

Analyses Factorielles et La Mal-Mesure de l'Homme

Joseph SAINT PIERRE,
Centre Interuniversitaire de Cacul de Toulouse
e-mail: Joseph.Saint-Pierre@cict.fr

Ecoles Doctorales
Vendredi 18 Avril 2003

Résumé

Cet exposé est une incitation à la lecture ou la relecture du livre de Stephen Jay GOULD et surtout aux passages consacrés à l'histoire de l'analyse factorielle. Le but est en partie de montrer les risques d'interprétation liés à cette méthode, mais aussi de montrer que cette méthode statistique est issue du domaine psychologique.

Pour débiter cet exposé je voudrais expliquer comment le sujet m'a été suggéré par les questions d'un utilisateur. Je me suis livré à une sorte d'exploration des divers messages électroniques que j'ai reçus durant l'année 2000. En fait j'ai regardé les diverses correspondances que j'ai entretenues durant une année et j'ai analysé celles ci, pour ainsi dire j'ai fait une analyse des correspondances. Ce jeu de mots n'est pas de moi, je l'ai entendu pour la première fois en septembre 1984 et en anglais. En fait ce jeu de mots est même plus pertinent en anglais car usuellement, en France la méthode statistique à laquelle fait référence le jeu de mots est désignée par Analyse Factorielle des Correspondances et abrégée en AFC. Il se trouve que les messages analysés sont des messages électroniques et qu'une des différences essentielles entre les messages traditionnels envoyés sous forme de lettres par la poste et les messages électroniques réside dans le rôle des facteurs, qui est essentiel dans la distribution des lettres et inexistant pour les messages électroniques. Le facteur est aussi d'ailleurs un mot qui a de multiples sens et est donc source de jeux de mots.

Voici le courrier que j'ai sélectionné, volontairement je n'ai pas laissé le nom de l'auteur, en fait il s'agit d'un message posté sur un forum usenet en

même temps qu'à une liste de de distribution, je ne suis pas abonné à la liste mais je lis régulièrement le forum (Newgroups). Je pense que l'auteur est sur la liste de distribution vu qu'il s'adresse aux membres de la liste. On peut noter que l'auteur appartient à l'Université d'Orléans.

Newsgroups: sci.stat.edu
Subject: Correspondence factor analysis (CFA)
Date: Sun, 31 Jan 2000 19:30:30 +0100
Organization: Université d'Orléans

Dear list members:

In order to check the results of a statistical software, does somebody can kindly send me the results of the Correspondence Factor Analysis of the following (6X6) table (Normally two main axis created by the first by r1-a and the second by r6-f) .

	a	b	c	d	e	f
r1	372	346	517	350	690	431
r2	207	414	621	413	828	512
r3	100	208	310	207	414	259
r4	180	345	520	345	690	431
r5	134	276	414	276	555	345
r6	212	418	621	410	828	317

Thanks for your help in advance. (Please e-mail me directly)

J'ai répondu au message posté dans le forum par un message directement adressé à l'auteur que voici sans le fichier attaché correspondant aux résultats.

Bonjour,

Je vous envoie l'AFC simple faite avec le logiciel SPAD.N sous forme d'un fichier attaché. En anglais l'AFC s'appelle CA pour correspondance analysis, CFA désigne habituellement l'analyse factorielle "confirmatoire", pour Confirmatory Factorial Analysis.

Je ne l'ai pas vérifié sur vos données mais en principe SPSS et SAS donnent exactement les mêmes résultats, si je vous l'ai fait avec SPAD c'est parce que j'ai vu que vous aviez une adresse en France...

Cordialement,

Joseph Saint Pierre,
Ingénieur Statisticien.

J'ai reçu la réponse suivante :

Merci beaucoup de votre mail, de votre promptitude et bien sur des résultats.

Ils confirment ceux que j'avais obtenus avec un logiciel qui est un add-on d'excel (XLSTAT) et dont je doutais un peu.

Merci aussi de l'information sur CFA/CA. En France on utilise AFC et dans nos documents en anglais on utilise CFA; je vais devoir changer cela.

Cordialement

Profitant assez souvent d'aides venant du réseau il m'arrive aussi de répondre à des utilisateurs du réseau demandant de l'aide en statistiques lorsque cela tombe dans mon domaine de compétence, car il me semble que c'est seulement comme cela que le réseau peut fonctionner correctement.

Bien sûr si j'ai choisi ce message c'est qu'il ressemble à d'autres demandes et qu'à partir de la discussion sur le nom même d'analyse factorielle peuvent

se greffer un certain nombre de réflexions sur les méthodes ayant un intérêt scientifique, à mon humble avis.

Il se trouve que l'expression analyse factorielle peut avoir divers sens et qu'elle offre une certaine ambiguïté. Lorsque j'ai commencé à travailler au CICT, en 1988, si un utilisateur demandait une analyse factorielle il s'agissait implicitement d'une analyse factorielle des correspondances et donc une méthode concernant des variables qualitatives, catégorielles, nominales (nationalité, CSP, etc. pour des questionnaires de sciences humaines). À cette époque là la pratique des statistiques était largement dominée, dans bien des secteurs, par l'école française dite d'Analyses de Données dont le fondateur était Jean-Paul Benzécri inventeur de l'analyse factorielle des correspondances. C'est autour des travaux de Benzécri qu'a été mise au point l'extension la plus populaire, dans les sciences humaines, de cette méthode, sous le nom d'analyse factorielle des correspondances multiples qui s'abrége en AFCM. Ces méthodes sont restées assez longtemps méconnues du monde en dehors de la France, elles étaient assez utilisées en France avec des logiciels spécifiques. Il est difficile de raconter toute l'histoire du passage de ces méthodes dans le reste du monde, mais c'est vers le début des années 1990 que ces méthodes ont été disponibles de façon standard dans la plupart des logiciels (BMDP, SAS ou SPSS). Mais voilà la méthode a changé de nom avec la traduction car le mot factoriel a disparu. Par ailleurs, il faut savoir qu'en anglais "factorial analysis" qui se traduit assez facilement par analyse factorielle désigne une méthode qui s'applique à des variables quantitatives, des scores, des variables de type "intervalle" ce qui veut dire que l'on doit supposer la continuité. L'analyse factorielle telle qu'elle est utilisée dans le reste du monde était par contre fort peu utilisée en France, tant et si bien que cette méthode est parfois appelée analyse en facteurs et cela pour faire clairement la distinction entre cette méthode et l'analyse factorielle des correspondances.

En ce qui concerne les utilisateurs de SPSS en France cela a provoqué un certain nombre de confusions, j'ai pu le constater avec les personnes de l'université du Mirail qui utilisent ce logiciel. Parmi les questions qui me sont le plus fréquemment posées, beaucoup sont relatives à la façon de faire des AFCM ou des ACP (Analyses en Composantes Principales) avec ce logiciel. Toutes ces méthodes sont relatives à un même problème qui consiste à réduire le nombre de nombres de variables dans une étude en s'appuyant sur les liaisons entre les variables initiales. Pour faire une ACP dans SPSS, il faut passer par le menu ou par la commande permettant de faire une analyse factorielle (au sens anglais, tout au long de l'exposé, j'utiliserai l'expression analyse factorielle pour cette méthode et non l'expression analyse en facteurs et j'expliquerai pourquoi.) Pour faire l'ACP il faut demander à ne pas faire de

rotation, depuis quelques années sur les versions interactives le défaut est de ne pas faire de rotation mais, si on utilise la syntaxe, la rotation dite varimax est la méthode par défaut. Le but de cet exposé n'est certainement pas de donner les différences mathématiques profondes entre ces diverses méthodes mais de donner une vision critique sur ces méthodes qui sont très souvent vues comme des boîtes noires faisant des opérations magiques sur les données.

J'ai souvent fait des remarques très critiques sur l'utilisation des méthodes d'analyses multivariées ou multidimensionnelles dans le traitement statistiques des données, mais ma position n'a pas grand chose de bien original. Le premier jour où j'ai commencé à travailler au CICT, c'était le 6 Juillet 1988, j'ai rencontré un utilisateur avec lequel je suis toujours resté en contact, cet utilisateur m'a parlé quelque temps après d'un livre qui s'appelle la Mal-Mesure de l'Homme par Stephen Jay Gould. L'utilisateur en question était un biologiste qui avait fait un DEA en anthropologie et faisait une thèse dans la même discipline, de par sa formation et de par son intérêt pour les statistiques j'ai compris qu'elle pouvait être sa proximité avec quelqu'un comme Stephen Jay Gould. Je n'ai pas lu ce livre qu'en 1995, d'abord en version originale puis en français. Ce livre décrit assez précisément mes positions sur l'utilisation des statistiques en sciences humaines et je ne cesse donc de conseiller sa lecture. Depuis cette lecture, le livre a été réédité, avec des nouveaux chapitres et surtout une explication sur l'histoire du livre, la seconde édition a été traduite en Français en 1997. La lecture du livre est intéressante pour aider à comprendre l'usage des statistiques mais pas seulement, cela permet de saisir l'origine des méthodes statistiques qui sont souvent issues d'autres disciplines que les mathématiques et cela permet d'entrevoir certains enjeux des démarches de recherche scientifique. Le livre reprend l'histoire de Sir Cyril Burt (1883-1971), le pionnier des tests mentaux, qui fut pendant 20 ans le psychologue officiel du London County Council, puis titulaire de la chaire de professeur de psychologie de l'University College de Londres de 1932 à 1950. Cette chaire avait été occupée par Charles Spearman et était considérée comme la plus prestigieuse de Grande Bretagne. En 1979 est parue une biographie de Burt par L.S. Hearnshaw, il y est démontré que Cyril Burt a triché pour faire ses études sur les vrais jumeaux (invention de sujets, de collaborateurs, d'articles etc...). La version américaine de la Mal-Mesure de l'Homme est parue en 1981 et reprend les "résultats" de l'enquête d'Hearnshaw. Burt a laissé un nom dans les méthodes statistiques et c'est justement dans la méthode qui s'appelle AFCM que l'on utilise un tableau qui s'appelle le tableau de Burt. L'AFCM n'est en fait qu'une AFC simple faite sur un tableau de Burt. Il est vraisemblable que Burt n'ait jamais entendu parler de l'AFCM, mais il est sûr que son nom n'a pas été une bonne publicité pour l'AFCM dans les pays où était connu Cyril Burt, le nom de Burt est associé à

celui de fraude scientifique majeure. J'ai moi même entendu, de la part d'anglais en 1984, des remarques sur le fait que le nom d'un tricheur anglais était associé à une méthode supposée purement française. Parmi ses mensonges scientifiques de Cyril Burt, celui s'est attribué le travail de Charles Spearman sur l'analyse factorielle. Les noms semblables des méthodes ainsi que la présence du nom de Burt dans l'histoire des deux méthodes ont contribué à ajouter à la confusion.

Les utilisateurs de l'analyse factorielle, telle qu'elle se fait dans les logiciels comme SAS, SPSS, BMDP ou autres peuvent prendre l'habitude de faire des rotations (varimax, oblimax etc.), mais surtout il faut savoir que les rotations sont apparues dans un contexte d'une forte critique de l'analyse définie par Spearman. Un des principaux artisans de cette critique fut Louis Leon Thurstone né en 1887, professeur de psychologie à Chicago de 1924 à sa mort en 1955. Les utilisateurs qui ont surtout fait de l'AFCM, ont rarement fait des rotations car celles ci ne se font pas facilement dans les logiciels pour cette méthode. Quelqu'un qui aura fait beaucoup d'analyses factorielles avec des rotations (varimax ou autres) aura la tentation de faire des rotations en voyant certaines sorties d'AFCM. Ce genre de statisticien, ou plutôt d'utilisateur des statistiques ne se rencontre pas trop en France.

À travers une note de Stephen Jay Gould on peut saisir l'intérêt de sa démarche et essayer de percevoir le décalage de la culture statistique française par rapport au reste du monde : « Les lecteurs qui ont abordé l'analyse factorielle à l'occasion d'un cours sur les statistiques ou la méthodologie appliquées aux sciences biologiques ou sociales (ce qui est très fréquent à notre époque où l'usage de l'ordinateur est si répandu), se souviendront peut être de la rotation des axes en position varimax. Comme à moi, on leur a sans doute enseigné ce procédé comme s'il s'agissait d'une déduction mathématique élaborée pour répondre au problème posé par les difficultés que l'on rencontre pour isoler des faisceaux à l'aide des composantes principales. En fait, cette méthode est apparue dans un contexte historique précis, à savoir le conflit opposant deux théories de l'intelligence, celle de Thurstone qui croyait à des aptitudes mentales primaires et celle de Spearman et Burt, partisans d'un système hiérarchisé (intelligence générale et facteurs subsidiaires). »

Cette analyse est très intéressante car elle permet de désacraliser les mathématiques mais surtout elle permet de bien comprendre que certaines méthodes mathématiques, telles que l'analyse factorielle, sont nées des sciences humaines, même si l'essentiel des outils sont issus des sciences physiques.

Pour saisir la critique essentielle de Stephen Jay Gould par rapport à l'utilisation des méthodes d'analyses factorielles, je vais me servir d'un exemple de données purement fictif qui a servi très souvent dans des cours ou des séminaires sur ces méthodes. À titre d'exemple ces données ont été utilisées dans la conférence du 15 Mai 1998, faite dans le cadre de l'IED, par Louis Ferré et Fred Ferraty, dans le cadre de cette conférence une présentation détaillée de la méthode a été donnée et cet exposé n'est absolument pas une reprise. Il s'agit de notes obtenues par 9 élèves dans 5 disciplines, Mathématiques, Sciences, Français, Latin, Travaux Manuels.

JE	6.0	6.0	5.0	5.5	8.0
AL	8.0	8.0	8.0	8.0	9.0
AN	6.0	7.0	11.0	9.5	11.0
MO	14.5	14.5	15.5	15.0	8.0
DI	14.0	14.0	12.0	12.5	10.0
AD	11.0	10.0	5.5	7.0	13.0
PI	5.5	7.0	14.0	11.5	10.0
BR	13.0	12.5	8.5	9.5	12.0
EV	9.0	9.5	12.5	12.0	18.0

Comme il s'agit de données fictives, il y a beaucoup de questions qui peuvent, et à mon avis qui doivent, se poser dans le cadre d'une recherche appliquée qui ne se posent pas réellement ici. Par exemple on peut se demander si les notes ont été construites à partir de plusieurs épreuves ou d'une seule, si il y a eu un seul correcteur ou plusieurs. On pourrait aussi se demander comment fonctionne l'établissement dont sont issus les élèves et quelle est la réussite des élèves issus de cet établissements à des examens comme le baccalauréat. Savoir replacer les données statistiques dans le contexte de leur constitution, de leur recueil, de leur codage etc. est une des tâches essentielles du statisticien appliqué. Il y a une autre question sous jacente qui est très importante et qui est souvent masquée : "Quel est le but de l'étude statistique?". Faire une étude statistique demande un certain travail et un certain temps, voire une toute petite compétence, il faut souvent savoir pourquoi on lance une étude qui peut avoir un certain coût. Le fichier fictif, a pour mission essentielle d'apprendre des méthodes statistiques, mais il faut alors se demander pourquoi on utilise telle méthode plutôt que telle autre.

Sur ce petit tableau de données je fais une ACP, en demandant l'extraction de 3 facteurs. Voici les résultats obtenus avec SPSS, en fait, choisir 3 fac-

teurs correspond à un critère raisonnable, éboulis des valeurs propres. La qualité de la représentation est extraordinairement bonne, on peut légitimement penser que les données ont été construites pour cela. On est dans une situation où l'on avait 5 variables observées et où l'on réduit à 3 dimensions. Les facteurs sont des variables construites par la méthode d'ACP qui ont la capacité de résumer au mieux les variables de départ. Le premier facteur est la variable qui synthétiserait le mieux l'ensemble des 5 variables de départ.

Variable	Communalité	* *	Facteur	Valeur propre	% Variance	% Cum
M	1.00000	*	1	2.87823	57.6	57.6
S	1.00000	*	2	1.13481	22.7	80.3
F	1.00000	*	3	.98361	19.7	99.9
L	1.00000	*	4	.00237	.0	100.0
DM	1.00000	*	5	.00098	.0	100.0

	Facteur 1	Facteur 2	Facteur 3
M	.81035	-.56321	.15985
S	.90089	-.42176	.09793
F	.75318	.61569	-.23005
L	.91585	.38395	-.11255
DM	.06236	.33651	.93961

Qualité de la représentation sur les 3 premiers facteurs:

Variable	Communalité
M	.99942
S	.99907
F	.99929
L	.99887
DM	1.00000

De façon naïve ou classique pour synthétiser 5 notes ce que l'on fait c'est de calculer la moyenne ou le total ce qui est absolument équivalent. C'est ce qui fait dès le cours préparatoire dans toutes les écoles et qui se prolonge jusqu'à l'université, cela est facile à faire dans ce cas là. Une idée assez simple consiste à comparer la moyenne des notes et le premier facteur calculé par l'ACP. Pour cela je me suis contenté de calculer le coefficient de corrélation entre la moyenne des 5 notes et le score factoriel obtenu par l'ACP. La valeur est de 0,9703, ce qui même pour 9 observations est exception-

nellement proche de 1. Les deux quantités sont pratiquement équivalentes, ce phénomène est classique en ACP, il est appelé facteur taille, ou effet taille. Stephen Jay Gould, donne un exemple de ce phénomène à partir de longueurs d'os chez des oiseaux, le nom de facteur taille provient d'ailleurs de statistiques biologiques. Dans ce fichier de données le facteur taille est important mais pas tant que dans certaines autres études.

Le point très important à noter c'est que calculer la moyenne entre 5 notes peut être fait par élève de l'école primaire, alors que le calcul du premier facteur d'une ACP se fait avec un outil mathématique, la décomposition en valeurs propres d'une matrice, qui s'enseigne en deuxième année d'études supérieures scientifiques (mathématiques spéciales ou DEUG scientifique par exemple), et que par ailleurs il est pratiquement toujours nécessaire d'utiliser un ordinateur pour faire cela.

Pour comprendre l'ACP il faut savoir que celle que nous avons faite est fondée sur la matrice de corrélation entre les diverses variables. Voici donc cette matrice :

	M	S	F	L	DM
M	1.0000	.9825	.2267	.5081	.0112
S	.9825	1.0000	.3967	.6515	.0063
F	.2267	.3967	1.0000	.9512	.0380
L	.5081	.6515	.9512	1.0000	.0805
DM	.0112	.0063	.0380	.0805	1.0000

Il est facile de constater qu'il y a deux coefficients de corrélation qui sont extrêmement forts, la corrélation entre Mathématiques et Sciences vaut 0,9825 et la corrélation entre Français et Latin vaut 0,9512. Cela doit pouvoir en principe s'interpréter, il se trouve que pour certaines raisons ces deux corrélations ne choquent en principe personne. Cela est un des thèmes principaux de la Mal-Mesure de l'Homme, auquel j'attache le plus d'importance, une corrélation ne doit pas s'interpréter uniquement en fonction de la valeur du coefficient de corrélation mais surtout par le sens concret de la liaison entre les deux concepts associés. La corrélation entre deux phénomènes peut être importante, sans que cela ait un sens interprétable, Stephen Jay Gould prend comme exemple la corrélation entre son âge et le prix de l'essence ou entre la population de Mexico et les distances entre galaxies. La plupart des corrélations observables ne sont pas explicables, même si elles peuvent donner des idées de recherche. En l'occurrence dans l'étude fictive qui nous intéresse, la corrélation entre Mathématiques et Sciences obéit au sens commun, comme quoi les mathématiques sont un outil indispensable aux autres savoirs scientifiques et il semble normal, quoique discutable, que la capacité à maîtriser un des savoirs aide à acquérir l'autre. On peut par exemple mettre deux types de relations entre les mathématiques et les sciences qui sont différentes, on peut par exemple considérer que la connaissance des mathématiques aide à apprendre les autres sciences et donc qu'un bon niveau en mathématique permette d'avoir des bonnes notes en sciences, mais on peut considérer qu'un intérêt pour les sciences et une réussite dans cette discipline incite à travailler les mathématiques et à y réussir. En tout cas il est difficile d'établir une relation de causalité à partir de la valeur d'un coefficient de corrélation linéaire entre deux mesures, la causalité doit être

justifiée pour des raisons extérieures au simple calcul statistique, il s'agit là d'un des principes fondamentaux exposés dans la Mal-Mesure de l'Homme.

L'observation de la matrice de corrélation, étayée par une connaissance empirique du système éducatif, nous incite à penser que le Latin et le Français ont des points communs, il se trouve que très souvent le professeur de Latin est souvent professeur de Français (parfois aussi de Grec), il serait donc parfaitement normal, sans passer par une méthode statistique sophistiquée de créer une note de Lettres en ajoutant les notes de ces deux disciplines. Il faut remarquer qu'une note de Français peut comprendre des sous disciplines différentes, Orthographe, Grammaire, Récitation, Dissertation et que l'on met une note globale pour mesurer des compétences qui peuvent être différentes. Rien ne permet de prouver que la corrélation entre une épreuve d'Orthographe et une épreuve de Dissertation Française soit forte. Ce type de raisonnement peut aussi s'appliquer aux Sciences qui peuvent regrouper de la Physique, de la Chimie, de la Technologie, ou des Sciences Naturelles et même aux Mathématiques qui peuvent regrouper de la Géométrie, de l'Arithmétique ou de la Logique. Il peut être intéressant de noter que le regroupement des disciplines peut parfois être sujet à discussion, la Géographie peut être considérée comme associée à l'Histoire et parfois associée à la Géologie, la Logique faisait souvent partie des cours de Philosophie avant d'être, avec l'avènement de l'Informatique, souvent rattachée aux mathématiques, dans l'enseignement du Moyen-Âge la musique faisait partie du même domaine que l'Arithmétique (le Quadrivium).

Le deuxième facteur de l'ACP, oppose les disciplines littéraires des disciplines scientifiques si on considère qu'il existe un décalage entre les savoirs scientifiques et les savoirs littéraires il peut être légitime de calculer la valeur de ce décalage à partir des notes obtenues dans les quatre disciplines, pour cela dans un premier temps on calcule la note Lettres en ajoutant la note de Français et la note de Latin, puis on calcule une nouvelle note Sciences en ajoutant la note de Mathématiques à celle de Sciences :

Lettres= Latin + Français
Sciences= Sciences + Mathématiques

Le décalage entre le niveau littéraire et le niveau scientifique peut être évalué en calculant la différence entre la note de Lettres et celle de Sciences.

Décalage= Lettres - Sciences

On peut essayer de calculer la corrélation entre ce décalage et le deuxième facteur de l'ACP cela donne une valeur de 0,9558 ce qui est extrêmement fort.

Le calcul de ce décalage est sans doute un peu plus compliqué que la somme totale mais ne fait intervenir que les signes plus et moins et cela est beaucoup plus simple que le calcul du deuxième facteur de l'ACP. Si il est possible d'interpréter facilement le deuxième facteur de l'ACP, c'est bien parce que la différence entre les savoirs scientifiques et littéraire est socialement acceptée. Il y a, en France, des baccalauréats littéraires et des baccalauréats supposés scientifiques, il y a aussi des universités littéraires et des universités scientifiques. Ce deuxième axe ou facteur, correspond géographiquement à la différence entre Le Mirail et Ranguel. C'est parce que cette différence a un sens usuel que ce deuxième facteur est facilement interprétable.

Bien évidemment, le livre de Stephen Jay Gould est un plaidoyer pour les méthodes utilisant des rotations on va donc voir sur ce petit exemple ce que donne la rotation varimax en demandant 3 facteurs.

VARIMAX a convergé en 4 itérations.

Rotated Factor Matrix:

	Facteur 1	Facteur 2	Facteur 3
M	.99107	.13099	.00638
S	.95176	.30531	-.00392
F	.09731	.99487	.00710
L	.39078	.91841	.05183
DM	.00077	.03098	.99952

L'intérêt de la méthode est évident, compte tenu des interprétations précédemment données. En effet, les nouvelles variables Sciences et Lettres ont déjà été calculées et il est donc possible de calculer les corrélations entre les facteurs de l'analyse factorielle avec rotation varimax et ces nouvelles variables. La corrélation entre le premier facteur et la note globale scientifique vaut 0,9769, la corrélation entre le deuxième facteur et la note globale littéraire vaut 0,9725, la corrélation entre le troisième facteur et la note de travaux manuels est de 0,9995. L'interprétation des axes factoriels devient plus facile, le premier facteur est le facteur scientifique, le deuxième facteur est le facteur littéraire et le troisième facteur est le facteur artistique. De manière encore plus prudente on pourrait dire que le premier facteur représente les Mathématiques et les sciences, le deuxième facteur représente le Français et le Latin et le troisième les Travaux Manuels. C'est un des principes essentiels exposés par Stephen Jay Gould pour privilégier les méthodes avec rotation à l'ACP.

Stephen Jay Gould dénonce le danger de la réification des objets mathématiques et dans la Mal-Mesure de l'Homme il développe une critique de l'analyse factorielle autour du danger de la possible réification des facteurs. L'exemple historique étudié très important est le g de Spearman, la réification consiste

a associer systématiquement un objet réel à un objet créé par une méthode théorique. Pour essayer de percevoir ce concept, on peut utiliser l'exemple des notes des élèves, sachant que dans l'analyse on a trouvé un facteur, avec la rotation varimax, qui regroupait les disciplines littéraires, la réification consiste à dire qu'il existe dans un premier temps un facteur littéraire, puis éventuellement un esprit littéraire, enfin des dispositions dans les convolutions du cerveau qui prédisposent aux compétences littéraires etc. Le phénomène de réification existe de manière populaire dans l'expression "avoir la bosse des mathématiques". Mettre une liaison entre les parties du cerveau et les compétences intellectuelles a été un des fantasmes du dix-neuvième siècle qui a pris une nouvelle ampleur au vingtième avec notamment Spearman. Toute cette histoire est très bien racontée dans la *Mal-Mesure de l'Homme*.

On peut toutefois noter que la démarche scientifique consiste souvent à réifier des objets qui n'ont pas d'existence tangible au sens ordinaire du terme. La Terre a un centre de gravité, et elle tourne sur elle même autour d'un axe d'inertie, en 24 heures. Cette axe n'est pas visible, il n'y a pas une tige passant par le pôle nord et le sud, les points correspondant aux pôles sont déterminés par une démarche calculatoire. Lorsqu'on représente la Terre sur une mappemonde on utilise une tige pour pouvoir faire tourner la représentation de la Terre sur elle même, cette tige est une sorte de réification de l'axe de rotation de la Terre. Il se trouve que les calculs mathématiques qui sont utilisés pour déterminer les axes de dans une ACP sont exactement les mêmes que ceux utilisés pour déterminer les axes d'inertie d'un solide physique de même que la moyenne statistique se calcule exactement de la même manière que le centre de gravité (ou d'inertie) d'un solide. Il ne faudrait pas pas considérer le livre de Stephen Jay Gould comme une critique systématique de la démarche scientifique, ni surtout utiliser ce livre pour dénigrer l'utilisation des mathématiques dans les autres sciences. En effet, il y a dans la *Mal-Mesure de l'Homme* énormément d'éléments quantitatifs simples et des idées intéressantes sur les statistiques. Il y a beaucoup d'autres réflexions sur les statistiques dans les autres livres de Stephen Jay Gould dont je recommande vivement la lecture.

Bien évidemment un tel exposé ne peut aborder tous les points intéressants d'un tel livre ni surtout s'y substituer et il ne peut être considéré que comme une incitation à sa lecture.