L'analyse factorielle des correspondances dans les sciences sociales

Autor(en): **Lorenzi-Cioldi, Fabio**

Objekttyp: **Article**

Zeitschrift: **Schweizerische Zeitschrift für Soziologie = Revue suisse de sociologie = Swiss journal of sociology**

Band (Jahr): **9 (1983)**

Heft 2

PDF erstellt am: **22.07.2024**

Persistenter Link: https://doi.org/10.5169/seals-814195

Schweiz. Z. Soziol. / Rev. suisse sociol. 2 (1983) 365–390

L'ANALYSE FACTORIELLE DES CORRESPONDANCES DANS LES SCIENCES SOCIALES

Fabio Lorenzi-Cioldi
Département de Sociologie – FSES
Faculté de Psychologie et des Sciences de l'Education
Université de Genève, CH–1211 Genève 4

ZUSAMMENFASSUNG

Dieser Artikel präsentiert eine Methode der Datenanalyse, die heute sehr weit verbreitet ist: die Faktorenanalyse der Korrespondenzen. Er ist in intuitiver, nichtmathematischer Sprache gehalten, weil er für den Humanwissenschafter gedacht ist, der zwar nicht über sehr intensive Statistikkenntnisse verfügt, aber doch eine multivariate Analyse seiner Daten ins Auge fasst. Nach einer allgemeinen Beschreibung der Methode werden die wichtigsten Etappen der Analyse nachvollzogen und kommentiert, und zwar von der Datenaufbereitung bis zur Einbeziehung von illustrativen Variablen und der Interpretation der Resultate. Die Darlegung stützt sich auf ein numerisches Beispiel, das sehr allgemein gehalten ist, und weist darüber hinaus auf mögliche Varianten hin, die zur Auswertung des selben Datensatzes dienen können.

RESUME

Cet article présente une méthode d'analyse des données dont il est largement fait usage aujourd'hui: l'analyse factorielle des correspondances. Ecrit dans un langage intuitif, non mathématique, il s'adresse au chercheur en sciences sociales qui n'a pas de connaissances approfondies en statistique, et qui envisage d'appliquer un traitement multivarié aux données dont il dispose. Après avoir exposé le sens général de la méthode, l'article retrace et commente les principales étapes de l'analyse, depuis la mise en forme des données, jusqu'à l'introduction de variables illustratives et à l'interprétation des résultats. L'exposé prend appui sur l'étude d'un exemple numérique ayant une portée générale. Il indique en outre quelques variantes possibles qui peuvent être utilisées pour l'analyse d'un même ensemble de données.

Le terme d'analyse des données regroupe une multitude d'algorithmes statistiques très différents mais solidaires d'un même but: la description simple et parcimonieuse de grands ensembles de données. De manière générale, ces méthodes ne nécessitent pas la formulation d'hypothèses très contraignantes au départ de l'analyse. Le choix de l'une ou de l'autre d'entre elles dépend pour une large part du type de question que le chercheur se pose face aux données qu'il a recueillies.

Ainsi, par exemple, l'analyse discriminante recherche la distinction maximale entre des groupes d'individus lorsque l'appartenance des individus à ces groupes (de sexe, de profession, etc.) est connue au préalable, alors que les méthodes typologiques ou les classifications automatiques visent à regrouper les individus en un nombre restreint de classes discontinues et relativement homogènes. L'analyse de la régression multiple tend à maximiser les relations entre une variable dépendante dont il s'agit d'expliquer la variance (par exemple une réponse ou une combinaison de réponses calculée à l'aide d'une autre analyse) et des variables indépendantes dont on teste le rôle causal. Les techniques d'analyses factorielles, qui comprennent plusieurs algorithmes dont il est parfois difficile d'apprécier les différences pratiques, se fondent sur des calculs de corrélations et concernent la mise en évidence de structures d'interrelations entre les réponses.

Or, de même que l'usage du thermomètre n'est pas réservé à ceux qui connaîssent la théorie physique de la dilatation des corps, celui des méthodes d'analyse des données ne doit pas obligatoirement se fonder sur une connaissance approfondie de l'aspect mathématique des techniques.

Le présent exposé se veut un texte de vulgarisation: il entend fournir une approche non mathématique de l'analyse factorielle des correspondances. Il s'adresse au chercheur en sciences sociales qui, n'étant pas familier avec le traitement multivarié des données, se propose d'évaluer le profit qu'il peut retirer de l'application de cette méthode aux données dont il dispose (des données d'enquête, des données sociodémographiques, expérimentales, ou d'autres encore). Dans ce contexte, certaines propositions que nous avancerons pourront choquer le statisticien, mais ce texte a été conçu pour ceux qui travaillent *à l'aide* d'une méthode et non pas *sur* la méthode elle-même. De nombreux traités détaillés sur l'analyse des données paraîssent chaque année. Toutefois, leur lecture suppose trop souvent la possession d'une attitude particulière face aux données, attitude qui se manifeste par la capacité à exprimer, sinon à résoudre, des problèmes dans un langage opérationnel, à formuler des questions méthodologiquement pertinentes, etc. A ce stade préliminaire de l'investigation le chercheur est ainsi laissé à lui-même. L'ambition de ce texte est précisément celle de contribuer à combler ce vide dans le domaine de l'analyse factorielle des correspondances.

1. INTRODUCTION

L'analyse factorielle des correspondances (AFC) est une méthode descriptive multivariée dont le champ d'application est extrêmement vaste. Proposée dans les années 60 par J.-P. Benzécri, elle donne une formulation originale de quelques principes de techniques alors déjà opérationnelles, comme l'analyse en composantes principales (Pearson, 1901; Hotelling, 1933), l'analyse canonique (Hotelling, 1936) ou l'analyse discriminante (Rao, 1952). L'originalité de la méthode réside dans le fait qu'elle permet la représentation simultanée de deux ou plusieurs ensembles de caractères qualitatifs décrivant un certain nombre d'individus ou d'entités quelconques. [1]

Historiquement, le premier champ d'application de l'AFC fut une branche de la linguistique, la lexicologie (cf. Benzécri, 1977, 9–40; Benzécri, 1981). De grands tableaux statistiques représentant les fréquences d'apparition des modalités de différentes unités grammaticales furent ainsi construits et analysés à l'aide de cette méthode (cf. par exemple Bergounioux *et ali,* 1982).

Soit une série I de pronoms personnels représentée par les modalités i_n [2] et J une série d'adjectifs représentée par les modalités j_m. On construit un tableau, appelé tableau de contingence, dans lequel, à l'intersection de la ligne i et de la colonne j, on écrit le nombre k_{ij} de fois que le pronom i est associé ou lié, dans un corpus donné, à l'adjectif j.

Or le nombre de modalités de pronoms (n) et surtout d'adjectifs (m) est généralement assez grand. La lecture immédiate du tableau de contingence qui en résulte laisse s'échapper les multiples relations d'attraction et d'éloignement, de similitude et de différence, d'implication et d'exclusion qui existent parmi tous les pronoms entre eux, parmi tous les adjectifs entre eux, ainsi qu'entre les pronoms et les adjectifs. L'AFC vise précisément à détecter et à représenter de manière rigoureuse et exhaustive les interrelations qui sont présentes dans un tel tableau. Il convient donc de partir de l'étude relativement simple d'un tableau de contingence. Nous évoquerons ensuite les extensions possibles de la méthode d'analyse dans un contexte plus général.

L'AFC contient toutes les prémisses d'une procédure d'inférence statistique: elle pose en effet une hypothèse nulle, selon laquelle les ensembles

1) Pour une description mathématique de l'AFC le lecteur peut se référer aux ouvrages spécialisés mentionnés dans la bibliographie.

2) On dit aussi que I est un *ensemble* ou un *caractère.* L'ensemble est constitué de *points* ou *éléments.*

Ensemble J (adjectifs)

Ensemble I (pronoms)

	j1	j2		jm	
i1	k_{i1j1}	k_{i1j2}		k_{i1jm}	$n_{i1.}$
i2	k_{i2j1}	k_{i2j2}		k_{i2jm}	$n_{i2.}$
in	k_{inj1}	k_{inj2}		k_{injm}	$n_{in.}$
	$n_{.j1}$	$n_{.j2}$		$n_{.jm}$	$n_{..}$

où: i1 = je
i2 = nous
. . . .
in = ils

j1 = tolérant
j2 = menaçant
. . . .
jm = ambitieux

TABLEAU 1:
tableau de contingence
(croisement de deux caractères qualitatifs:
les pronoms personnels I et les adjectifs J).

I et J sont indépendants l'un de l'autre (les effectifs des modalités d'un ensemble sont distribués de manière aléatoire sur les modalités de l'autre ensemble), dans le but de la refuter et de décrire ainsi la forme et l'intensité de la (des) dépendance(s). Comment s'y prend-elle pour refuter cette hypothèse nulle? Pour répondre à cette question, nous allons aborder l'AFC comme la recherche de la meilleure représentation simultanée des deux ensembles I et J.

2. RECHERCHE DE LA MEILLEURE REPRESENTATION SIMULTANEE

Contrairement aux modèles des analyses factorielles et en composantes principales, celui de l'AFC accorde des rôles parfaitement symétriques aux lignes et aux colonnes d'un tableau des données. L'analyse procède par le reclassement des lignes i et des colonnes j du tableau de contingence de manière à faire apparaître des blocs qui se correspondent, ainsi que par l'attribution d'une valeur numérique à chacune des modalités des ensembles I et J, afin de maximiser un coefficient d'association entre les deux ensembles (le coefficient de corrélation "r" de Pearson). En quantifiant les deux caractères, elle opère le passage d'un niveau de mesure nominal ou ordinal (dans le tableau d'origine) à un niveau de mesure intervalle (sur chacun des facteurs de I et de J d'une dimension extraite par l'analyse). C'est ce calcul appelé *codage* ou quantification qui permettra d'évaluer les proximités et les éloignements des modalités à l'intérieur de chaque ensemble et entre les ensembles différents. [3]

Une notion importante en statistique bivariée est celle de *coefficient de détermination:* ce coefficient exprime la proportion de variance partagée par deux caractères. Il peut s'interpréter comme la probabilité avec laquelle la valeur (inconnue) d'un individu sur un caractère peut être prédite, générée, à partir de la valeur (connue) du même individu sur un autre caractère. Or à chaque couple de facteurs (facteur de I et facteur de J) extraits en AFC est associée une valeur propre (lambda: λ) qui indique précisément la proportion de la variance totale partagée par les deux caractères. On a alors:

$$\lambda = r_{IJ}^2 \qquad \text{d'où:} \quad \sqrt{\lambda} = r_{IJ}$$

Le lambda est un indice symétrique qui mesure le degré de liaison ou dépendance entre deux caractères I et J, et donc le degré de prévision de l'un

3) Le codage des modalités d'un tableau qualitatif possède réellement toutes les propriétés associées à un niveau de mesure intervalle. Des indices de dispersion et de concentration peuvent être calculés à l'aide des valeurs des modalités d'un même ensemble.

par l'autre; il existe un tel coefficient pour chaque couple de facteurs (facteur de I et facteur de J) c'est-à-dire pour chaque dimension calculée par l'analyse.

Une seule dimension ne peut généralement pas rendre compte de la totalité de l'information contenue dans le tableau. Après avoir effectué le premier codage, on en cherchera un autre, de corrélation maximum, qui soit indépendant (orthogonal) du premier, et ainsi de suite.

Nous allons illustrer la notion de codage à l'aide d'un exemple fictif. Soit un tableau de contingence (I x J) qui comporte cinq modalités pour chaque caractère. Dans chaque case de ce tableau figure le nombre k_{ij} d'individus ayant répondu à la modalité i de la question I et à la modalité j de la question J, pour un total de 811 individus (voir tableau 2). Nous cherchons une première quantification des caractères I et J afin que leur corrélation soit maximum.

Ensemble J

Ensemble I	j1	j2	j3	j4	j5	
i1	10	70	5	10	40	135
i2	40	1	70	55	20	186
i3	15	40	5	10	70	140
i4	40	5	45	70	35	195
i5	70	5	15	25	40	155
	175	121	140	170	205	811

TABLEAU 2:
exemple numérique: tableau de contingence (I x J); correspondance statistique. Les cases avec des effectifs d'au moins 40 unités sont hachurées.

L'opération d'extraction d'une première dimension comporte deux moments : l'analyse commence par réordonner les lignes et les colonnes de ce tableau, de sorte que depuis la première ligne jusqu'à la dernière, de même que depuis la première colonne jusqu'à la dernière, il s'instaure *simultanément* des gradations qui s'accordent entre elles. Les lignes du haut vont surtout contenir des valeurs élevées vers leur début, les lignes du bas vers leur fin. Le tableau reclassé va contenir les valeurs les plus élevées autour de sa diagonale principale. L'analyse va ensuite quantifier ce tableau en attribuant des nombres à chaque modalité reclassée des caractères I et J. On lit au tableau 3 le produit de cette double opération dans notre exemple.

Ensemble J

Ensemble I	j3: 0,65	j4: 0,49	j1: 0,34	j5: −0,41	j2: −1,24 → Facteur de J	
i2: 0,64	70	55	40	20	1	186
i4: 0,45	45	70	40	35	5	195
i5: 0,23	15	25	70	40	5	155
i3: −0,72	5	10	15	70	40	140
i1: −1,05	5	10	10	40	70	135
↓ Facteur de I	140	170	175	205	121	811

λ = 0,42 ; r = 0,65.

TABLEAU 3 :

exemple numérique : tableau de contingence (I x J), réordonné et quantifié par l'analyse : dimension 1. Les nombres attribués à chaque ligne et à chaque colonne sont indiqués entre paranthèses car, situés sur une échelle d'intervalle, il faudrait représenter les distances qui séparent les modalités entre elles. Les cases d'au moins 40 unités sont hachurées.

La valeur propre indique l'existence d'une bonne association entre les deux caractères ($\lambda = 0{,}42$; $r = 0{,}65$).

L'analyse va nous permettre, outre l'examen de l'ordre et des distances qui séparent les modalités à l'intérieur de chaque ensemble I et J, celui des proximités ou de la correspondance entre les modalités des deux ensembles. La *représentation simultanée* situe un point i et un point j d'autant plus près l'un de l'autre que leur degré d'association est élevé. Un élément i (respectivement j) est entouré des éléments j (respectivement i) avec lesquels il est fortement associé. On dira que i (respectivement j) est en position *barycentrique* par rapport au système des j (respectivement i). Il con-convient cependant de préciser la signification du principe barycentrique. On pourrait croire en effet que chaque modalité d'un ensemble peut être correctement lue, située, interprétée par rapport à chacune des modalités de l'autre ensemble prises séparément. Or, un point i de l'ensemble I n'est pas mis en position par rapport à un seul point j de l'ensemble J qui lui est proche, mais par rapport à tous les points de J. La position d'une modalité i dépend de la position de la *totalité* des modalités j. Une modalité est proche d'une modalité d'un autre ensemble aussi bien parce qu'attirée par elle, que parce que repoussée par d'autres. La lecture des proximités de sous-groupes de points n'est légitime que si l'on se place à l'intérieur d'un même ensemble I ou J. Elle devient totalement erronée dès qu'elle porte sur la correspondance de points issus d'ensembles différents.

L'AFC peut-elle nous amener d'autres informations intéressantes? Pour répondre à cette question, nous allons examiner une deuxième dimension. Nous constatons ainsi qu'à la dimension 2 est associée une valeur propre de 0,10 ($r = 0{,}32$). Le codage de cette dimension est reporté tableau 4.

		Facteur 2			Facteur 2
Ensemble I :	i1	0,25	Ensemble J :	j1	−0,44
	i2	0,27		j2	0,27
	i3	−0,14		j3	0,38
	i4	0,12		j4	0,19
	i5	−0,57		j5	−0,20

TABLEAU 4:
exemplaire numérique; codage des ensembles I et J (facteurs de I et de J) sur la dimension 2.

La valeur atteinte par le lambda étant un des indicateurs possibles de la validité des facteurs extraits, [4] nous renonçons à l'examen d'une troisième dimension.

Afin de disposer d'une vision d'ensemble des résultats de cette analyse, nous pouvons tracer, en utilisant le nouveau codage, un diagramme plan dans lequel sont représentés simultanément les nuages des ensembles I et J sur les axes 1 et 2. On remarque figure 1 la forme particulière de ce nuage. Bien que non corrélées entre elles, les deux dimensions montrent un lien de dépendance mutuelle: alors que le premier axe distribue les modalités extrêmes du tableau réordonné, l'axe suivant s'interprète comme une intensité et figure la conjonction des extrêmes vers le centre; pour le vérifier, il suffit de comparer la distrubution des points des nuages I et J (figure 1) avec l'ordre des lignes et des colonnes du tableau réordonné (tableau 3). Cette dépendance fonctionnelle entre les facteurs, appelée *effet Guttman,* constitue une forme caractéristique des résultats en AFC.

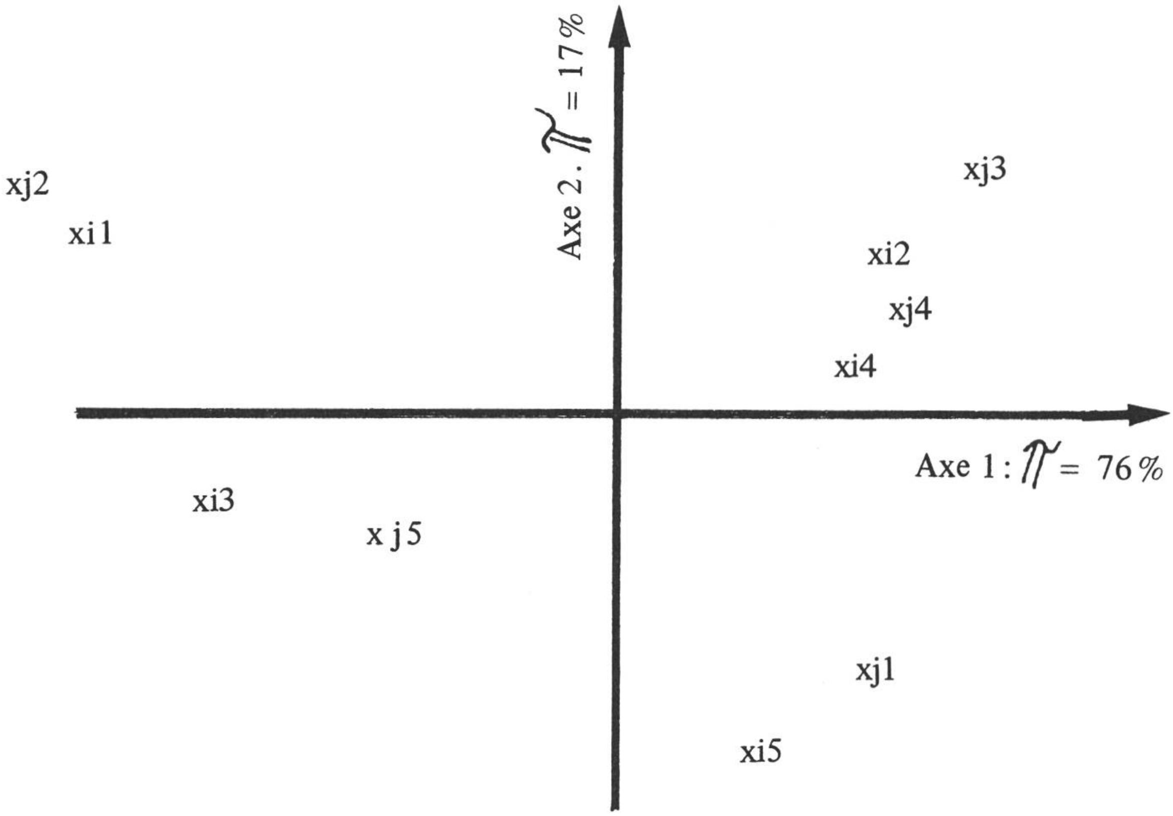

FIGURE 1:
*exemple numérique; représentation simultanée des ensembles I et J sur les axes 1 et 2. Le nuage des points a une forme en croissant parabolique. C'est ce qu'on appelle l'*effet Guttman.

4) Du moins dans le cas des tableaux de contingence.

3. LA PREPARATION DU TABLEAU DES DONNEES

L'AFC permet l'analyse d'un grand nombre de tableaux différents. On peut schématiquement relever les mises en forme des données suivantes.

a) Tableaux de contingence ou de fréquence: l'analyse porte sur une table rectangulaire, où k_{ij} est entier et positif, et représente la fréquence d'association de i à j. Notre exemple numérique concerne une table de contingence (voir tableau 3). Cette table s'applique aisément lorsque l'on dispose d'une population différenciée seulement sur deux critères. Ceux-ci peuvent se rapporter aux réponses à deux questions comportant chacune plusieurs modalités (questions ouvertes, à choix multiple, échelles d'attitude, etc.); un des critères peut aussi représenter la position des répondants sur des critères socio-démographiques tels le sexe, la profession, etc. L'analyse d'un tel tableau est généralement appelée *correspondance statistique.*

b) Tableaux de proportions: k_{ij} n'est pas entier (e. g. une proportion), la somme de toutes les cases du tableau étant égale à 1. L'AFC appliquée aux tableaux a– et b– est dite binaire, car elle étudie les liaisons entre deux caractères.

c) Tableaux logiques: k_{ij} ne prend que les valeurs 0 et 1. Si l'association de i à j est impossible (e. g. /———————————/ une réponse "non"), $k_{ij} = 0$; si l'association est présente (e. g. /———————————/ une réponse "oui"), alors $k_{ij} = 1$. L'importance de ce tableau découle du fait qu'il permet l'extension de l'AFC binaire à plus de deux caractères: on effectue alors *une analyse factorielle* des correspondances multiples (AFCM). Le tableau des données à analyser est appelé "tableau logique disjonctif complet" ou "tableau des indicatrices de l'ensemble qualitatif" (voir tableau 5). Logique: les cases du tableau ne contiennent que des 0 et des 1. Disjonctif: les différentes modalités d'un même ensemble (j' de J', j" de J", etc.) s'excluent mutuellement, chaque ensemble n'admettant qu'une seule réponse par individu. Complet: chaque individu répond effectivement à chaque question (on peut en effet prévoir une colonne j pour coder de manière logique une éventuelle non réponse).

Notons par ailleurs que, moyennant quelques précautions, des données quantitatives peuvent être soumises à l'AFCM. Il suffit pour cela de rendre toutes ces données positives et de les rassembler en classes discrètes mutuellement exclusives.

Le tableau logique a en outre la propriété importante de l'homo-

	Ensemble J = 10										
	J' = 5					J'' = 5					
I = Observations (individus)	j'i	j'2	j'3	j'4	j'5	j''1	j''2	j''3	j''4	j''5	
individu 1	0	0	1	0	0	1	0	0	0	0	2
individu 2	0	1	0	0	0	0	0	1	0	0	2
individu 3	0	0	0	0	1	0	0	0	0	1	2
individu 4	1	0	0	0	0	0	0	1	0	0	2
individu 5	0	0	0	1	0	0	1	0	0	0	2
individu 6	0	0	1	0	0	0	1	0	0	0	2
individu 7	0	0	0	0	1	1	0	0	0	0	2
individu i	0	1	0	0	0	0	0	0	1	0	2
individu 808	0	0	0	1	0	1	0	0	0	0	2
individu 809	1	0	0	0	0	0	0	1	0	0	2
individu 810	0	0	1	0	0	0	1	0	0	0	2
individu 811	0	1	0	0	0	0	0	0	1	0	2
											1622

TABLEAU 5:
tableau logique: correspondance ensembliste. Le tableau analysé précédemment (voir tableau 2) est ici présenté sous forme disjonctive complète.

généité, car toutes les grandeurs recencées sont de même nature (il permet ainsi la réduction de quantité de données hétérogènes). [5] L'analyse d'un tel tableau est généralement appelée *correspondance ensembliste.*

On remarque tableau 5 que, contrairement à la table de contingence classique, chaque ligne i du tableau des données représente l'ensemble des réponses fournies par un seul individu à une batterie de questions dont toutes les modalités sont décrites par les colonnes (dans notre exemple, deux questions comportant chacune cinq modalités de réponse). Le chercheur est intéressé avant tout par la description de comportements de réponse globaux, propres à l'ensemble des individus ou propres à des sous-groupes d'individus interrogés. La mise en correspondance d'ensembles différents, de façon analogue à l'analyse factorielle canonique, va dans la plupart des cas concerner uniquement les colonnes de ce tableau car, comme nous allons le voir, une description exhaustive des caractéristiques des individus peut être introduite dans les colonnes en tant que caractère passif. Il s'avère en outre qu'en AFCM la valeur propre perd en grande partie le sens que nous lui avons attribué jusqu'ici. Notre exemple a montré que les forts effectifs étant placés par l'analyse autour d'une diagonale, l'ordre de grandeur du lambda exprime en quelque sorte le contraste entre les profils du tableau, contraste difficile à produire dans le cas d'un tableau logique (codé en 0 et 1). [6] Les valeurs propres associées à chaque facteur en AFCM seront donc en général assez petites. "Une valeur propre voisine de 1 (toute valeur propre est inférieure ou égale à 1) indique souvent une dichotomie au niveau des données (tableaux en blocs diagonaux); ou seulement quelques isolés qui s'opposent à tout l'ensemble. Des valeurs propres très faibles (e. g. 0,0006) ne doivent pas arrêter l'interprétation: si les profils individuels s'écartent très peu du profil moyen type (profil marginal) les valeurs propres seront très faibles; mais les axes de dispersion n'en seront pas moins interprétables" (Benzécri, 1980, 298).

d) Deux cas particuliers: L'AFC et l'AFCM permettent l'étude de tableaux très différents. Les mises en forme des données recencées ci-dessus se rapportent aux études les plus courantes. La méthode autorise en outre l'examen direct de tableaux ayant une portée moins générale.

Tableaux de préférences: les objets ou item j de J sont classés ou ordonnés par chaque sujet sur un continuum quelconque. A l'intersection de la ligne i et de la colonne j de la table on trouve le rang attribué par le sujet i à l'objet j.

Tableaux de rangs: il s'agit d'un tableau complémentaire au précédent. Une colonne j du tableau contient les rangs associés aux individus I.

5) Pour d'autres propriétés concernant la forme disjonctive, cf. Bouroche et Saporta, 1980, 107–108.

6) Voir Bouroche et Saporta, 1980, 103.

Cette brève revue n'a aucune prétention d'exhaustivité. La construction d'autres tableaux reste envisageable car l'AFC est une technique robuste. Un tableau logique peut par exemple être converti en un *tableau de Burt* (cf. Fénélon, 1981, 117–119). Il convient cependant de réserver de telles applications aux chercheurs qui ont acquis une certaine aisance avec l'analyse des données. Une attention particulière doit en tout cas être accordée à la construction du tableau des données. Il est souvent possible d'analyser un même ensemble de réponses à l'aide de tableaux très différents (par exemple un tableau de contingence ou une table logique). Le chercheur devrait effectuer des exercices sur des tableaux différents, afin de comparer les résultats obtenus au terme de ces analyses.

4. CARACTERES ACTIFS ET CARACTERES ILLUSTRATIFS

Les caractères actifs, comme le suggère leur nom, sont ceux qui contribuent à la définition des axes factoriels (dans les exemples précédents, les ensembles I et J du tableau de contingence ou encore les ensembles J' et J" du tableau logique). L'analyse proprement dite ne porte que sur ces caractères dont les modalités sont appelées *éléments principaux.*

Les caractères illustratifs ou passifs sont des lignes et / ou des colonnes du tableau qui ne participent pas à l'analyse et qui sont projetés sur les axes factoriels déjà calculés en fonction d'une similitude de distribution. Les modalités de ces caractères sont appelées *éléments supplémentaires.*

Une conséquence importante de la technique des variables illustratives est que, ayant été insérées dans l'ensemble des calculs et des graphiques après avoir établi les différents facteurs, elles ne perturbent pas l'analyse. Des individus ou des réponses qui sont, pour une raison ou pour une autre, quelque peu étrangers au problème traité en AFC peuvent ainsi être rajoutés à titre comparatif sans que l'interprétation des résultats de base en soit affectée. Cependant, placer des éléments sur des facteurs déjà définis signifie en même temps ne pas interpréter les relations qu'entretiennent ces éléments supplémentaires *entre eux;* il sera prudent de se limiter à l'interprétation en termes barycentriques des relations qu'entretient chacun de ces éléments avec l'ensemble des éléments principaux.

L'insertion d'éléments supplémentaires en AFC permet la prise en compte d'une grande diversité de problèmes pratiques de recherche en sciences sociales. Nous en mentionnons ici quelques uns :

- l'analyse étant faite, on recueille les réponses relatives à un nouveau groupe d'individus. On situera ce nouveau groupe d'individus par rapport aux facteurs extraits.
- cas d'une étude longitudinale: on analyse les réponses recueillies au temps 1 en projetant sur les facteurs ainsi obtenus les réponses recueillies au temps 2.
- la variance de quelques réponses semble entachée d'erreur (par exemple, le poids de quelques réponses est beaucoup plus faible ou beaucoup plus fort que le poids moyen de l'ensemble des réponses). On refait l'analyse en mettant ces réponses en éléments supplémentaires.
- dans une recherche par questionnaire on dispose d'informations sur l'état civil des répondants (sexe, profession, âge, etc.). Dans une recherche expérimentale on a mesuré l'effet de l'appartenance des sujets à des groupes différents. Toutes les informations concernant les catégories d'appartenance des sujets pourront être ajoutées en tant qu'éléments supplémentaires. Le tableau 6 fournit une illustration de ce cas. [7]

7) Comme nous l'avons déjà évoqué, l'AFC offre des options pour la construction du tableau des données ainsi que pour l'adjonction d'éléments supplémentaires. Le choix définitif de l'analyse la plus satisfaisante repose cependant sur l'expérience même du chercheur. Les cas d'opinions divergentes sont fort nombreux. Relevons par exemple que, pour ce qui est de l'appartenance des sujets à des groupes sociodémographiques, nous avons adopté la position de Benzécri (1980, 218), alors que Bouroche et Saporta (1980, 104) renversent le schéma et proposent d'analyser les informations sociodémographiques en caractères actifs et les réponses en caractères passifs.

			J' = 2		J'' = 2		J''' = 2			J^{IV} = 3		
			Réponse 1 : oui	Réponse 1 : non	Réponse 2 : oui	Réponse 2 : non	Réponse 3 : oui	Réponse 3 : non	TOTAL MARGINAL	Groupe 1	Groupe 2	Groupe 3
I' = 12	Individus du groupe 1	individu 1	0	1	0	1	1	0	3	1	0	0
		individu 2	0	1	1	0	1	0	3	1	0	0
		individu 3	1	0	0	1	1	1	3	1	0	0
		individu 4	1	0	1	0	0	1	3	1	0	0
	Individus du groupe 2	individu 5	1	0	1	0	0	1	3	0	1	0
		individu 6	1	0	1	0	0	1	3	0	1	0
		individu 7	1	0	0	1	0	1	3	0	1	0
		individu 8	0	1	0	1	0	1	3	0	1	0
	Individus du groupe 3	individu 9	1	0	1	0	0	1	3	0	0	1
		individu 10	0	1	1	0	1	0	3	0	0	1
		individu 11	1	0	0	1	1	0	3	0	0	1
		individu 12	0	1	1	0	1	0	3	0	0	1
TOTAL MARGINAL			7	5	7	5	6	6	36	4	4	4
I'' = 3		Réponses groupe 1	2	2	2	2	3	1	12			
		Réponses groupe 2	3	1	2	2	0	4	12			
		Réponses groupe 3	2	2	3	1	3	1	12			

TABLEAU 6 :
tableau logique disjonctif complet représentant les réponses de trois groupes expérimentaux ; l'appartenance des sujets à des groupes différents, de même que la distribution des réponses à l'intérieur de chaque groupe, sont décrits par des éléments supplémentaires.

5. L'INTERPRETATION DES RESULTATS

Il est généralement admis que l'essentiel des résultats d'une AFC est fourni par les graphiques plan qui croisent deux à deux les différentes dimensions. Or des biais de perspective peuvent résulter de la projection simultanée dans un même espace de facteurs issus d'ensembles différents. Il est donc recommandé d'accompagner la lecture de ces graphiques par celle de quelques indices fournis par l'analyse. On peut alors procéder par les étapes mentionnées ci-dessous.

1)

Examen de la *valeur propre* (λ) sur chaque facteur extrait, en termes à la fois de coefficient de détermination et relatifs, par rapport à la somme de toutes les valeurs propres. Ce dernier indice, exprimé en %, est couramment appelé *taux d'inertie* (proportion de la variance totale expliquée par le facteur). Dans notre exemple numérique (voir tableau 3) nous avons retenu deux facteurs dont le taux d'inertie est respectivement 76 % (la première valeur propre représente 76 % de la somme de toutes les valeurs propres) et 17 % (la deuxième valeur propre représente 17 % de la somme de toutes les valeurs propres). Ceci indique, avec un total de 93 %, une bonne reproduction des relations dans nos données. Le taux d'inertie est donc un indicateur de l'importance relative des associations présentes sur le facteur et, "même si elle ne fournit pas d'indications impératives, la suite des valeurs propres et des taux doit être examinée" (Benzécri, 1980, 298). Ce coefficient est un des critères possibles pour la détermination du nombre de facteurs à extraire. Cependant, et surtout dans le cas des tableaux disjonctifs complets (cf. supra), il s'avère que peu d'importance est donnée à ce critère, et que "l'interprétation des facteurs successifs procède puis s'arrête principalement d'après le sens des associations qui apparaîssent, d'après la forme typique des résultats (. . .)" (idem, 297). [8] La valeur propre représente donc la part de la variance totale associée à chaque dimension; elle est ainsi une mesure de l'étendue du nuage de points sur une dimension.

8) Nous savons que, à propos du nombre de facteurs à extraire et à propos d'autres questions encore, une multitude de critères, à caractère probabiliste ou pragmatique, ont été imaginés par les théoriciens de l'analyse factorielle (critère de Kaiser, scree test, etc.). L'AFC est, en revanche, relativement dépourvue de tests probabilistes. Soucieux de faire une "synthèse multidimensionnelle de faits complexes acceptés tels quels" (cf Benzécri, 1980, 289 et 304), Benzécri prône une statistique multivariée qu'on pourrait qualifier de prudente (un autre symptôme de cette prudence étant le refus d'effectuer une quelconque rotation des facteurs extraits).

2)

Les graphiques plan constituent certainement la forme la plus élégante de présentation des résultats (un exemple de représentation simultanée est montré figure 1). Les axes orthogonaux donnent à chaque fois deux dimensions sur lesquelles sont projetés, à l'aide de leur nouveau codage, les éléments de l'ensemble I et/ou ceux de l'ensemble J ; l'origine des axes (point zéro) fournit le centre de gravité de tous les points (l'endroit où le sens du facteur disparaît). D'éventuelles formes typiques (effet Guttman, etc.) peuvent être aisément détectées par l'examen visuel des résultats. On pourra interpréter avant tout les proximités entre éléments d'un même ensemble I ou J en termes de similarité ou d'association. Il faut par contre se rappeler du principe barycentrique (cf. supra) pour interpréter la proximité de deux éléments appartenant à des ensembles différents. [9]

3)

L'étude des *contributions* permet une véritable appréciation de la solution retenue. Outre la coordonnée de chaque point, qui constitue le codage calculé par l'analyse, deux séries de coefficients sont calculés pour chaque facteur et pour tous les éléments des différents ensembles analysés (voir tableau 7). Nous avons vu que, à la différence des modèles classiques de l'analyse factorielle, la coordonnée ou codage d'une modalité i ou j s'interprète comme l'emplacement de cette modalité sur une échelle métrique établie sous la condition de maximiser la corrélation entre les modalités des deux ensembles I et J. La coordonnée n'est donc pas directement interprétable en termes corrélationnels. [10]

Dans les modèles classiques, le carré du poids factoriel d'une modalité fournit immédiatement le coefficient de détermination, indiquant la part de variabilité de cette modalité expliquée par le facteur (ou: partagée avec le facteur). L'AFC quant à elle nous permet d'apprécier la part de la dispersion d'une modalité expliquée par chaque facteur et donc la qualité de la représentation des points dans les plans principaux, à travers l'étude des *contributions relatives* du facteur à l'élément (abréviation : COR). Une COR exprime en un certain sens le degré d'efficacité de l'explication de la variance d'une modalité par un axe. La somme des COR pour une modalité et tous les facteurs est

9) Cf sur ce point Lebart, Morineau, Tabard, 1977, 60; Bouroche et Saporta, 1980, 98.

10) Elle est interprétable en termes corrélationnels dans le cas des coordonnées (plus proprement appelées "saturations" ou poids factoriels) de l'analyse factorielle ou en composantes principales. Notons par ailleurs qu'en AFC quelques coordonnées dépassent parfois les limites du coefficient de corrélation (−1; +1) (voir tableau 7).

FACTEUR 1 :	Ordre	Coordonnée (codage)	CTR (%)	COR (%)
Ensemble I	1 : i2	0,64	22,0	81,3
	2 : i4	0,45	11,5	80,6
	3 : i5	0,23	2,4	13,4
	4 : i3	–0,72	20,9	82,4
	5 : i1	–1,05	43,2	90,7
Ensemble J	1 : j3	0,65	17,4	71,6
	2 : j4	0,49	12,1	77,3
	3 : j1	0,34	5,8	32,9
	4 : j5	–0,41	10,0	62,5
	5 : j2	–1,25	54,8	93,4

FACTEUR 2 :				
Ensemble I	1 : i2	0,27	17,5	14,6
	2 : i1	0,25	11,2	5,3
	3 : i4	0,12	3,4	5,4
	4 : i3	–0,14	3,4	3,0
	5 : i5	–0,57	64,4	81,3
Ensemble J	1 : j3	0,38	25,6	23,7
	2 : j2	0,27	11,7	4,5
	3 : j4	0,19	8,0	11,5
	4 : j5	–0,20	10,7	15,1
	5 : j1	–0,44	43,9	56,1

TABLEAU 7:
exemple numérique; résumé des résultats de l'analyse factorielle des correspondances (facteur 1 et facteur 2).

égale à 100 %. [11] La COR amène des renseignements sur chaque variable en particulier; elle ne nous renseigne que très imparfaitement sur le sens du facteur lui-même. Il existe alors un autre coefficient, appelé *contribution absolue* (abréviation: CTR) qui sert à l'interprétation des axes. [12] La CTR exprime la part prise par une modalité dans la variance expliquée par un facteur. De manière inverse et complémentaire à la COR, elle indique la proportion dans laquelle une modalité donne un sens au facteur. La somme des CTR pour toutes les modalités sur un facteur est égale à 100 %.

Il est courant de fonder l'interprétation d'un axe sur un nombre restreint de modalités, et précisément celles qui apportent à cet axe les plus fortes CTR. Si une modalité a à elle seule une forte CTR (voir tableau 7 l'élément i5 sur le facteur 2), il convient de refaire l'analyse en mettant cette modalité en supplémentaire (ce n'est pas le but d'une analyse qui concerne des associations et des covariations de fournir un axe caractérisé par une seule réponse). Une modalité qui a une faible CTR et une COR très élevée n'est pas intervenue de manière significative lors de la construction de l'axe, tout en étant une modalité très caractéristique de cet axe (voir tableau 7 l'élément i4 sur le facteur 1). Il est inutile de multiplier les exemples d'interprétation que seule la pratique de l'analyse est à même de fournir. Notons seulement que l'étude des contributions dans notre exemple permet, elle aussi, de détecter un effet Guttman dans les résultats, car les plus fortes CTR sur le facteur 1 sont associées aux éléments situés aux extrémités du tableau réordonné, tandis que les plus fortes CTR sur le facteur 2 sont associées aux éléments situés au centre du tableau (voir tableaux 3 et 7).

L'AFC est une technique robuste de description de grands tableaux de données. L'intérêt majeur de son utilisation en sciences sociales découle certainement du fait qu'elle fournit les bases d'une étude rigoureuse de la nature des dépendances entre des modalités issues d'ensembles qualitatifs différents. En analysant les données reportées tableau 6, le chercheur pourrait tester les hypothèses portant sur les covariations des différentes réponses entre elles (J', J", J'''), ainsi que des hypothèses portant sur les dominantes de "patterns" de réponse à l'intérieur de chaque groupe d'individus (J^{IV}); mais il pourrait aussi étudier la position de chaque individu en particulier par rapport à l'ensemble des réponses (I'). L'AFCM est en ce sens un complément et parfois un substitut de l'analyse de variance multivariée, des classifications automatiques, de l'analyse discriminante.

11) La COR est calculée par rapport aux axes retenus, ainsi que par rapport aux axes qui, n'étant pas relevants aux yeux du chercheur, n'ont été rejetés. La somme des COR est appelée *qualité de la représentation* (abrévation: QLT).

12) Certains auteurs préfèrent la dénomination plus explicite de *contribution relative de l'élément au facteur.*

6. UNE ILLUSTRATION EMPIRIQUE

Nous avons soumis à l'AFC un tableau de fréquences qui croise le Sexe, l'Origine, la Couche sociale (ensemble I = 11 modalités) et les Sections scolaires (ensemble J = 8 modalités) fréquentées par les élèves dans les Cycles d'Orientation de l'Enseignement Public Secondaire du Canton de Genève. Nous ne pourrons pas fournir dans ce paragraphe une interprétation exhaustive des résultats obtenus; nous nous bornerons à quelques commentaires visant à procurer au lecteur les moyens nécessaires pour une lecture plus approfondie des résultats présentés ci-dessous. Nous voyons tableau 8 les données à analyser.

Le traitement de ce tableau au moyen de l'AFC répond aux questions suivantes:

- la description des ressemblances et des différences de plusieurs positions sociodémographiques entre elles (ensemble I) ainsi que de plusieurs modes de scolarisation entre eux (ensemble J);
- l'étude systématique des liens entre d'un côté les positions sociodémographiques des élèves (ensemble I) et de l'autre leur scolarisation (ensemble J).

De par leur statut quelque peu excentrique par rapport à l'ensemble considéré ou leur composition hétérogène, nous avons préféré omettre de l'analyse les Sections "Niveaux-Options" et "Accueil" (ensemble J = 2 modalités), et l'Origine "Autres Pays" et la Couche sociale "Divers et Sans Indication" (ensemble I = 2 modalités). Ces modalités ne vont pas affecter la structure des facteurs dégagés, car elles ne seront placées sur ces derniers qu'au terme de l'analyse selon la technique des éléments supplémentaires. De surcroît, l'emplacement sur chaque dimension par rapport à l'ensemble actif va permettre une meilleure appréciation de leur nature.

L'examen des valeurs propres (λ) associées aux dimensions issues de l'analyse nous renseigne sur l'étroitesse du lien entre les facteurs des deux ensembles I et J, ainsi que sur l'importance relative de chacune des dimensions. Les corrélations entre les ensembles I et J sur chaque dimension, exprimées par les racines carrées de chaque valeur propre, sont respectivement de 0,23, de 0,17 et de 0,05. La faiblesse de ces corrélations indique que le tableau des données n'a pu être ordonné en blocs diagonaux suffisamment contrastés. S'agissant d'un tableau de fréquences, ce premier constat appellera une attitude prudente lors de l'interprétation des proximités des modalités issues d'ensembles différents.

	Latine	Scientifique	Moderne	Générale	Pratique	Atelier	Niveaux-Options	Accueil
Homme	635	2094	271	1434	214	53	1134	65
Femme	1272	1052	971	1240	131	6	969	55
Origine Suisse	1381	2184	844	1531	181	24	1350	10
Origine Espagnole	100	214	84	309	44	11	140	15
Origine Italienne	168	332	159	553	86	7	266	14
Origine Française	78	119	57	130	17	8	108	––
Couche Supérieure	639	646	191	176	9	3	377	23
Couche Moyenne	803	1418	622	1043	113	25	1005	36
Couche Inférieure	424	1018	395	1394	213	28	656	48
Origine Autres Pays	180	297	98	151	17	9	239	81
Couche Divers S. I.	41	64	34	61	10	3	65	13

TABLEAU 8:
illustration empirique. Enseignement Public du Canton de Genève:
Cycles d'Orientation.
Nombre d'élèves selon la Section, le Sexe, l'Origine et la Couche Sociale
(Source: Annuaire Statistique de l'Education,
Département de l'Instruction Publique, Genève, 1982, p. 64).

En termes relatifs, la valeur propre associée à la dimension 1 représente 63,6 % de la somme de toutes les valeurs propres; ce taux, couramment appelé taux d'inertie, est de 32,8 % pour la dimension 2 et 2,7 % pour la dimension 3. Les deux premières dimensions expriment ainsi dans leur ensemble 96,4 % de l'inertie du nuage des points. Le taux d'inertie est un critère parmi les plus couramment employés dans le cas des tableaux de fréquences pour déterminer le nombre de dimensions à retenir. Ce taux peut être considéré comme satisfaisant et l'on renoncera donc à l'extraction d'une troisième dimension. Dans notre illustration toutefois, nous avons préféré présenter les résultats relatifs à cette dernière dimension, car elle offre des caractéristiques que l'on rencontre fréquemment dans la pratique pour des dimensions qui ont une valeur propre très faible (cas des tableaux logiques disjonctifs complets).

L'interprétation des facteurs de I et de J pour chaque dimension se fonde principalement sur l'étude des contributions, et en particulier des contributions des éléments ou modalités aux facteurs (contributions absolues; abrév.: CTR). Il convient tout d'abord de fixer une valeur minimale de la CTR au dessus de laquelle on considérera qu'une modalité fournit un sens appréciable au facteur. Nous savons que la CTR est exprimée en termes de pourcentage: la somme de toutes les CTR sur le facteur d'un *ensemble actif* est en effet égale à 100. Connaissant le nombre de modalités actives de chaque ensemble (I = 11–2 = 9 modalités; J = 8–2 = 6 modalités), nous pouvons calculer la valeur minimale que ce coefficient doit atteindre pour que l'apport de sens d'une modalité au facteur puisse faire l'objet d'une attention particulière. Cette valeur de la CTR sera ainsi, dans notre illustration, 11,2 % (c'est-à-dire 100/9) pour l'ensemble I, et 16,7 % (c'est-à-dire 100/6) pour l'ensemble J. Cela ne signifie cependant pas que l'interprétation d'un facteur se fonde exclusivement sur les modalités qui ont une très forte CTR. Des modalités qui ont une CTR inférieure à la valeur minimale ainsi calculée pourront être utilisées pour étayer une interprétation préalablement esquissée. Il s'agit en quelque sorte de procéder par approximations successives: on donne d'abord un sens et une interprétation d'un facteur à l'aide des modalités qui ont une CTR plus élevée que la valeur minimale; on ajuste et on complète ensuite cette interprétation en vérifiant que les positions du restant des modalités du même ensemble s'incorporent de manière cohérente dans l'interprétation fournie.

Nous voyons tableau 9 le résultat de l'AFC de notre tableau de fréquences.

Le facteur 1 de J est caractérisé par l'opposition des Sections "Latine" sur le pôle positif (CTR = 44,2 %) et "Générale" sur le pôle négatif (CTR = 26,8 %). Ce facteur semble exprimer un aspect de la hiérarchie de prestige qui lie les différentes Sections. Aucune autre CTR de cet ensemble

	FACTEUR 1				FACTEUR 2				FACTEUR 3			
	Ordre	Coordonnée	CTR (%)	COR (%)	Ordre	Coordonnée	CTR (%)	COR (%)	Ordre	Coordonnée	CTR (%)	COR (%)
ENSEMBLE I:												
Homme	8	-0,24	19,3	54,9	10	-0,22	30,7	45,1	7	-0,00	0,0	0,0
Femme	2	0,25	20,5	58,0	2	0,22	28,8	42,0	6	-0,00	0,0	0,0
Or. Suisse	4	0,09	3,2	67,0	8	-0,06	2,5	27,1	8	-0,02	3,3	2,9
Or. Espagnole	10	-0,28	4,2	73,3	3	0,14	2,1	18,8	3	0,05	3,4	2,5
Or. Italienne	9	-0,28	6,9	58,0	1	0,23	8,7	37,7	2	0,07	9,2	3,2
Or. Française	7	-0,07	0,1	11,2	6	0,08	0,3	14,5	5	0,02	0,1	0,4
Couche Sup.	1	0,48	25,8	71,0	11	-0,28	17,4	24,6	1	0,12	37,7	4,4
Couche Moy.	5	0,06	0,9	32,1	7	-0,02	0,2	3,9	11	-0,08	40,3	63,2
Couche Inf.	11	-0,28	19,0	79,1	4	0,14	9,2	19,8	4	0,03	6,0	1,1
Or. Autres Pays	3	0,13	0,9	31,8	9	-0,17	2,9	52,1	10	-0,04	2,1	3,1
Couche Divers-S.I.	6	-0,01	0,0	0,7	5	0,08	0,2	28,3	9	-0,03	0,3	3,9
ENSEMBLE J:												
Latine	1	0,34	44,2	96,4	5	-0,02	0,4	0,4	3	0,06	34,2	3,2
Scientifique	4	-0,06	1,9	6,7	8	-0,21	50,9	91,5	6	-0,03	11,3	1,7
Moderne	2	0,23	13,6	45,8	1	0,24	27,8	48,2	8	-0,08	41,6	5,9
Générale	6	-0,22	26,8	74,3	3	0,13	17,6	25,2	4	0,02	4,3	0,5
Pratique	7	-0,39	10,4	85,9	2	0,14	2,5	10,7	2	0,07	8,6	3,0
Atelier	8	-0,52	3,1	63,1	7	-0,20	0,9	8,9	5	-0,01	0,1	0,1
Niveaux-Options	3	0,01	0,1	1,4	6	-0,06	3,1	43,7	7	-0,05	24,6	28,1
Accueil	5	-0,16	0,5	9,4	4	0,04	0,1	0,6	1	0,22	21,3	18,2

TABLEAU 9:

illustration empirique; résultat de l'analyse factorielle des correspondances (Facteurs 1, 2 et 3). Les coefficients qui se rapportent aux éléments supplémentaires sont indiqués en italique. *Les CTR qui dépassent le seuil minimum sont entourées: l'interprétation des facteurs 'effectue principalement sur la base des modalités qui leur correspondent.*

sur le facteur 1 n'atteint la valeur minimale de 16,7 %. Si nous considérons alors la coordonnée de chacune des modalités actives restantes, nous constatons que la section "Moderne" s'oppose de manière isomorphe à "Pratique" et "Atelier" (de même qu'à la modalité passive "Accueil"). L'hypothèse que nous faisions sur le sens de ce facteur est renforcée par ce dernier constat. La lecture des contributions du facteur aux modalités (contributions relatives; abrév.: COR) indique en outre que les Sections "Latine" d'un côté, "Pratique" et "Générale" de l'autre, sont tout particulièrement caractéristiques de ce facteur. Prenons par exemple le cas de la Section "Latine"; la CTR de cette modalité est très élevée, et sa dispersion est expliquée presque intégralement par le facteur 1 (COR = 96,4 %). On peut dire que "Latine" est une modalité très caractéristique de cet axe (d'après sa COR) et que, de surcroît, elle fournit le sens principal du facteur (d'après sa CTR). Si l'on considère maintenant le facteur 3, on voit qu'à la même modalité est associée une faible COR et une forte CTR: "Latine" participe de nouveau à la construction du facteur, mais pour une très faible portion seulement de sa dispersion. Un tel rapport entre les deux contributions associées à une modalité, analogue à celui de la modalité "Moderne" sur ce même facteur, indique le peu d'importance relative de l'information dégagée par un facteur; il comporte une baisse notable de la fiabilité et de la stabilité de l'interprétation du facteur que le chercheur est à même de fournir. Nous avions par ailleurs déjà relevé la faiblesse du taux d'inertie de l'axe 3 (2,7 %) et proposé son abandon.

Les facteurs de I s'interprètent de manière analogue car, comme nous l'avons mentionné – cf supra – l'AFC accorde des rôles parfaitement symétriques aux deux ensembles I et J. Ainsi sur le facteur 1, les plus fortes CTR sont associées aux modalités Couche "Supérieure" et "Femme" sur le pôle positif, et Couche "Inférieure" et "Homme" sur le pôle négatif. On est tenté de conclure que l'échantillon analysé se compose principalement de filles de couche sociale favorisée et de garçons de couche sociale défavorisée. Or l'AFC n'autorise pas une telle conclusion. L'analyse opère un reclassement des lignes *et* des colonnes du tableau des données, et attribue une valeur numérique à chacune des modalités, dans le but de maximiser un coefficient de corrélation linéaire. L'organisation des modalités de l'ensemble I sur le facteur 1 n'a donc de sens que par rapport à l'opposition que nous avons décrite dans l'ensemble J (Sections prestigieuses et littéraires versus Sections moins prestigieuses et manuelles) sur le même facteur. La valeur propre associée au facteur 1 (λ = 0,0537) exprime alors la force que possède la correspondance entre l'ordre des Sections (ensemble J) et l'ordre des positions sociodémographiques (ensemble I).

Il apparaît en définitive que la lecture des résultats d'une AFC se fait avant tout par l'étude des contributions et de la coordonnée associées à chacune des modalités d'un ensemble sur chaque facteur. La lecture si-

simultanée de plusieurs ensembles sur une même dimension, ou des modalités supplémentaires d'un même ensemble sur un facteur, constituent probablement les apports les plus originaux et les plus féconds de cette méthode d'analyse, à la condition toutefois que les interprétations effectuées par le chercheur ne contredisent pas certaines propriétés mathématiques des facteurs calculés.

7. CONCLUSION

L'AFC a désormais acquis une légitimité certaine en sciences sociales. Dans les *Cahiers pour l'Analyse des Données* paraissent périodiquement depuis 1976 des textes fondamentaux sur la méthode, des études de cas dans différents domaines, ainsi que des programmes.

La pratique de l'analyse des données implique cependant beaucoup d'exercice. Un ensemble de données appelle souvent non pas une, mais plusieurs démarches statistiques qui peuvent être, selon les cas, complémentaires ou mutuellement exclusives. Il conviendra d'opérer des essais comparatifs à l'aide de méthodes différentes, ou d'adopter des stratégies différentes dans le cadre d'une seule méthode. Supposons par exemple de disposer des données relatives à une question à choix multiple comportant une dizaine d'items. Les réponses ont été recueillies auprès de sujets garçons et filles, peu et très scolarisés, d'origine sociale basse et élevée.

On peut, dans un premier temps, construire des tableaux de fréquences croisant chaque item avec chaque position sociodémographique des répondants. On aura 30 tableaux (2 x 2: deux lignes et deux colonnes) à partir desquels on calcule un test du χ^2 . Ce procédé ne permet toutefois pas l'étude systématique des *relations* entre les réponses. On peut alors effectuer une analyse factorielle de la matrice des corrélations des 10 réponses pour tous les sujets; après l'examen et éventuellement la rotation des facteurs retenus, on calculera des scores factoriels qui indiquent l'emplacement de chaque individu sur chaque facteur, et on effectuera une analyse de la variance de ces scores sur chaque facteur en prenant comme source de variation les positions sociodémographiques des repondants.

Ce procédé comporte deux étapes distinctes: l'analyse porte avant tout sur les colonnes du tableau (les items) et ensuite sur les lignes du même tableau (les sujets) (cf. pour les illustrations Deschamps, Lorenzi-Cioldi, Meyer, 1982). Ceci découle du fait que le modèle de l'analyse factorielle accorde des rôles asymétriques aux lignes et aux colonnes d'un tableau des données, rendant par là-même impossible la représentation simultanée des deux ensembles. Si l'on est intéressé par une telle représentation, on effectuera une AFCM sur le tableau logique qui comporte l'ensemble actif des réponses à la question, et l'ensemble passif des positions sociodémographiques des répondants; ou encore, on analysera à l'aide de cette méthode la table de contingence qui croise les six appartenances sociodémographiques avec les dix réponses. Il existe encore bien des possibilités offertes au chercheur (analyse discriminante, analyse typologique, multidimensional scaling, analyse de la régression multiple, etc.). Chacune de ces méthodes devrait être essayée, et les résultats systématiquement comparés. Une convergence de ceux-ci indique une bonne fiabilité de l'interprétation définitive.

BIBLIOGRAPHIE

BENZECRI, J.-P. (1973), "L'Analyse des Données"; Tome 1 : La Taxinomie; Tome 2 : L'Analyse des Correspondances (Dunod, Paris).

BENZECRI, J.–P. (1977), Histoire et préhistoire de l'analyse des données, *Cahiers pour l'Analyse de Données*, **2** / 1 (1977) 9–40.

BENZECRI, J.-P. et coll. (1980), "Pratique de l'Analyse des Données";
Volume 1 : Analyse des correspondances, exposé élémentaire (1980) (Dunod, Paris);
Volume 2 : Abrégé théorique : études de cas modèles (1980) (Dunod, Paris);
Volume 3 : Linguistique et lexicologie (1981) (Dunod, Paris).

BERGOUINOUX, A.; LAUNAY, M. F.; MOURIAUX, R.; SUEUR, J.-P. & TOURNIER, M. (1982), "La Parole Syndicale" (PUF, Paris).

BOUROCHE, J.–M. & SAPORTA, G. (1980), "L'Analyse des Données" (PUF, collection "Que sais-je ?", Paris).

DESCHAMPS, J.–C.; LORENZI–CIOLDI, F. & MEYER, G. (1982), "L'Echec Scolaire, Elève modèle ou Modèles d'élève" (P.–M. Favre, Lausanne).

FENELON, J.–P. (1981), "Qu'est-ce que l'Analyse des Données?" (Lefonen, Paris).

HOTELLING, H. (1933), Analysis of a Complex of Statistical Variables into Principal Components, *Journal of Educational Psychology*, **24** (1933) 417–441 et 498–520.

LEBART, L.; MORINEAU, A. & TABARD, N. (1977), "Techniques de la description statistique, méthodes et logiciels pour l'analyse des grands tableaux" (Dunod, Paris).

PEARSON, K. (1901), On Line and Planes of Closest Fit to a System of Points in Space, *Philosophical Magazin*, **2** (1901) 559–572.

RAO, C. R. (1952), "Advanced Statistical Methods in Biometric Research" (J. Wiley, New York).