

# APPLICATION DE L'ANALYSE FACTORIELLE MULTIPLE POUR LE TRAITEMENT DE CARACTÈRES EN ÉCHELLE DANS LES ENQUÊTES

Sergio Camiz et Jérôme Pagès<sup>1</sup>

## RÉSUMÉ

On rencontre fréquemment dans les questionnaires des batteries de questions sollicitant chacune une réponse sous forme d'une note mesurant un degré d'accord, un degré de satisfaction, etc. L'analyse descriptive exploratoire d'un tel ensemble de questions se fait couramment soit par une analyse en composantes principales (une note étant considérée comme une valeur d'un caractère quantitatif) soit par une analyse des correspondances multiples (une note étant considérée comme une modalité d'un caractère qualitatif). À l'aide d'un exemple, nous comparons ces deux démarches via une analyse factorielle multiple, prenant en compte les mêmes caractères simultanément en tant que quantitatifs et qualitatifs. On met alors en évidence les structures communes aux deux points de vue ainsi que les structures spécifiques de chacun. L'*AFM* apparaît alors comme un outil commode de comparaison de méthodologies reposant sur une analyse factorielle, mais aussi comme un compromis entre ces deux approches.

## 1. INTRODUCTION

On rencontre dans beaucoup de questionnaires des batteries de caractères « en échelle » (à savoir, quantitatifs discrets ayant un faible nombre de modalités, disons moins d'une dizaine pour fixer les idées). Ce sont des notes que l'on demande aux interviewés afin qu'ils expriment par exemple leur satisfaction sur une série d'aspects d'un produit ou, autre exemple classique, leur degré d'accord avec une série d'items.

Dans le traitement de ce type de données par analyse factorielle, en préalable ou non à une classification automatique, les praticiens sont partagés entre deux méthodologies :

- considérer les caractères comme continus et mettre en œuvre une analyse en composantes principales (*ACP*) ; ce point de vue suppose linéaire l'échelle de notation ce qui est quelquefois critiqué ; il n'envisage que les liaisons linéaires entre les caractères au travers de leur matrice des corrélations, ce qui peut s'avérer une restriction forte ; enfin, la présence de données manquantes pose un problème, généralement résolu en pratique en important la moyenne du caractère concerné, bien que cette pratique relève plus d'un expédient commode que d'une réelle solution. En contrepartie de ces inconvénients assez lourds, cette méthodologie est très synthétique en ce sens qu'elle fournit des graphiques peu chargés ; un caractère n'étant représenté que par un seul point.

- considérer les caractères comme qualitatifs et mettre en œuvre une analyse des correspondances multiples (*ACM*) ; ce point de vue ne suppose aucune hypothèse sur les modalités d'un même caractère, pas même celle de l'ordre des modalités, ce qui est évidemment une lacune, mais ne s'avère pas trop grave en pratique, cet ordre étant généralement « retrouvé » par la méthode, du fait de la structure des données ; le point fort de cette analyse est la possibilité de mettre en évidence des liaisons non linéaires ; elle gère avec élégance le cas des données manquantes à l'aide d'une modalité ad hoc ; mais, en contrepartie, la lecture des graphiques est complexe, chaque caractère étant représenté par autant de points que de modalités ; en particulier, dans le cas de caractères en échelle, on observe souvent sur les premiers axes un effet Guttman, qui rend plus difficile l'interprétation des axes suivants (Camiz, 2004, sous presse).

Nous présentons dans ce travail une comparaison de ces deux méthodologies sur un exemple. La confrontation directe des résultats des deux traitements séparés est toujours fastidieuse et quelquefois insuffisante, notamment

---

<sup>1</sup> Sergio Camiz, Università di Roma La Sapienza, Dipartimento di Matematica Guido Castelnuovo, Piazzale Aldo Moro 2, Roma, Italy, I 00185 (sergio.camiz@uniroma1.it)  
Jérôme Pagès, AGROCAMPUS, Laboratoire de Mathématiques Appliquées, 65 rue de Saint-Brieuc CS 84215, Rennes cedex France, F 35042 (pages@agrocampus-rennes.fr)

lorsque les structures communes aux deux analyses ne s'expriment pas sur des axes de même rang. Aussi, utilisons-nous ici une analyse factorielle multiple (*AFM*; Escofier & Pagès 1998), méthode conçue pour la recherche de structures communes à plusieurs tableaux de caractères. Cette méthode peut prendre en compte à la fois des caractères quantitatifs et qualitatifs (Pagès, 2002). Elle s'applique donc bien à notre problème dans lequel un même ensemble de caractères doit être vu d'une part comme quantitatif et d'autre part comme qualitatif. Notre objectif est donc d'illustrer comment l'*AFM* constitue un outil commode (idéal ?) pour comparer deux méthodologies : c'est-à-dire de fournir une hiérarchie des structures communes et spécifiques.

Du point de vue de l'utilisateur, on verra que l'utilisation de l'*AFM* sur le même tableau considéré une fois comme qualitatif et une fois comme quantitatif, fournit un aperçu de la structure des données beaucoup plus riche et imagé que par les deux analyses appliquées séparément. Ceci s'avère d'autant plus intéressant car ce sont soit les facteurs communs soit ceux propres d'une structure seule, mais alors très importants, qui apparaissent en premier et ceci peut aider l'utilisateur, qui ira rarement examiner au delà des premiers deux ou trois plans factoriels.

## 2. LES DONNÉES D'EXEMPLE ET LA MÉTHODOLOGIE

Il s'agit des questionnaires d'une enquête portée en 1992 sur les étudiants de la faculté d'Economie de l'Université de Brescia (Italie du Nord), concernant leur situation, soit personnelle soit universitaire, et d'une part leur jugements sur l'université et les services qu'elle fournissait et de l'autre leur satisfaction sur des aspects de leur situation personnelle. Le questionnaire était composé de 125 questions fermées et 16 ouvertes, groupés en cinq sections ; il avait été distribué aux étudiants lors de leur inscription ; 3416 questionnaires ont été retournés, soit 76.6% des étudiants inscrits. Les données ont été étudiées soit par des statistiques descriptives (Camiz et al., 1994), soit par des analyses multidimensionnelles, afin de caractériser les groupes homogènes d'interviewés par rapport à leurs réponses (Bottiroli Civardi et Camiz, 1997), soit par des analyses textuelles, pour avoir un meilleur aperçu de leur jugements vis-à-vis des études et de leurs projets de vie (Camiz et Tagliacozzo, 1997).

Pour cette étude, on a sélectionné deux groupes de questions : un groupe signalétique concernant le sexe des répondants (2 modalités), s'il travaillent ou pas (3 modalités) et la moyenne des notes reçues (6 modalités) ; l'autre groupe est formé de dix échelles, dont six concernent un jugement, de 1 (très mauvais) à 9 (excellent), sur les espaces fournis par l'université pour les cours et pour l'étude, les horaires, les aides (telles que la bibliothèque, le laboratoire de calcul, etc.), l'administration en général et leurs rapports avec les enseignants, et quatre concernent leur satisfaction vis-à-vis de leur situations économique et familiale, des rapports amicaux et de l'environnement universitaire. Dans notre étude on s'est borné aux étudiants ayant été inscrits depuis un an au mois, car autrement il n'auraient pas pu donner un jugement sur les structures universitaires. Il s'agit finalement d'un tableau de 2403 étudiants et 10 caractères en échelle de 1 à 9 plus 3 caractères qualitatifs signalétiques.

Le tableau initial des caractères en échelle est dupliqué pour constituer un tableau  $2403 \times 20$ . À l'intersection de la ligne  $i$  et de la colonne  $k$ , on trouve  $x_{ik}$ , valeur attribuée par l'individu  $i$  à la question  $k$ . Cette valeur est un numéro de modalité dans le premier sous-tableau (point de vue qualitatif) et une valeur numérique dans le second (point de vue quantitatif). A ces deux sous-tableaux, on ajoute trois caractères du signalétique (qualitatif) (Tableau 1).

	10 questions de satisfaction		3 signalétiques			
	1	10	1	10	1	3
$2403 = I$	Approche ACM : valeur = modalité		Approche ACP : valeur = nombre		Approche ACM : caractères qualitatifs illustratifs	

Tableau 1 - Structure du tableau des données soumis à l'*AFM*.

Les deux premiers sous-tableaux sont introduits en actif. En raccourci, l'*AFM* « se comporte comme une *ACM* » pour le premier et « comme une *ACP* » pour le second. Aussi, pour simplifier, nous parlerons de « groupe *ACM* » et de groupe *ACP* ». Les grandes différences d'inertie (la première valeur propre est toujours inférieure à 1 en *ACM* et supérieure à 1 en *ACP*) sont neutralisées par la pondération de l'*AFM*. Enfin, le signalétique constitue un groupe illustratif. Toutes les analyses ont été effectuées avec le logiciel *SPAD* (Spad, 2003).

## 3. LES RÉSULTATS

### 3.1 Les analyses séparées

Dans l'*ACP* des données quantitatives on trouve deux axes sensiblement plus grands que 1 : on est porté à se borner à considérer ces deux-là, pour un total d'inertie expliquée de 49.75% ; sur ce plan, non représenté mais très

proche

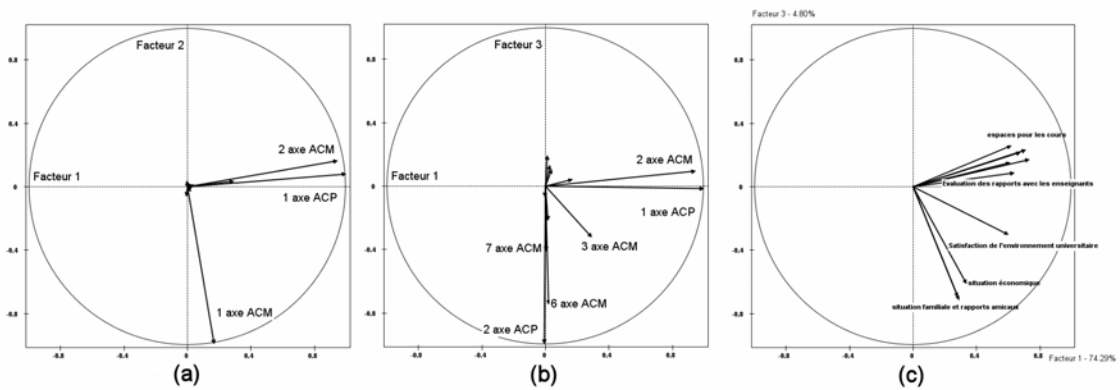


Figure 1 – les axes partiels de l'ACP et de l'ACM sur les plans factoriels de l'AFM : (a) plan 1-2, (b) plan 1-3, (c) les caractères quantitatifs sur le plan 1-3.

de la Figure 1c, on remarque l'effet de taille sur le premier axe et une opposition entre les deux groupes de réponses sur le deuxième, avec la question sur l'environnement universitaire en position plus centrale. La distribution des interviewés sur ce plan (non représenté) apparaît assez uniforme, sans suggérer de véritables groupes.

Dans l'ACM on se trouve avec deux axes très importants, résumant 11,19% de l'inertie totale, les autres descendant lentement. Il faut préciser que, après avoir réévalué l'inertie des axes suivant Benzécri (1979), ce taux saute à 82,07% et 89,02% si on y ajoute le troisième axe : inutile de prendre en compte plus de deux ou trois axes selon les critères courants...

Or, que se passe-t-il dans les plans de l'ACM? L'axe 1 est très clairement dû aux non-réponses, les non-répondants étant distribués sur nombre de strates parallèles, suivant le nombre de non-réponses fournies (Figure 2a), l'axe 2 montrant les séquences des différentes notes avec des aller-retour pas clairs du tout. Effectivement, c'est dans le plan engendré par les axes 2 et 3 qu'un aperçu plus intéressant se montre (Figure 2b): un bel effet Guttman où toutes les trajectoires des items se ressemblent sans aucune distinction apparente... La distribution des répondants est encore une fois fort uniforme sur ce plan ; néanmoins, c'est avec de l'imagination qu'on entrevoit une forme un peu parabolique du nuage.

### 3.2 Les résultats de l'AFM

La matrice des corrélations entre les facteurs des analyses séparées montre deux structures communes : 1) la proximité du deuxième facteur de groupe ACM au premier facteur du groupe ACP ; 2) les facteurs 6 et 7 de l'ACM légèrement corrélés au deuxième facteur de l'ACP. À part cela, les autres facteurs ne sont pas corrélés entre eux.

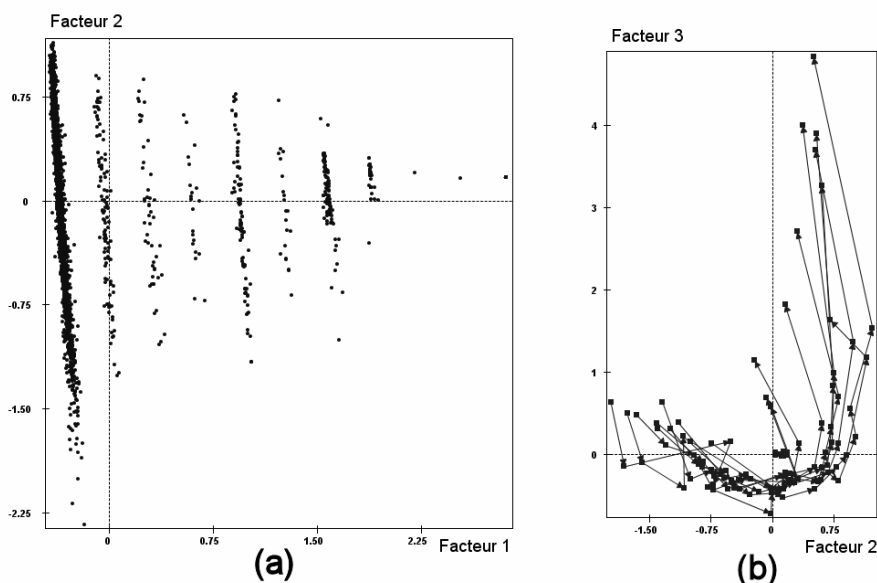


Figure 2 – Les plans factoriels de l'ACM : (a) la distribution des répondeurs sur le plan 1-2, (b) l'effet Guttman des trajectoires des caractères en échelle sur le plan 2-3.

	axe 1	axe 2	axe 3	axe 4	axe 5	axe 6	axe 7	axe 8	axe 9	axe 10
Ensemble en %	9,107	6,002	4,758	2,612	2,366	2,232	2,083	2,024	1,879	1,793
Ensemble	1,508	0,994	0,788	0,433	0,392	0,370	0,345	0,335	0,311	0,297
Groupe 1 (ACM)	0,516	0,985	0,280	0,340	0,218	0,186	0,221	0,185	0,170	0,157
Groupe 2 (ACP)	0,992	0,009	0,508	0,093	0,174	0,184	0,124	0,150	0,141	0,140

Tableau 2 - Inertie des axes de l'AFM, globale et ventilée par groupe

Les valeurs propres de l'AFM mettent en évidence trois facteurs bien marqués. Leur décomposition par groupe (Tableau 2) montre que le premier et le troisième facteur correspondent à des directions d'inertie importante des deux groupes (Figure 1b), le deuxième étant spécifique du groupe ACM (Figure 1a). Ce second facteur isole les modalités « non-réponses » (et donc les non-répondants), une information manquante en ACP, où les données manquantes sont remplacées par la moyenne. Cependant, dans le plan 1-2 on retrouve les strates de Figure 2a, mais cette fois en vertical. C'est donc sur le plan factoriel 1-3, que notre intérêt va se concentrer, le troisième facteur correspondant au second axe de l'ACP et en partie aux deux facteurs 6 et 7 de l'ACM (Figure 1b): en effet, il s'agit d'informations qui n'auraient pas été prises en compte par un chercheur qui aurait préféré utiliser l'ACM. Sur ce plan on retrouve, du point de vue quantitatif (Figure 3a), les caractères d'évaluation des structures universitaires fortement corrélés au premier facteur, alors que ceux de satisfaction personnelle sont beaucoup plus proches du troisième, seule la satisfaction de l'environnement universitaire se trouvant entre les deux groupes. Son interprétation est donc facile : les évaluations personnelles semblent peu liées aux notes affichées aux structures universitaires, excepté l'environnement universitaire qui synthétise les deux aspects interagissant entre eux. Si maintenant on projette les centres de gravité des modalités des mêmes caractères ainsi que leurs trajectoires (Figure 3a), on voit bien distingués les deux ensembles : les évaluations ayant une allure horizontale assez proche du premier facteur et les satisfactions plus diagonales, descendant du second quadrant vers le quatrième ; une fois encore l'environnement universitaire suit une trajectoire intermédiaire entre les deux groupes. Sur ce plan, les interviewés se distribuent de façon quasiment uniforme, sinon dans le second quadrant où la densité semble nettement plus faible.

En projetant sur ce plan les centres de gravité des caractères qualitatifs constituant le groupe illustratif (Figure 3b), on remarque une opposition significative entre hommes et femmes et une entre travailleurs (soit réguliers soit occasionnels) et non travailleurs : en effet, les hommes semblent fort plus critiques vis-à-vis de leur condition personnelle que les femmes, notamment en ce qui concerne l'environnement universitaire ; de même, les travailleurs sont fort plus critiques que les non-travailleurs : il faut souligner que les travailleurs occasionnels sont plus des hommes que des femmes. Quant aux notes, on remarque un effet étrange, car si ceux avec de bonnes notes (28-30) se trouvent du côté des répondants généralement satisfaits, ceux avec les notes les plus faibles (18-21) sont encore dans le premier quadrant, grâce surtout aux bons rapports qu'ils déclarent avoir avec les enseignants.

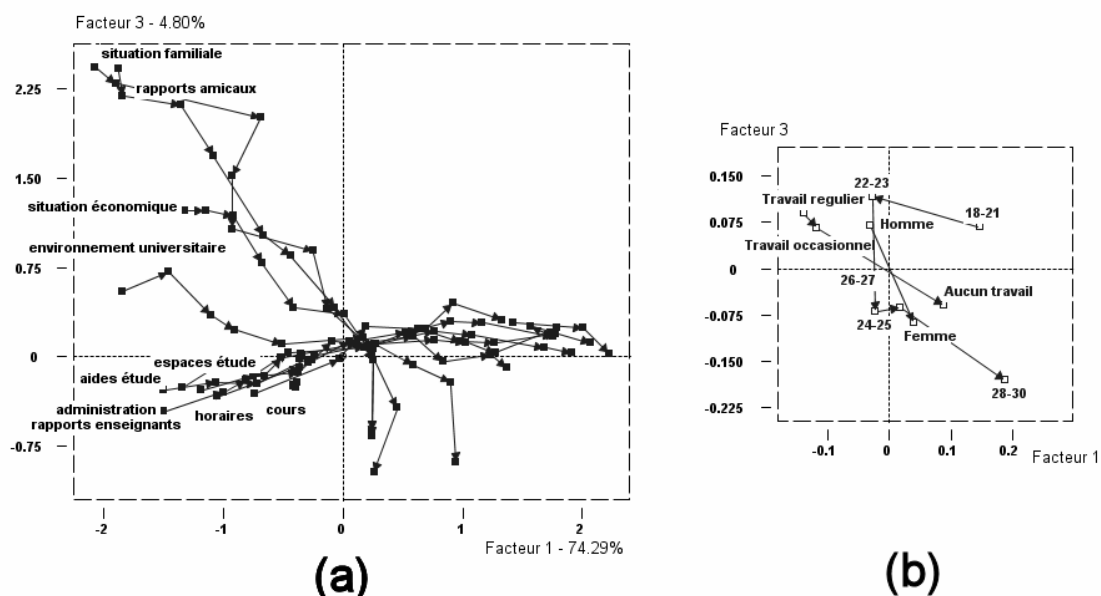


Figure 3 - Les trajectoires des caractères qualitatifs sur le plan 1-3 de l'AFM : (a) les évaluations et satisfactions actives, (b) trois signalétiques illustratives.

## 4. CONCLUSION

Dans cet exemple, l'AFM a mis en évidence les principales structures, communes aux deux codages ou spécifiques de l'un d'eux. Soit :

- un facteur de satisfaction globale, qui est la structure forte des deux tableaux (1<sup>er</sup> axe de l'ACP, 2<sup>e</sup> axe de l'ACM) ;
- un facteur mettant en exergue la satisfaction personnelle, structure forte de l'ACP (2<sup>e</sup> axe) mais qui apparaît à un rang moins important dans l'ACM (axes 6 et 7), analyse dans laquelle il serait sans doute passé inaperçu ;
- un facteur non réponse, comme attendu structure spécifique de l'ACM.

Remarquons que, dans cet exemple où les liaisons entre les échelles sont plutôt linéaires, l'ACP est bien adaptée, et ce malgré l'imputation de la moyenne pour les données manquantes, assez nombreuses (entre 10% et 15% selon les questions).

Cette analyse suggère aussi un autre intérêt de l'AFM : celui de compromis entre les deux points de vue. Les premiers facteurs sont clairement interprétables et ce, quand c'est possible, à partir des caractères codés de chacune des façons. Au delà de la comparaison entre les points de vue, il semble donc intéressant d'utiliser ces plans factoriels pour présenter les résultats de l'enquête. On dispose ainsi d'une méthode à la fois enrichie par la présence simultanée des deux points de vue et relativement « robuste » vis à vis du codage des données en échelle.

## RÉFÉRENCES

- Benzécri, J.P. (1979), « Sur le Calcul des taux d'inertie dans l'analyse d'un questionnaire, Addendum et erratum à [BIN.MULT.] », *Les Cahiers de l'Analyse des Données*, 4, p. 377 -378.
- Bottiroli Civardi, M., et S. Camiz (1997), « Analisi esplorativa e di conferma per lo studio della propensione al lavoro degli studenti d'Economia e Commercio dell'Università di Brescia », dans M. Bottiroli Civardi et S. Camiz (eds.), *La popolazione studentesca e le Università Italiane: indagini, modelli e risultati*, Padova : CLEUP Editrice, p. 73-113.
- Camiz, S. (2004, sous presse), « The Guttman Effect: its Interpretation and a New Redressing Method », *Data Analysis Bulletin* (sous presse), Università di Roma \*La Sapienza+, Dipartimento di Matematica \*Guido Castelnuovo+, preprint n. 5/2004.
- Camiz, S., M. Civardi, R. Cremonesi, P. Falbo, S. Stefani et S. Zoppetti (1994), « Indagine sulla popolazione studentesca dell'Università di Brescia », Rapport de recherche du Dipartimento di Metodi Quantitativi dell'Università di Brescia, Quaderno n. 62.
- Camiz, S. , et G. Tagliacozzo (1997), « Analisi testuale delle risposte a testo libero dei questionari degli studenti di Brescia: primi risultati », dans M. Bottiroli Civardi et S. Camiz (eds.), *La popolazione studentesca e le Università Italiane: indagini, modelli e risultati*. Padova : CLEUP Editrice, p. 211-241.
- Escofier, B., et J. Pagès (1998), *Analyses factorielles simples et multiples*, Paris : Dunod.
- Pagès, J. (2002), « Analyse factorielle multiple appliquée aux variables qualitatives et aux données mixtes », *Revue de Statistique Appliquée*, 50(4), p. 5-37.
- SPAD (2003), version 5.6 , diffusé par DECISIA , 30 rue Victor Hugo, 92532 LEVALLOIS-PERRET cedex France.