

MÓNICA BÉCUE-BERTAUT

JÉRÔME PAGÈS

Analyse simultanée de questions ouvertes et de questions fermées. Méthodologie, exemple

Journal de la société française de statistique, tome 142, n° 4 (2001), p. 91-104

http://www.numdam.org/item?id=JSFS_2001__142_4_91_0

© Société française de statistique, 2001, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). 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/>

ANALYSE SIMULTANÉE DE QUESTIONS OUVERTES ET DE QUESTIONS FERMÉES MÉTHODOLOGIE, EXEMPLE

Mónica BÉCUE-BERTAUT *, Jérôme PAGÈS **

RÉSUMÉ

Usuellement, dans l'analyse de résultats d'enquêtes, les réponses aux questions ouvertes sont traitées par analyse des correspondances binaires et les réponses aux questions fermées sont traitées par analyse des correspondances multiples. On peut souhaiter introduire dans une même analyse ces deux types de données en actif. Se posent alors plusieurs problèmes, en particulier celui de l'équilibre, au sein d'une analyse globale, entre les différents ensembles de données. Nous présentons ici une extension de l'analyse factorielle multiple pour ce type de données, à la fois dans son principe méthodologique et dans une application.

Mots clés : question ouverte, analyse des correspondances, analyse des correspondances multiples, analyse factorielle multiple.

ABSTRACT

Usually, in the analysis of results from surveys, the answers to open-ended questions are treated by correspondence analysis and the answers to the closed questions are treated by multiple correspondence analysis. One may wish to introduce these two types of data as active into a single analysis. This poses several problems, in particular the balance between the various sets of data in a global analysis. Here, an extension of multiple factor analysis is presented through its methodological principle and an application.

Keywords : open-ended questions, correspondence analysis, multiple correspondences analysis, multiple factor analysis.

1. Introduction

Lors d'une enquête par questionnaire, il est très utile d'aborder certains thèmes au moyen d'une batterie de questions fermées complétée par une ou deux questions ouvertes. Le problème se pose alors de traiter globalement cette information pour obtenir une synthèse de l'ensemble des questions.

* EIO. Universitat Politècnica de Catalunya, Barcelona, Spain, Edifici FME. C/Pau Gargallo, 5; 08028 Barcelona. E-mail : monica.becue@upc.es

** ENSA/INSFA, 65, rue de Saint-Brieuc, CS 84215, F-35042 Rennes cedex, France.
E-mail : Jerome.Pages@agrorennes.educagri.fr

Cet article décrit une méthodologie qui permet de traiter simultanément des questions ouvertes et des questions fermées en leur donnant le statut actif. Cette méthodologie est illustrée au moyen d'une enquête espagnole, *Les enfants et la lecture* (paragraphe 2). En premier lieu, on considère les données comme un tableau multiple qui juxtapose des tableaux de contingence, correspondant chacun à une question ouverte, et un tableau disjonctif complet construit à partir des réponses aux questions fermées (paragraphe 3). On montre d'abord les résultats obtenus avec l'approche classique, qui considère ou bien les questions fermées ou bien les questions ouvertes comme actives (paragraphe 4). Ensuite, on rappelle brièvement les principes de l'analyse factorielle multiple (AFM), de l'analyse factorielle multiple pour tableaux de contingence (AFMTC) et de leur extension à un tableau mixte juxtaposant tableaux de contingence et tableaux disjonctifs complets (paragraphe 5). Le paragraphe 6 présente, pour l'enquête précitée, les résultats obtenus au moyen de cette extension de l'AFM.

2. Exemple : les enfants et la lecture

Afin de mieux connaître les pratiques et les attitudes envers la lecture des enfants, une étude a été conduite dans une ville de la banlieue de Barcelone (Rajadell, 1990). 895 enfants de 5^e année d'Enseignement Général de Base (10-11 ans) de cette ville ont répondu à un questionnaire fermé comportant une batterie de questions sur les habitudes de lecture ainsi que les deux phrases suivantes à compléter :

- Q_1 : Para mí, leer es... (Pour moi, lire est...)
- Q_2 : Creo que leer es importante porque... (Je crois que lire est important parce que...)

TABLEAU 1. – Exemples de réponses.

—A101; Q_1 :	<i>fantástico! a mí me gusta leer mucho para luego contar lo que he leído. si cada día leo, cuando me toque leer en clase sabré leer rápido</i>
— Q_2 :	<i>sino luego al final de curso no sabré leer bien</i>
—A122; Q_1 :	<i>una diversión porque así no se pierde el tiempo</i>
— Q_2 :	<i>después te encuentras algo peligroso como perro peligroso en una fabrica y lo ponen en un cartel y es bueno para el estudio y la imaginación</i>
—A130; Q_1 :	<i>es un rollo porque hay muchas letras</i>
— Q_2 :	<i>si no sabes leer no sabes escribir</i>

En outre, on dispose des notes scolaires de chaque enfant et de quelques informations sur la famille comme le niveau d'études des parents.

ANALYSE SIMULTANÉE DE QUESTIONS OUVERTES ET QUESTIONS FERMÉES

Dans ce travail, on cherche, dans un premier temps, à analyser l'attitude et les habitudes des enfants envers la lecture en prenant en considération à la fois les réponses fermées et les réponses ouvertes et, dans un deuxième temps, à mettre en évidence les individus (ou les catégories d'individus) dont le comportement diffère selon la question.

3. Tableau de données et notations

3.1. Tableau de données

Le tableau 2 schématise le tableau multiple construit à partir des données. Aux questions fermées actives (cf. tableau 3), on associe un tableau disjonctif complet. À chaque question ouverte, on associe le tableau de contingence qui contient la fréquence avec laquelle chacun des mots est employé par chacun des individus pour répondre à la question.

TABLEAU 2. – Tableau multiple à analyser.

	Questions fermées	Question ouverte 1 <i>Para mí, leer es...</i>	Question ouverte 2 <i>Creo que leer es importante porque...</i>
Individus ($I = 895$)	Variabes indicatrices	Fréquence des mots	Fréquence des mots
	K_1 colonnes-modalités $K_1 = 23 = 7 \times 3 + 2$	K_2 colonnes-mots $K_2 = 91$	K_3 colonnes-mots $K_3 = 92$

3.2. Notations

Pour le tableau correspondant aux réponses aux questions fermées, de type *individus* \times *variables indicatrices*, x_{ik} indique si le répondant i présente ($x_{ik} = 1$) ou non ($x_{ik} = 0$) la modalité-colonne k . On traite ici l'ensemble des questions fermées simultanément mais, plus généralement, les variables qualitatives peuvent être structurées selon J_q groupes (ici $J_q = 1$).

Pour les J_c tableaux de contingence associés chacun à une des J_c questions ouvertes (ici $J_c = 2$), f_{ikj} indique la fréquence avec laquelle la réponse-ligne i emploie le mot-colonne k pour répondre à la question j . Cette fréquence est relative à l'ensemble des tableaux de contingence : $\sum_{ikj} f_{ikj} = 1$. En outre, conformément à l'usage, les termes marginaux sont notés en remplaçant par un point l'indice sur lequel on fait la sommation (ex : $f_{i.j} = \sum_k f_{ikj}$).

4. Démarche classique : un seul sous-tableau actif, les autres en supplémentaires

Les méthodes les plus classiques d'analyse de données permettent d'analyser un tableau de données de même nature (tableau de contingence, tableau de variables continues ou tableau disjonctif complet) en donnant aux lignes et aux colonnes de ce tableau un statut actif et d'illustrer ensuite les résultats à l'aide de colonnes (ou de lignes) supplémentaires, éventuellement de nature différente. Ainsi, on peut donner aux questions fermées relatives à la lecture le statut actif dans une analyse des correspondances multiples (ACM) sur le tableau disjonctif correspondant, puis illustrer les plans principaux en projetant les colonnes-mots comme colonnes supplémentaires. On peut aussi analyser le tableau de contingence correspondant à l'une des questions ouvertes au moyen d'une analyse des correspondances binaires (ACB) et ensuite projeter les variables indicatrices des différentes opinions sur la lecture comme éléments supplémentaires.

Les deux paragraphes suivants décrivent les principaux résultats obtenus au moyen de cette démarche.

TABLEAU 3. – Questions fermées relatives à la lecture.

1. <i>A l'école, nous lisons</i>	(peu, assez, beaucoup)
2. <i>Chez moi, nous avons</i>	(peu, assez, beaucoup de livres)
3. <i>Je lis</i>	(peu, assez, beaucoup)
4. <i>Je lis avec</i>	(facilité, quelques difficultés, de grandes difficultés)
5. <i>Les livres donnés par l'instituteur</i>	(me plaisent, ne me plaisent pas)
6. <i>Je lis quand</i>	(j'en ai envie, je fais un travail scolaire, les deux)
7. <i>Je préfère lire</i>	(en silence, à voix haute, des deux façons)
8. <i>Lire les manuels scolaires</i>	(me plaît, ne me plaît pas, me plaît quelquefois)

4.1. Analyse avec questions fermées actives

On réalise une ACM en prenant comme modalités actives les modalités de réponse aux questions fermées relatives à l'attitude envers la lecture (*cf.* tableau 3). Les mots employés dans les réponses ouvertes sont alors considérés comme des colonnes-fréquences supplémentaires. La figure 1 présente le premier plan principal obtenu.

4.1.1. Modalités actives

On choisit d'interpréter les bissectrices ce qui n'est pas en contradiction avec les deux premiers pourcentages d'inertie assez proches.

La seconde bissectrice suit la progression de l'intensité de lecture : *je lis peu, je lis assez, je lis beaucoup*. Les élèves qui déclarent *je lis peu* rencontrent

ANALYSE SIMULTANÉE DE QUESTIONS OUVERTES ET QUESTIONS FERMÉES

de grandes ou moyennes difficultés à lire (*je lis peu* occupe une position intermédiaire entre les modalités *je lis avec de grandes difficultés* et *je lis avec quelques difficultés*). Par contre, les « gros » lecteurs (*je lis beaucoup*) lisent avec facilité.

La première bissectrice concerne l'intensité de lecture scolaire : elle oppose la modalité à l'école *je lis peu* aux modalités à l'école *je lis assez* et à l'école *je lis beaucoup*. Les élèves qui déclarent lire peu à l'école sont souvent des enfants qui déclarent *lire les manuels scolaires ne me plaît pas* et *les livres donnés par l'instituteur ne me plaisent pas*; à l'opposé ceux qui lisent assez ou beaucoup à l'école ont une opinion plus positive sur les textes de lecture scolaire : ils déclarent fréquemment *lire les manuels scolaires me plaît* et *les livres donnés par l'instituteur me plaisent*.

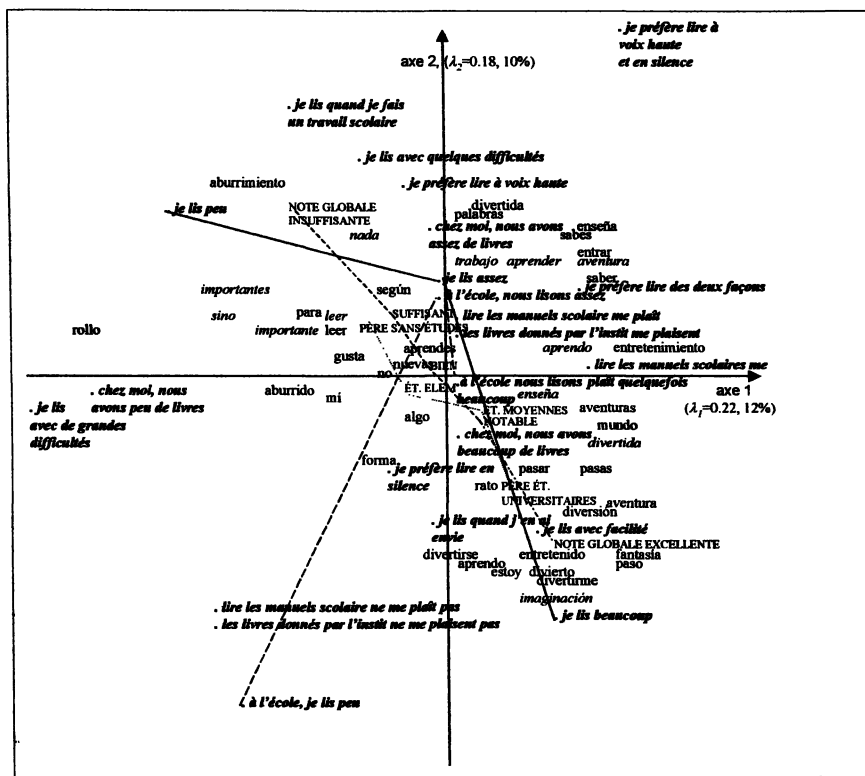


FIGURE 1. – Premier plan principal de l'ACM décrite section 4.1. Représentation des questions fermées sur la lecture (en gras), de quelques autres questions fermées (majuscules), de quelques mots de la question ouverte 1 (droit) et 2 (italiques).

Les trois modalités de la variable *je lis quand* (*quand je fais un travail scolaire*, *quand j'en ai envie*, *les deux*) occupent des positions clairement différenciées, dans les premier, deuxième et troisième quadrants. Les « petits » lecteurs ont tendance à ne lire que dans le cadre scolaire. Parmi les « moyens » et « gros »

lecteurs, se trouvent des enfants qui associent clairement la lecture à une activité de loisir (*je lis quand j'en ai envie*) et d'autres qui considèrent la lecture comme un loisir mais aussi comme liée au travail scolaire (*les deux*).

4.1.2. Modalités de questions fermées illustratives

Certaines questions fermées illustratives ont été représentées. Ainsi, par exemple :

- le niveau scolaire augmente le long de la deuxième bissectrice : l'intensité de lecture et les résultats scolaires sont très liés, alors qu'ils ont peu à voir avec la lecture scolaire ;
- le niveau d'études du père progresse dans la même direction que les qualifications scolaires ;
- les bons élèves, dont le père est souvent plus diplômé que la moyenne, ont tendance à se déclarer « gros » lecteurs ; mais que signifie lire pour eux ? C'est ce que l'on cherche à connaître à travers les réponses aux questions ouvertes.

4.1.3. Questions ouvertes illustratives

Les mots employés au moins 10 fois pour répondre à l'une ou l'autre des deux questions ouvertes (91 mots pour la première ; 92 mots pour la seconde) servent aussi à illustrer l'ACM. Les mots soulignés sur la figure 1 correspondent à la deuxième question ouverte.

On ne s'étonne pas du fait que beaucoup de « petits » lecteurs qualifient la lecture de très ennuyeuse (avec les mots *rollo*, *aburrimiento* dans des expressions comme *es un aburrimiento* ou *es un rollo*). Interrogés sur l'importance de la lecture, ces enfants emploient fréquemment le mot *sino* (*sinon*) : *porque sino no sabemos leer, sino serias analfabeto, sino no tendrás trabajo*, etc. Les « petits » lecteurs scolaires (quadrant inférieur gauche) ne sont guère caractérisés par les mots qu'ils emploient : en fait, ils sont relativement peu nombreux (86 enfants choisissent la modalité à l'école *je lis peu*). Dans la partie positive du premier axe, zone des « moyens » et « gros » lecteurs, on trouve des mots qui expriment bien une opinion positive sur la lecture : *divertido, divertida, bonito entretenimiento, entretenido, fantasía, aventura, imaginación, entrar, mundo*, etc. Néanmoins, ces mots se trouvent mélangés avec d'autres :

- qui montrent un intérêt beaucoup plus modéré pour la lecture (*pasar rato*, de l'expression *pasar el rato*, « faire passer le temps »),
- qui insistent sur le rôle de la lecture dans l'apprentissage scolaire (*aprendo, aprender*, etc.).

En outre, les mots qui semblent mieux traduire un réel goût pour la lecture (*fantasía, aventura, imaginación, entrar, mundo*) occupent des positions moyennes sur la deuxième bissectrice, axe d'intensité de lecture.

4.2. Analyse des réponses avec une question ouverte active

Dans une deuxième analyse dont nous ne reproduisons pas les plans principaux, la seule première question ouverte est active. Le tableau de contingence croisant les réponses individuelles et les mots prononcés au moins 10 fois dans l'ensemble des réponses à cette question est soumis à une analyse des correspondances binaires. Cette analyse présente des difficultés particulières, liées au grand nombre de zéros que contient le tableau de données (chacune des réponses ne mentionne qu'un nombre réduit de mots) et au caractère particulier des données textuelles : si l'on conserve les mots-outils (articles, prépositions, etc.), des réponses qui expriment un « contenu » différent sous une même « forme » peuvent être considérées très semblables; ainsi, par exemple, lors de l'analyse de la première question ouverte, les enfants qui ont prononcé les réponses *es un aburrimiento* et *es un entretenimiento* se retrouvent fortement rapprochés par la forme semblable alors qu'ils expriment des opinions opposées.

Une stratégie habituelle consiste à éliminer ces mots-outils, pour ne conserver que les mots porteurs de contenu. Dans cet exemple, après ce traitement préliminaire, le très petit groupe d'individus qui donnent la réponse *es un aburrimiento* s'oppose alors à l'ensemble de l'échantillon sur le premier et le deuxième axes; cette opposition est due au fait qu'ils emploient exactement le même mot (*aburrimiento*) et que ce mot est très peu employé par les autres répondants. Ce petit groupe (composé de sept enfants) est responsable de 89.4 % de l'inertie du premier axe.

Si on déclare le mot *aburrimiento* comme mot supplémentaire, un autre mot tend à jouer un rôle prépondérant pour le premier axe. Ce problème est fréquent lors de l'analyse directe des réponses : lorsque les réponses sont courtes, il est habituel d'obtenir des premiers axes dus à de petits noyaux d'individus. Ceci nous a conduit à rejeter cette méthodologie pour cette étude.

4.3. Analyse avec questions fermées et ouvertes actives

Lors de la première analyse, la dispersion des individus était uniquement due aux questions fermées. Pour réaliser une analyse qui tienne à la fois compte des questions fermées et des questions ouvertes, il est nécessaire d'adopter une méthodologie qui puisse traiter comme actives les colonnes de tableaux de diverses natures tout en équilibrant leur influence dans la construction des axes.

Ceci nous conduit à une méthodologie de type analyse factorielle multiple (AFM) (Escofier & Pagès, 1988, 1998). Nous présentons successivement :

- les principes de l'AFM usuelle, qui permet l'introduction de tableaux de contingence à condition que leur dimension commune ait la même marge (Abdessemed et Escofier 1996);
- son extension à l'analyse simultanée de plusieurs tableaux de contingence ayant des marges différentes pour leur dimension commune, appelée analyse factorielle multiple pour tableaux de contingence (AFMTC) (Bécue & Pagès, 1999);

- la prise en compte simultanée de tableaux de contingence et de variables qualitatives.

5. Analyse de tableaux multiples

5.1. Brève présentation de l'AFM

L'analyse factorielle multiple analyse un tableau multiple dans lequel un ensemble d'individus est décrit par J groupes de variables, quantitatives ou qualitatives. L'AFM réalise une analyse en composantes principales (ACP) non normée du tableau juxtaposé, en pondérant les variables de façon à équilibrer l'influence de chacun des groupes. Pour cela, le poids des colonnes-variables du groupe j est divisé par λ_1^j , première valeur propre obtenue lors de l'analyse séparée – ACP ou ACM selon la nature des variables – du sous-tableau j . Cette méthode offre des résultats :

- analogues à ceux de l'ACP ou de l'ACM; principalement une représentation globale des lignes (individus) et des colonnes (variables ou modalités);
- spécifiques des tableaux multiples; principalement, la représentation superposée des structures induites par chacun des sous-tableaux sur l'ensemble des individus (ou catégories d'individus) – appelées structures partielles – et la représentation des facteurs issus des analyses séparées.

5.2. AFM sur Tableaux de Contingence (AFMTC)

Une méthodologie pour l'analyse simultanée d'un ensemble de tableaux de contingence a été proposée par Bécue & Pagès (1999). Elle reprend les principes de l'analyse des correspondances binaires intra-tableaux (Benzécri, 1983; Escofier & Drouet 1983) généralisée (Cazes & Moreau, 1991, 2000) sous le nom d'analyse de correspondances internes (ACI), terminologie que nous adoptons, et prend ainsi en compte les différences entre les marges en lignes. Elle reprend aussi ceux de l'AFM pour équilibrer l'influence des différents tableaux et pour fournir des représentations graphiques spécifiques de la structure en groupes de colonnes. Nous en présentons les principes de base dans les lignes qui suivent.

Chaque tableau *Individus* \times *Mots* constitue un tableau de contingence. Pour les introduire dans une AFM, on remplace chacun d'eux par le tableau de terme général x_{ikj} (ligne i , colonne k , sous-tableau j) :

$$x_{ikj} = \frac{f_{ikj} - \left(\frac{f_{i..} \cdot f_{.kj}}{f_{..}}\right)}{f_{i..} \cdot f_{.kj}} = \frac{1}{f_{i..}} \left[\frac{f_{ikj}}{f_{.kj}} - \frac{f_{i..}}{f_{..}} \right]$$

L'ACP de ce tableau conduit aux résultats de l'ACI. On reconnaît, entre crochets, l'écart entre un profil colonne et le profil colonne marginal du tableau auquel elle appartient.

En AFM, l'équilibre entre sous-tableaux est obtenu en surpondérant chaque colonne du tableau j par $1/\lambda_1^j$ inverse de la première valeur propre de l'analyse séparée du sous-tableau j .

Dans l'AFMTC, les analyses séparées des sous-tableaux ne sont pas des ACB usuelles mais des analyses de correspondances binaires imposant aux lignes un poids constant dans toutes ces analyses, poids égal à la marge en ligne du tableau juxtaposé (analyses dites pseudo-séparées). Ce sont ces analyses de correspondances binaires avec marges imposées qui sont utilisées pour calculer les premières valeurs propres λ_1^j . La déformation inhérente à la substitution des ACB séparées par les ACB pseudo-séparées, minime si les marges en ligne diffèrent peu selon les tableaux, résulte du compromis nécessaire pour comparer des tableaux de contingence dont les marges diffèrent.

Dans l'exemple, les analyses séparées et pseudo-séparées conduisent à des facteurs quasiment identiques. Dans la suite, par souci de clarté, on ne mentionne plus cette distinction.

On réalise ensuite l'analyse globale qui consiste en une ACP non normée du tableau juxtaposant l'ensemble des sous-tableaux de contingence avec :

- terme général : x_{ikj} défini ci-dessus ;
- poids des lignes-individus (et métrique dans l'espace des colonnes) : $\{f_{i..}; i = 1, \dots, I\}$, $f_{i..}$ étant le poids relatif moyen des lignes-individus sur l'ensemble des tableaux de contingence considérés ;
- poids des colonnes du groupe j (et métrique dans l'espace des individus) : $\{f_{.kj}/\lambda_1^j; k = 1, \dots, K_j; j = 1, \dots, J_c\}$.

Les graphiques obtenus et les aides à l'interprétation sont similaires à ceux de l'AFM.

5.3. Analyse du tableau multiple comportant des sous-tableaux de contingence et des sous-tableaux de variables indicatrices

Pour introduire à la fois des tableaux de contingence et des tableaux de variables indicatrices dans une même analyse, on considère les tableaux de contingence comme en ACI et les tableaux de variables indicatrices comme en ACM.

Dans toutes les analyses séparées – ou plutôt pseudo-séparées – (ACB ou ACM, selon le type de tableau), les lignes ont le même poids, calculé à partir de la marge en ligne des tableaux de contingence juxtaposés. On reprend ainsi la solution proposée par Abdessemed & Escofier (1996) pour analyser conjointement des tableaux de variables qualitatives et un tableau de fréquence (ou plusieurs tableaux de fréquence, mais de marges égales).

Ensuite, on adopte la démarche de l'AFM en effectuant une ACP non normée sur le tableau juxtaposant l'ensemble des sous-tableaux avec :

- comme poids des lignes-individus (et métrique dans l'espace des colonnes) : $\{f_{i..}; i = 1, \dots, I\}$, où $f_{i..}$ est le poids relatif moyen des lignes-individus sur l'ensemble des tableaux de contingence considérés ;

- comme poids des colonnes du groupe j (et métrique dans l'espace des individus) le poids initial de la colonne divisé par λ_1^j , donc $\{(I_k/IJ)/\lambda_1^j; k = 1, \dots, K_j; j = 1, \dots, J_q; f_{.kj}/\lambda_1^j; k = 1, \dots, K_j; j = J_q + 1, \dots, J_c\}$.

Cette analyse globale offre des résultats :

- analogues à ceux de l'ACB appliquée aux tableaux juxtaposés (principalement une représentation globale des lignes-individus et des colonnes-modalités et colonnes-mots);
- spécifiques des tableaux multiples (principalement une représentation superposée des structures induites par chacune des questions ouvertes et par chacun des groupes de questions fermées – appelées structures partielles – et la représentation des facteurs des analyses séparées).

La lecture des résultats est facilitée par les nombreuses aides à l'interprétation de l'AFM.

6. Application

6.1. Analyses séparées

Dans les analyses séparées, les premières valeurs-propres valent respectivement 0.22 dans le groupe des questions fermées, 0.51 dans le groupe de la première question ouverte et 0.49 dans le groupe de la deuxième question ouverte. La différence entre ces valeurs nécessite d'équilibrer l'influence des groupes.

6.2. Analyse globale

La figure 2 montre un extrait de la représentation des colonnes sur le premier plan principal. Nous ne donnons que quelques commentaires, en cherchant surtout à illustrer les règles d'interprétation de cette analyse qui, à très peu près, sont celles d'une ACM pour les questions fermées et celles d'une ACB pour les questions ouvertes.

Les trois groupes de colonnes contribuent à la construction du premier axe avec respectivement 46 %, 31 % et 23 % de l'inertie. Cette direction de dispersion est commune aux trois groupes et non confondue avec les premiers axes des analyses séparées; en effet, la première valeur propre de l'analyse globale (1.4) est éloignée de la valeur maximum (ici 3) obtenue lorsque les premiers axes des analyses séparées sont identiques.

L'apport du groupe 1 (questions fermées) à la construction du deuxième axe est faible (18.5 %), les groupes 2 et surtout le 3 (questions ouvertes) ayant une plus forte contribution (respectivement 32 % et 49 %).

Le premier axe de l'AFM oppose les mots qui indiquent un fort goût pour la lecture (à gauche de l'axe) aux mots qui signifient un franc rejet de celle-ci. La représentation simultanée des individus, permet de voir que cet axe oppose des enfants qui offrent des réponses comme « Amusant et intéressant » (Q1), « On apprend des choses nouvelles » (Q2) à d'autres enfants qui répondent « Assez rasoir » (Q1), « Pour moi, ce n'est pas trop important » (Q2) ou « Un ennui » (Q1), « Pour savoir » (Q2). Ce premier axe est largement apparenté à

ANALYSE SIMULTANÉE DE QUESTIONS OUVERTES ET QUESTIONS FERMÉES

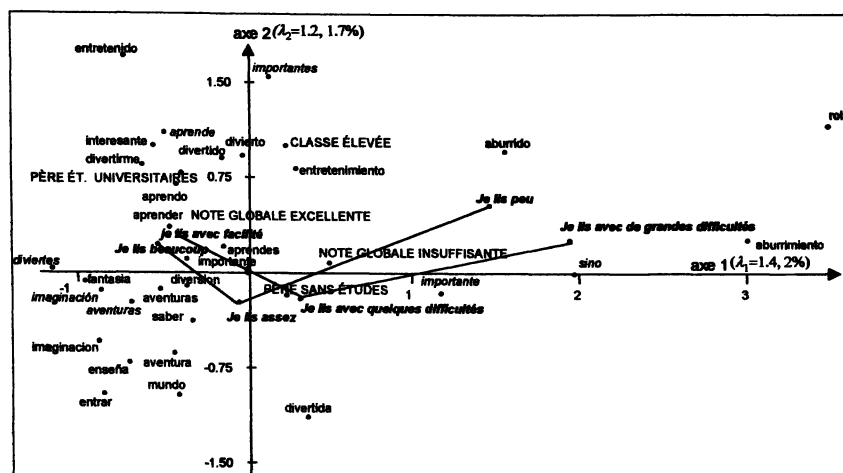


FIGURE 2. – Premier plan principal de l'AFM du tableau multiple décrite section 6 : quelques questions fermées sur la lecture (en gras), quelques autres questions fermées (majuscules), quelques mots de la question ouverte 1 (droit) et 2 (italiques).

la deuxième bissectrice de l'ACM des questions fermées relatives à la lecture : il oppose les enfants qui lisent peu aux enfants qui lisent beaucoup : ainsi, *je lis peu* est à la fois bien caractérisé dans les deux analyses. Néanmoins ces deux directions ne coïncident pas : *je lis beaucoup* est surtout bien caractérisé dans l'ACM.

Du point de vue des questions ouvertes, les élèves qui ont répondu *je lis beaucoup* et *je lis assez* diffèrent peu, d'où la position moins excentrique de *je lis beaucoup* dans l'AFM qui prend en compte l'ensemble des données. Le deuxième axe de l'AFM, principalement dû à la deuxième question ouverte, oppose une vision scolaire de la lecture à une vision non scolaire. Ainsi cet axe oppose des enfants qui soulignent l'importance de la lecture « Apprendre » (Q1), « On apprend des choses nouvelles » (Q2) à d'autres qui insistent sur le plaisir de la lecture « C'est très intéressant parce que tu peux aller au monde de la fantaisie. Par exemple, tu peux voyager au moyen âge ou te retrouver dans la préhistoire » (Q1), « Sans la lecture, je ne pourrais pas voyager » (Q2). Cette opposition n'est pas sans rappeler la première bissectrice de l'ACM, mais les variables qualitatives liées à cette deuxième bissectrice (à l'école nous lisons, les livres donnés par l'instituteur et lire les manuels scolaires) ne sont pas liées au deuxième facteur de l'AFM. Ce ne sont pas les mêmes élèves qui ont une vision non scolaire de la lecture selon le type de question, ouverte ou fermée.

6.3. Corrélations entre les facteurs des analyses séparées

Les corrélations entre les premiers facteurs des trois analyses séparées sont faibles : ces analyses ne peuvent pas être facilement comparées. L'apport

relativement équilibré de chaque groupe au premier axe global de l'AFM montre que cet axe correspond à une direction de dispersion commune aux trois groupes, direction qui n'est pas mise en évidence par les analyses séparées.

6.4. Représentation superposée des nuages partiels

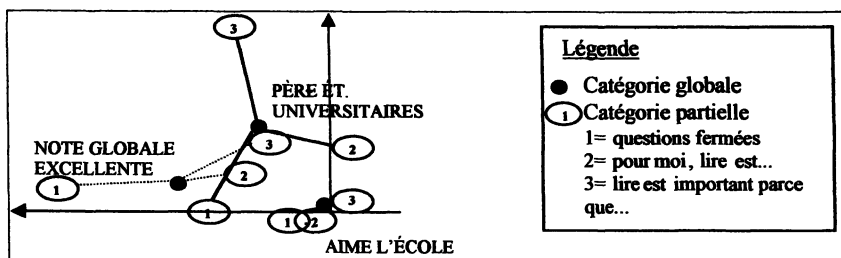


FIGURE 3. – Extrait de la superposition de la représentation globale des catégories et de leur représentations partielles, induites par chacun des groupes.

L'AFM fournit une représentation dans laquelle chaque individu (ou catégorie d'individus) apparaît représenté à la fois par un point issu de l'analyse globale et par autant de points, appelés point partiels, qu'il y a de groupes de colonnes. Pour cela, on introduit, en tant que lignes supplémentaires du tableau juxtaposé, les lignes de chacun des sous-tableaux complétées par des zéros, de façon à ce qu'elles aient le même nombre de colonnes que le tableau global. Il est alors possible de projeter ces lignes-individus supplémentaires sur les axes ou les plans principaux globaux. Ces projections constituent les représentations partielles des individus. Sur le graphe superposant les représentations globale et partielles, ces dernières sont dilatées par un coefficient égal au nombre de tableaux. Ainsi, le point global se retrouve au centre de gravité des points partiels.

Quand les individus sont nombreux, cas des données d'enquête, on préfère représenter non pas les individus mais plutôt les catégories d'individus définies par les modalités des questions fermées. Dans cette optique, une modalité est considérée comme le barycentre des individus qui la possèdent.

A titre d'exemple, examinons la modalité *note globale excellente*. Les bons élèves sont de très « gros » lecteurs selon les questions fermées, alors qu'ils occupent une position plus centrale selon les questions ouvertes. Il semble que, dans cet exemple, il existe une tendance très marquée à donner la « réponse correcte », à se déclarer « gros » lecteur, peut-être à cause du cadre scolaire dans lequel s'est réalisé l'enquête. Le mode de questionnement ouvert permet moins de se cacher derrière des réponses conventionnelles.

7. Conclusion

Dans les questionnaires, on utilise fréquemment à la fois des questions ouvertes et des questions fermées. Les réponses ouvertes sont très riches mais leur analyse directe est souvent sensible à de petits groupes d'individus, d'où l'intérêt d'une analyse simultanée qui prenne en compte les deux types de question pour profiter à la fois de la solidité des structures induite par les questions fermées et de la richesse procurée par les questions ouvertes.

L'AFM étendue à des tableaux mixtes (tableaux de contingence et tableaux disjonctifs complets) offre cette possibilité : elle équilibre l'influence de plusieurs groupes de variables de nature différente dans une analyse globale, fournit une représentation simultanée de l'ensemble des colonnes, fréquences et modalités, et pour chaque individu (ou centre de gravité de groupes d'individus) une représentation selon chaque groupe de colonnes. Cette démarche permet une analyse construite à partir des réponses individuelles et non pas, comme cela est fait souvent, à partir des réponses regroupées par catégories d'individus.

Logiciel utilisé

L'AFMTC utilisée ici est incluse dans le logiciel ADDAD et peut être obtenue, auprès de l'ADDAD, sous forme d'un programme DOS indépendant.

Bibliographie

- ABDESSEMED L. et ESCOFIER B. (1996), Analyse factorielle multiple de tableaux de fréquences ; comparaison avec l'analyse canonique des correspondances, *Journal de la Société de Statistique de Paris*, 137(2), pp. 3-18.
- BÉCUE M. et PAGÈS J. (1999), Intra-Sets Multiple Factor Analysis. Application to textual data, in *Proc. of the 9th International Symposium on Applied Stochastic Models and Data Analysis*, edited by J. Jansen et al., pp. 51-60. Universidade de Lisboa, Lisboa.
- BENZÉCRI J.-P. (1983), Analyse de l'inertie intraclasse par l'analyse d'un tableau de contingence, *Les Cahiers de l'Analyse des Données*, 8(3), pp. 351-358.
- CAZES P. and MOREAU J. (1991), Analysis of a contingency table in which the rows and the columns have a graph structure, in *Symbolic-numeric data analysis and learning*, edited by E. Diday and Y. Lechevallier, pp. 271-280. Nova Science Publishers, New York.
- CAZES P. et MOREAU J. (2000), Analyse des correspondances d'un tableau de contingence dont les lignes et les colonnes sont munies d'une structure de graphe bistoche, in *L'analyse des correspondances et les techniques connexes. Approches nouvelles pour l'analyse statistique des données*, edited by J. Moreau, P.A. Daudin and P. Cazes, pp. 87- 103. Springer-Verlag, Berlin-Heidelberg.
- ESCOFIER B. et DROUET D. (1983), Analyse des différences entre plusieurs tableaux de fréquence, *Les Cahiers de l'Analyse des Données*, 8(4), pp. 491-499.
- ESCOFIER B. et PAGÈS J. (1988-1998), *Analyses factorielles simples et multiples ; objectifs, méthodes et interprétation*, Dunod, Paris.

ANALYSE SIMULTANEE DE QUESTIONS OUVERTES ET QUESTIONS FERMEES

ESCOFIER B et PAGES J (1994), Multiple Factor Analysis AFMULT package,
Computational statistics & data analysis, 18, pp 121-140

LEBART L et SALEM A (1994), *Statistique Textuelle*, Dunod, Paris

RAJADELL PUIGGROS N (1990), Les attitudes envers la lectura, un model d'anàlisi
a l'Educació Primària, These de doctorat (en catalan), Universitat de Barcelona,
Barcelona