isobarycentre de ses images partielles  $i^j$ .

##### **4.2. Dans STATIS**

La représentation superposée peut aussi s'interpréter à l'aide d'une projection des  $N_I^j$  sur la suite des axes principaux de  $N_I$ . Mais cette représentation est conçue de façon à obtenir la relation :

$$F_s(i) = \sum_j \alpha_j F_s(i^j)$$

ce qui, d'une certaine manière, permet de visualiser les  $\alpha_j$  sur les graphiques.

Compte tenu de la surpondération par  $\alpha_j$  des variables du groupe  $j$ , on s'attend plutôt au coefficient  $\sqrt{\alpha_j}$  dans la relation ci-dessus (c'est ce que l'on obtiendrait en projetant les  $N_I^j$  sans tenir compte des  $\alpha_j$ , et  $N_I$  en en tenant compte); cette relation peut être interprétée ainsi : les nuages  $N_I^j$  subissent une homothétie de rapport  $1/\sqrt{\alpha_j}$  avant d'être projetés.

### Remarque

En général :  $\sum \alpha_j \neq 1$  ( $\sum \alpha_j \geq 1$  si les groupes sont normés).

### 4.3. Comparaison

En comparaison avec celle de l'AFM, la représentation de STATIS accentue la proximité de  $i$  avec les  $i^j$  des groupes les plus proches de la structure commune, quel que soit l'axe factoriel, y compris ceux, en principe peu nombreux et de rang élevé, auxquels ces groupes ont peu contribué.

Techniquement,  $\sum \alpha_j \neq 1$  est quelquefois gênant, par exemple lorsque les points  $i^j$  sont confondus entre eux et diffèrent de leur compromis  $i$ .

### 4.4. Illustration

Nous reprenons les données choisies du tableau 2 en précisant cette fois les valeurs des variables pour les individus (cf. Tableau 4).

TABLEAU 4  
Données choisies : valeurs des 8 individus (A... H)  
pour les variables de base (V1...V6).

|   | V1 | V2 | V3 | V4 | V5 | V6 |
|---|----|----|----|----|----|----|
| A | 1  | 1  | 1  | 1  | 1  | 1  |
| B | 1  | -1 | -1 | -1 | 1  | 1  |
| C | -1 | -1 | 1  | -1 | 1  | -1 |
| D | -1 | 1  | -1 | 1  | 1  | -1 |
| E | 1  | 1  | 1  | -1 | -1 | -1 |
| F | 1  | -1 | -1 | 1  | -1 | -1 |
| G | -1 | -1 | 1  | 1  | -1 | 1  |
| H | -1 | 1  | -1 | -1 | -1 | 1  |

La figure 1 présente côte à côte le plan (F3, F6) de l'AFM et le plan (F5, F4) de STATIS; ils correspondent aux mêmes variables (V3 et V4) et ne diffèrent que par leur inertie.

Les deux représentations de  $N_I$  sont très voisines; les différences entre leurs inerties, dues aux  $\alpha_j$ , ont déjà été commentées. Nous analysons les représentations



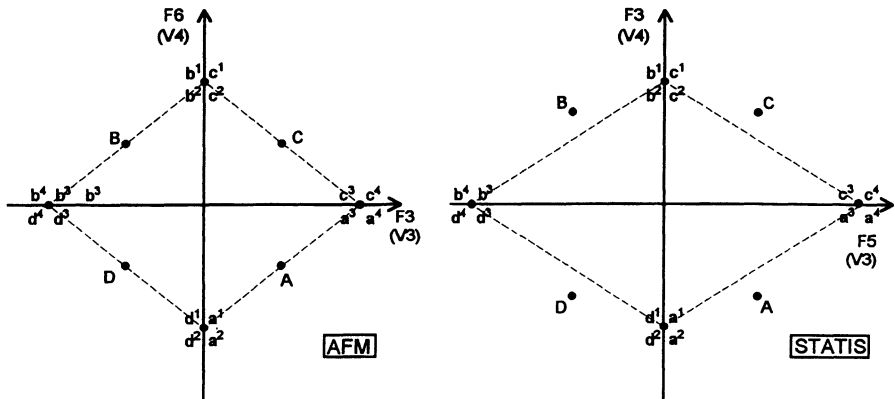


FIGURE 1

*Données choisies : représentations superposées des  $N_I^j$ , sur le plan (F3,F6) de l'AFM et sur le plan (F5,F3) de STATIS.*

Sur ces plans, les individus suivants sont confondus : A et G, B et H, C et E, D et F. Pour alléger les graphiques, seuls les individus A, B, C et D sont figurés.

des  $N_I^j$  en illustrant tour à tour les deux spécificités de STATIS (par rapport à l'AFM) déjà signalées.

#### 4.4.1. Somme des $\alpha_j$ différente de 1 dans STATIS

Sur le graphique de STATIS, l'individu compromis A, par exemple, est à l'extérieur de l'enveloppe convexe (ici un segment) de ses homologues partiels  $a^j$ . Cette caractéristique, qui vaut également pour tous les individus, n'est pas gênante dans cet exemple.

#### 4.4.2. Proximité privilégiée entre un individu compromis et ses homologues partiels de fort poids

En AFM l'individu  $i$  figure à l'isobarycentre de ses homologues partiels  $i^j$ .

Dans STATIS ce point  $i$  figure plutôt du côté des  $i^j$  ayant un fort poids. Or le poids de  $i^j$  ne dépend que de  $j$  et vaut donc pour tous les axes, que ces axes dépendent du groupe  $j$  ou non.

- Le long de l'axe vertical, les points compromis ressemblent plus aux individus partiels des groupes 1 et 2 qu'à ceux des groupes 3 et 4. Or seuls les groupes 1 et 2 contribuent à cet axe. Ceci traduit bien l'esprit de la méthode qui vise à accorder aux groupes une importance d'autant plus grande qu'ils possèdent la structure commune.

- Le long de l'axe horizontal, de nouveau les points compromis ressemblent plus aux individus partiels des groupes 1 et 2. Or ces deux groupes ne contribuent pas à cet axe. On retrouve ici sous un autre angle le fait qu'une dimension commune à certains groupes seulement (ici les groupes 3 et 4) est émoussée dans le compromis si par ailleurs ces groupes participent peu à l'ensemble de la structure commune.

## 5. Mesure de liaison entre groupes de variables

Dans les deux méthodes, la liaison entre les deux groupes de variables  $K_j$  et  $K_l$  est mesurée par le produit scalaire  $\langle W_j D, W_l D \rangle$ , introduit initialement par Escoufier (1973) sous le nom de  $\text{CovV}$ .

### 5.1. Dans STATIS

En pratique, les groupes sont normés; la liaison entre les groupes  $K_j$  et  $K_l$  s'écrit :

$$RV(K_j, K_l) = \left\langle \frac{W_j D}{\|W_j D\|}, \frac{W_l D}{\|W_l D\|} \right\rangle$$

### 5.2. En AFM

Du fait de la pondération de l'AFM, la norme du groupe  $K_j$  vaut :

$$\mathcal{N}_g^2 = \left\| \frac{W_j D}{\lambda_1^j} \right\|^2 = \sum_s \left( \frac{\lambda_s^j}{\lambda_1^j} \right)^2$$

Cette norme est d'autant plus grande que le groupe  $K_j$  possède un grand nombre de facteurs d'importance comparable au premier d'entre eux. Si le groupe ne comporte que des variables réduites non corrélées deux à deux, cette norme est égale au nombre de variables. Elle peut s'interpréter comme un «nombre de dimensions», chacune étant «pondérée» par son inertie; elle constitue de ce fait un indicateur de multidimensionnalité.

La liaison entre les groupes  $K_j$  et  $K_l$  s'écrit :

$$\mathcal{L}_g(K_j, K_l) = \left\langle \frac{W_j D}{\lambda_1^j}, \frac{W_l D}{\lambda_1^l} \right\rangle$$

Cette mesure est d'autant plus grande que  $K_j$  et  $K_l$  sont multidimensionnels et présentent des directions communes nombreuses et proches des directions d'inertie importante de chaque groupe. C'est en quelque sorte un «nombre de dimensions communes», chacune étant «pondérée» par son inertie.

### 5.3. Comparaison

Dans les deux cas, on évalue dans quelle mesure on est en présence d'une structure commune à deux groupes de variables :

- dans le cas de RV, sans se préoccuper de la dimension de cette structure commune;

- dans le cas de  $\mathcal{L}_g$ , en prenant en compte la dimension de cette structure commune et son inertie relativement à celle des groupes.

Les deux mesures sont donc parfaitement complémentaires. Point de repère : on choisira RV ou  $\mathcal{L}_g$  selon que l'on souhaite ou non distinguer le cas de deux structures homothétiques unidimensionnelles de celui de deux structures homothétiques multidimensionnelles (à bien des égards, le second cas est plus remarquable que le premier).

La figure 2 donne les valeurs de RV et de  $\mathcal{L}_g$  dans quelques cas de référence.

## 6. Représentation des groupes de variables (interstructure)

### 6.1. En AFM

$N_J$  est projeté sur les directions de  $R^{I^2}$  induites par les composantes principales de  $N_K$ . En notant  $z_s$  le vecteur unitaire associé au  $s^e$  axe factoriel de  $N_K$ , le  $s^e$  vecteur sur lequel on projette  $N_J$  s'écrit :  $z_s z'_s D$ .

Cette représentation est optimale au sens suivant : on cherche dans  $R^{I^2}$ , sous contrainte qu'ils soient de rang 1 (*i.e.* associés à une seule direction de  $R^I$ ), une suite de vecteurs orthonormés liés le plus possible à l'ensemble des groupes de variables. En notant  $zz'D$  un élément de rang 1 (*i.e.* associé à une direction  $z$  de  $R^I$ ),  $z_s z'_s D$  rend maximum :

$$\mathcal{L}_g(z, K) = \sum_j \langle zz'D, W_j D \rangle$$

### 6.2. Dans STATIS

$N_J$  est projeté sur ses axes principaux d'inertie. On recherche donc  $WD$  qui rend maximum :

$$\sum_j \langle WD, W_j D \rangle^2$$

### 6.3. Comparaison

#### 6.3.1. Critère

Le critère de l'inertie maximum utilisé par STATIS s'impose pour des raisons techniques. Le critère *somme des projections* a une interprétation plus naturelle en terme de liaison et ne pose aucun problème technique dès lors que l'on impose aux éléments de  $R^{I^2}$  cherchés d'être de rang 1. Cela étant, il ne semble pas que l'un ou l'autre critère soit plus adapté à telle ou telle problématique d'application.

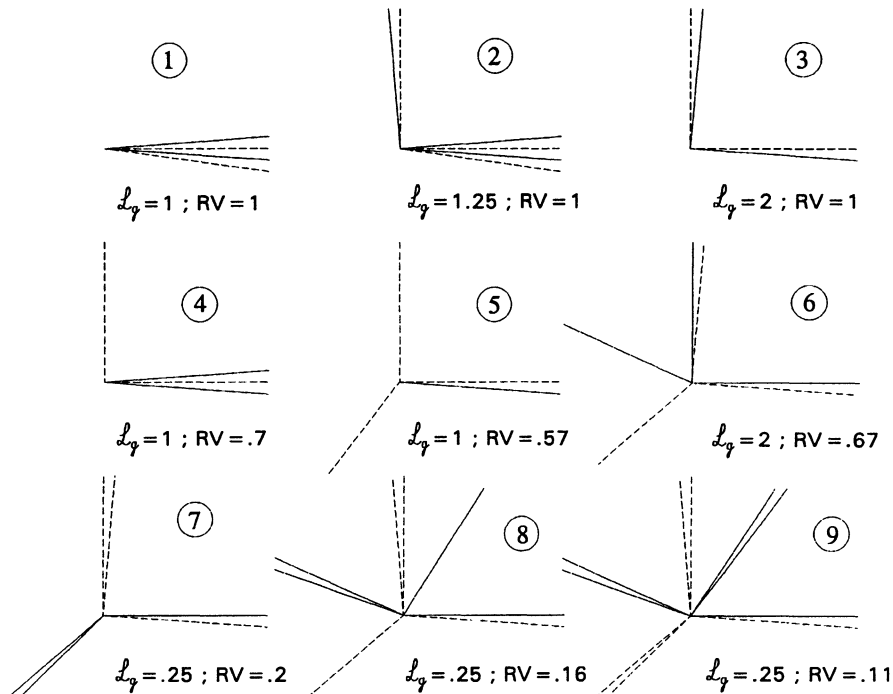


FIGURE 2

Valeur des mesures  $\mathcal{L}_g$  et RV dans quelques cas de références.

Chaque segment représente une variable, appartenant au groupe 1 (traits pleins) ou au groupe 2 (tirets). Dans ces exemples, les angles entre variables valent  $0^\circ$  ou  $90^\circ$ , ce que les dessins ne respectent pas tout à fait afin de figurer toutes les variables.

Cas 1, 2 et 3 : dans ces trois cas, les structures engendrées par les deux groupes sont homothétiques : RV vaut toujours 1 ;  $\mathcal{L}_g$  croît avec le nombre et l'importance (relativement à la première composante principale de chaque groupe) des dimensions communes.

Cas 1, 4 et 5 : dans ces trois cas, les deux groupes ont leur première composante principale, et seulement elle, en commun :  $\mathcal{L}_g$  vaut toujours 1 ; RV décroît au fur et à mesure que le nombre de dimensions non communes croît.

Cas 1 et 6 : entre 1 et 6, le nombre de dimensions communes ainsi que le nombre de dimensions non communes croissent :  $\mathcal{L}_g$  croît et RV décroît.

Cas 7, 8 et 9 : dans ces trois cas, la dimension commune des deux groupes est différente de leur première composante principale et est associée, pour chaque groupe, à une inertie constante relativement à celle de la première composante principale :  $\mathcal{L}_g$  est constant, et vaut ici 0,25 ; RV décroît au fur et à mesure que le nombre et l'importance des dimensions non communes croissent.

### 6.3.2. Représentation de $N_J$

En AFM, on sait interpréter les axes de représentation de  $N_J$ , et donc exprimer axe par axe les proximités entre les groupes en termes de variables. En revanche, la qualité globale de représentation des groupes est souvent mauvaise. En effet la

représentation du groupe  $K_j$  sur le sous-espace engendré par les  $s$  premiers axes s'écrit :

$$\sum_s q_s^j z_s z'_s D \text{ (avec } q_s^j \text{ la coordonnée de } W_j D \text{ sur l'axe de rang } s)$$

expression dans laquelle on reconnaît une écriture du modèle INDSCAL : même si l'on conserve tous les axes, la représentation n'est parfaite que si les données vérifient exactement ce modèle.

Dans STATIS, la qualité globale de représentation de  $N_J$  (mesurée par l'inertie projetée) est optimale, et en pratique bien meilleure qu'en AFM; le premier axe, combinaison linéaire à coefficients positifs des  $W_j D$ , est interprétable en tant que compromis. En revanche les axes suivants ne sont pas interprétables, en ce sens que l'on ne sait pas traduire en termes de données initiales la proximité entre les groupes le long de tel ou tel axe.

#### 6.4. Illustration

##### 6.4.1. Les deux représentations de $N_J$ issues des données choisies (figure 3)

Le nuage  $N_J$  comporte 4 points dont 2 (groupes 1 et 2) sont confondus. STATIS, qui réalise une analyse factorielle non centrée de ce nuage, en fournit une représentation parfaite à l'aide de 3 axes. Sur ces mêmes données, la représentation parfaite par l'AFM nécessite tous les 6 axes (attention ! en général, même en utilisant tous les axes de l'AFM, on n'obtient pas une représentation parfaite de  $N_J$ ). Les trois premiers axes de l'AFM représentent 78% de l'inertie totale. Ce résultat était attendu : par rapport à STATIS, la contrainte à laquelle les axes de l'AFM sont soumis diminue fortement la qualité de représentation de  $N_J$  (au sens usuel de l'inertie projetée rapportée à l'inertie totale).

**Remarque :** dans STATIS la qualité de représentation décroît nécessairement avec le rang de l'axe puisqu'elle correspond à la quantité maximisée; ce n'est pas le cas en AFM.

##### 6.4.2. Interprétation de la représentation fournie par STATIS

La représentation fournie par STATIS montre essentiellement que les groupes 1 et 2 sont identiques et plus liés à chacun des groupes 3 et 4 que ces derniers ne le sont entre eux. Ceci est en accord avec l'examen direct des données (cf. tableau 2) et avec celui des coefficients RV (cf. Tableau 5).

Nous sommes en présence d'une visualisation globale des coefficients RV, d'autant plus précieuse que le nombre de groupes est grand.

Le premier axe montre que les groupes 1 et 2 sont très voisins du compromis ( $RV(W_1 D, W D) = .965$ ), et que les groupes 3 et 4 sont également liés à ce compromis ( $RV(W_3 D, W D) = .636$ ;  $RV(W_4 D, W D) = .603$ ). Cet axe 1 correspond à une structure multidimensionnelle : le fait que toutes les coordonnées des  $W_j D$  le long de cet axe soient positives (ce qui est toujours le cas) n'implique pas l'existence d'une dimension commune à tous les groupes (ce qu'illustre cet exemple).

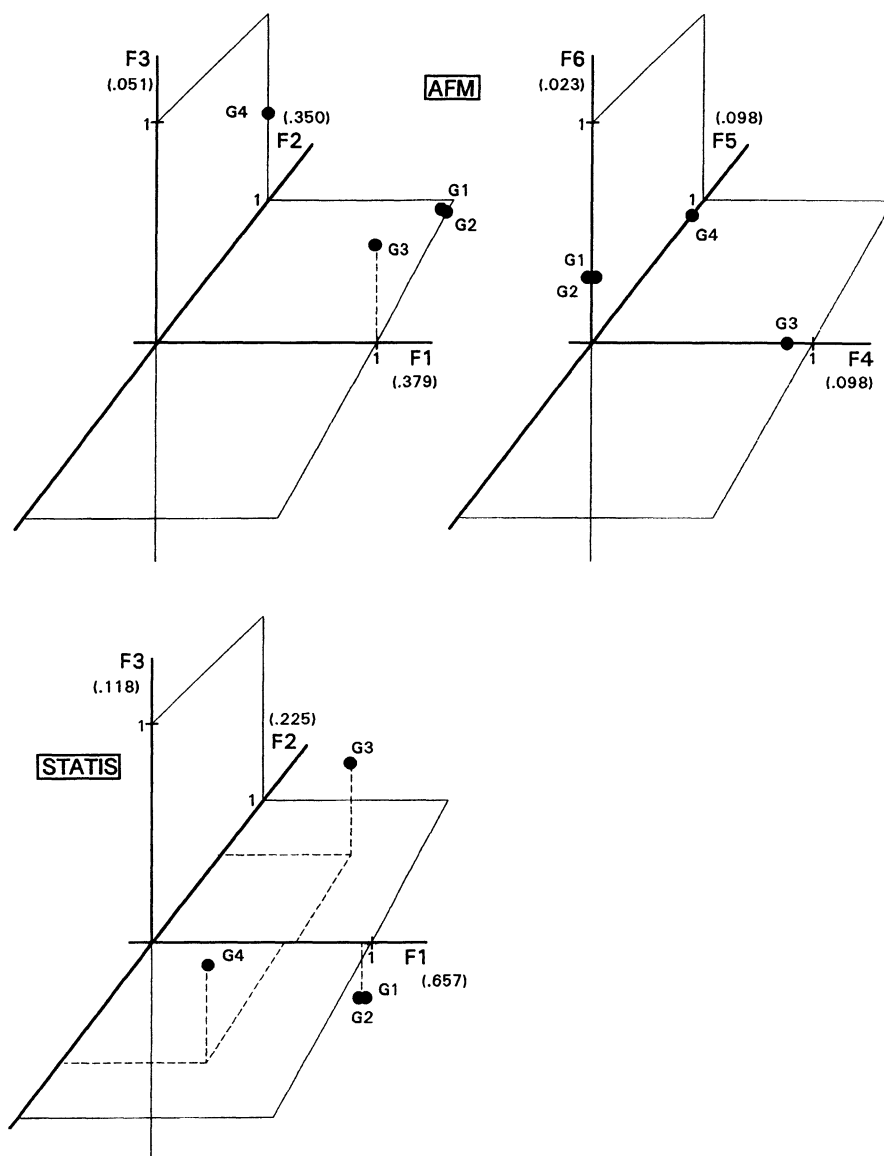


FIGURE 3  
*Jeu des données choisies*

Représentation du nuage des groupes de variables par l'AFM et par STATIS.

Entre parenthèses : pourcentages d'inertie

TABLEAU 5  
Données choisies : coefficients RV.

| Groupe | 1     | 2     | 3     | 4     |
|--------|-------|-------|-------|-------|
| 1      | 1,000 |       |       |       |
| 2      | 1,000 | 1,000 |       |       |
| 3      | 0,505 | 0,505 | 1,000 |       |
| 4      | 0,475 | 0,475 | 0,102 | 1,000 |

(Remarque : dans les données initiales, les normes des  $W_j D$  sont constantes; les mesures  $\mathcal{L}_g$  sont proportionnelles aux RV)

On ne peut rien déduire, en termes de données, des autres axes.

#### 6.4.3. Interprétation de la représentation fournie par l'AFM

##### Remarque préliminaire

Dans ces données, l'interprétation est particulièrement simple puisque chaque axe  $s$  correspond à une variable initiale  $\nu$  : la coordonnée non nulle du groupe  $j$  sur l'axe  $s$  implique que cette variable  $\nu$  est présente dans le groupe  $j$ .

Les représentations fournies par l'AFM conduisent aux mêmes conclusions que celle de STATIS en ce qui concerne les positions relatives globales des points. Rappelons que sur ce point les données sont favorables à l'AFM qui, en général, ne fournit pas une image exacte de  $N_J$ . En revanche, ce qui est général, c'est l'interprétabilité axe par axe de cette représentation, due au fait qu'un axe de cette représentation correspond à une seule dimension des nuages  $N_I$  et  $N_K$ . Ainsi, il apparaît que :

- les groupes 1, 2 et 3 ont en commun leur principale direction d'inertie (axe 1);
- les groupes 1 et 2 ont chacun une direction d'inertie importante en commun avec la principale direction d'inertie du groupe 4 (axe 2);
- les groupes 3 et 4 ont en commun une direction d'importance moyenne (axe 3) et présentent chacun une direction spécifique d'inertie importante (axes 4 et 5);
- les groupes 1 et 2 ont aussi en commun une direction de faible inertie (axe 6).

### Conclusion

Au-delà de considérations techniques, deux points de repère peuvent guider l'utilisateur dans le choix entre les deux méthodes pour traiter un problème particulier.

### *Extraction d'une structure commune*

On utilisera l'AFM lorsque l'on souhaite extraire *une par une* les dimensions communes à plusieurs groupes. Cette problématique est celle de l'analyse multicanonique; l'AFM peut d'ailleurs être vue comme une analyse multicanonique particulière (chaque dimension extraite est en un certain sens – cf. §6.1 – la plus liée à l'ensemble des groupes) qui bénéficie pleinement des relations de dualité de l'analyse factorielle.

On utilisera STATIS lorsque l'on souhaite *extraire d'un coup* la structure commune multidimensionnelle, et *analyser ensuite* cette structure.

Les principales structures communes aux groupes de variables sont nécessairement mises en évidence quel que soit le point de vue adopté. La différence entre les deux stratégies porte sur des structures marginales, par exemple une liaison entre deux groupes qui par ailleurs ne ressemblent pas aux autres. La première stratégie fait ressortir une telle structure en l'interclassant parmi les autres en fonction de son importance. La seconde stratégie minimise son importance du fait qu'elle ne participe pas à la structure commune d'ensemble.

En pratique, les données réelles, sur lesquelles nous avons confronté les deux méthodes, ont conduit à des configurations d'individus très voisines. Il semble que les deux stratégies aboutissent à des résultats très proches dès lors que les données possèdent une structure un tant soit peu forte.

### *Représentation des groupes*

On utilisera STATIS si l'on attend de cette représentation uniquement une *visualisation globale* des ressemblances entre groupes.

On utilisera l'AFM si l'on attend une *visualisation analytique*, même partielle, des ressemblances entre groupes.

En pratique, on observe toujours des représentations de groupes très différentes entre les deux méthodes.

### **Bibliographie**

- ESCOFIER B., PAGES J. (1993). *Analyses factorielles simples et multiples; objectifs, méthodes et interprétation*. Dunod. Paris.
- ESCOUFIER Y. (1973). Le traitement des variables vectorielles. *Biometrics*, 29, 751-760.
- LAVIT C. (1988). *Analyse conjointe de tableaux quantitatifs*. Masson. Paris.
- LAVIT C. (1988). ACT : *Analyse conjointe de tableaux quantitatifs, méthode STATIS*. Logiciel diffusé par CISIA, 1 av. Herbillon 94160 Saint-Mandé.
- PAGES J. (1993). *AFMULT : Analyse factorielle multiple* : logiciel diffusé par ADDAD, 22 rue Charcot 75013 Paris.
- PAGES J. (1994). *Analyse factorielle multiple*. Document d'habilitation. Université de Rennes 2.