Statistique et analyse des données

Roger Lafosse

Ressemblance et différence entre deux tableaux totalement appariés

Statistique et analyse des données, tome 14, n° 2 (1989), p. 1-24.

<http://www.numdam.org/item?id=SAD_1989__14_2_1_0>

Statistique et Analyse des Données
1989 - Vol. 14 n° 2 - p. 1 - 24

RESSEMBLANCE ET DIFFERENCE ENTRE DEUX TABLEAUX TOTALEMENT APPARIES

Roger LAFOSSE

Laboratoire de Statistique et Probabilités
U.A. 745
Université Paul Sabatier - 31062 TOULOUSE CEDEX

Résumé : *Une étude de ressemblance entre deux tableaux totalement appariés, à la fois quant aux individus et quant aux variables, est menée à l'aide de l'analyse interbatterie de Tucker, de la rotation procruste orthogonale, de changements d'échelles et d'un indice d'association. Par la suite, tout en introduisant des variables synthétiques expliquées dans un tableau par les variables d'un autre tableau, on définit une étude de différence étroitement liée à l'analyse précédente. On mesure ainsi l'effet produit sur la ressemblance par suppression de la contrainte d'appariement des variables. Ce travail permet, dans un contexte homogène, une présentation d'analyses connues et une description de propriétés de l'indice utilisé.*

Abstract : *We first investigate the agreement between two paired tables (obtained by measuring twice the same variables on the same cases), by means of Tucker's interbattery analysis, of orthogonal procrustes rotation, of change of scalings and of association index. Then, a study of the similarity is proposed. Introducing artificial variables explained in a table by the variables of the other table, we can derive a study of the dissimilarity between the two tables. Here we attempt to assess the effect of the matching of the variables. This leads to a unified presentation of some known analyses and to the description of the properties of the considered association index.*

Mots clefs : Analyses, en composantes principales, canonique, interbatterie, procruste, et de redondance.

Indices STMA : 06 : 010 , 07 : 020

Manuscrit reçu le 2 mars 1989, révisé le 28 août 1989

1 - INTRODUCTION

Les deux tableaux soumis à l'analyse se correspondent dans un appariement des lignes et aussi des colonnes. Ils pourraient être relatifs à un même ensemble de mesures effectuées à deux reprises sur les mêmes individus; ainsi deux ensembles de notes attribuées par deux juges à un groupe d'élèves ayant subi les mêmes épreuves, constituent un exemple souvent repris par la suite.

Dans une première partie, on s'intéresse à ce que les deux tableaux de variables peuvent avoir en commun, sans réduire le problème à une comparaison d'espaces vectoriels respectifs engendrés, c'est à dire en proposant autre chose que l'analyse canonique classique. En effet, vu l'appariement, le discours sur la ressemblance des deux tableaux peut être détaillé et enrichi, comme peut l'être celui concernant deux objets de même nature.
Nous expliquons pourquoi un traitement préliminaire des tableaux est souhaitable, en vue de rendre la comparaison abordable. Le centrage des variables, la rotation procruste, un changement d'échelle, et l'extraction de la redondance, sont les opérations décrites pour ce faire. Nous intégrons ainsi à notre démarche l'ajustement de Schönemann et Carroll (1970); cependant un autre changement d'échelles se justifie chez nous, les tableaux en présence étant par ailleurs plus particuliers. Après ce traitement, une méthode comparative permettant de proposer des résumés successifs, est introduite à partir de l'analyse interbatterie de Tucker (1958), qui nous semble dans ces conditions un outil bien adapté à l'étude d'une ressemblance. L'analyse alors définie est nommée analyse de communité.

Le travail précédent est effectué en prenant pour support des couples de variables principales caractéristiques d'une communité des deux tableaux. Il est repris dans une deuxième partie, mais cette fois en prenant pour support des couples de variables synthétiques appropriés à la recherche d'un ajustement. On aborde ainsi une notion de différence entre les deux tableaux, différence introduite par rapport à l'analyse précédente en supprimant l'appariement sur les variables. Les nouveaux couples sont définis à partir de variables principales successivement les mieux expliquées dans un tableau par les variables de l'autre tableau. Il s'agit ici d'une nouvelle présentation de l'analyse de Johansson (1981) reprise par Tyler (1982). Les variables explicatives, instrumentales, de l'A.C.P.V.I. de Rao (1964) -- confer aussi à ce sujet le travail de synthèse de Sabatier (1983) -- apparaissent comme sous-produit de notre analyse, alors que c'est plutôt l'inverse chez Johansson et Tyler. Cette présentation conduit à l'obtention du modèle d'ajustement considéré par Lebart, Morineau et Fénelon (1979) dans leur analyse des covariances partielles.

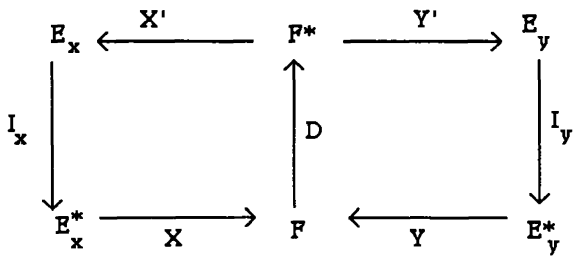
Dans un même temps, nous montrons combien l'indice d'association de Lingoes et Shönemann (1974), est adapté à une mesure de l'intensité des ajustements proposés, ou de l'évaluation globale de la communité, ou de celle de la redondance d'un tableau par rapport à l'autre. Certaines étapes de la démarche suivie dans notre exposé sont construites à partir d'une optimisation de cet indice.
Dans un article précédent (Lafosse (1985)), l'étude associée à deux tableaux portait sur la comparaison de deux nuages d'individus appariés, en usant de métriques éventuellement différentes pour les représenter. Cette fois, il s'agit de comparer deux tableaux de variables, les variables étant également appariées, du moins au départ, ce qui induit un autre discours et de nouveaux développements.
Le dernier paragraphe fournit les résultats obtenus sur un exemple d'école.

2 - PRESENTATION DES TABLEAUX ET NOTATIONS

Notre étude se limite à un ensemble de mesures effectuées sur les mêmes individus, à deux reprises, pour donner un tableau X et un tableau Y.
Les n lignes de X forment l'ensemble des individus $\{x_i\}$, i=1,...,n, vecteurs de l'espace engendré E_X, formant un nuage N_X de points (de $\mathbf{R}^p$ muni de la métrique identité) . Celles de Y forment l'ensemble des individus $\{y_i\}$, i=1,...,n, vecteurs de E_y, formant un nuage N_y de $\mathbf{R}^p$. Un poids p_i est affecté à l'individu i dans ses deux représentations x_i et y_i , avec $\Sigma p_i = 1$

Les p colonnes de X et celles de Y forment deux ensembles de variables centrées, à partir desquels sont engendrés les espaces respectifs F_X et F_y. L'espace $F = F_X + F_y$ est muni de la métrique diagonale D des poids p_i. Les variables peuvent être non réduites. Une j-ème colonne de X et une j-ème colonne de Y constituent deux représentations de la même variable.

Toute application considérée dans l'exposé, l'est en accord avec le schéma de dualité suivant (Cazes,1977 et Cailliez et Pagès, 1976) :



Le choix des métriques identité permet de simplifier l'exposé, sans empêcher des généralisations éventuelles liées à d'autres choix.
Avec le même objectif, nous supposons les applications X et Y injectives.

On note les matrices de covariances : $V_X = X'DX$ et $V_Y = Y'DY$
et celle des intercovariances : $V_{XY} = X'DY$ ou $V_{YX} = (V_{XY})'$
(la transposée d'une matrice A est notée A').

Matriciellement, un vecteur noté u correspond à un vecteur colonne et, noté u', à un vecteur ligne.
La somme étendue à l'indice i courant de 1 à n est notée, par exemple, $\sum_{1,n} p_i$.

3 - ANALYSE DE LA RESSEMBLANCE

3.1 Traitement préliminaire des tableaux

Deux objets sont plus aisément comparables si, mentalement, on a d'abord cherché à les superposer au mieux, de façon à faire coincider les traits les plus caractéristiques de leur ressemblance. C'est ce que nous avons commencé avec le centrage des variables, ce qui revient à donner aux nuages N_X et N_Y le même centre de gravité. C'est ce que nous poursuivons ci-dessous avec un changement d'échelles et le choix d'un repère commun.
Ainsi les notes données par deux juges à n élèves ayant réalisé chacun p travaux, sont ramenées à un même système de référence, par exemple celui de l'un des deux juges, avant comparaison. Les différences de notations correspondront alors à une façon différente de positionner les élèves les uns par rapport aux autres dans chacune des matières.

3.1.1 Changements d'échelles

En analyse en composantes principales (A.C.P.), il est parfois souhaitable d'opérer auparavant une standardisation des variables. Nous avons ici ce même problème et, par exemple, on peut homogénéiser la façon de noter les différentes matières pour le premier juge, en ramenant à l'unité l'étendue observée pour chacune. La même opération pourrait être effectuée sur le tableau associé au deuxième juge, mais le choix d'une métrique dans R^p ne peut être considéré de façon indépendante pour les deux juges.
En effet, dans une étude comparative, nous décidons aussi que le standard, la référence, est constitué par l'un des tableaux et que l'objet de l'étude consiste à voir en quoi l'autre tableau s'y conforme ou non. Effectuer des changements d'échelle sur l'un des tableaux, qui seraient

différents de ceux appliqués à l'autre tableau, sans se préoccuper des liaisons entre variables, et modifier alors le résultat d'une analyse de ressemblance, ne nous semble pas correspondre à une démarche suffisamment objective.

C'est pourquoi, soit aucun des deux tableaux ne subit de changement d'échelle avant analyse, soit chacun des deux subit exactement le même traitement : lorsqu'on éprouvera le besoin de réduire toutes les variables d'un tableau pris pour référence, on opérera exactement les mêmes changements d'échelles sur l'autre tableau, en divisant chacune des variables correspondantes dans l'appariement par les écarts-type respectifs calculés sur le tableau référence. De la sorte, les variables du second tableau ne sont pas réduites, et les deux tableaux ont subi exactement la même transformation; on use donc d'une même métrique dans R^p relativement aux deux juges.

Nous reviendrons sur cette notion de changement d'échelles dans le § 4. Pour le moment, nous adoptons celui venant d'être décrit, *X et Y désignant désormais les tableaux alors obtenus*. Nous supposons que *c'est le tableau X qui est adopté comme référence, et qu'il conservera ce rôle .*

3.1.2 Choix d'un repère commun

Les deux triplets statistiques en présence sont les triplets (X,I,D) et (Y,I,D), où Y est le tableau déjà éventuellement modifié en prenant X pour référence.

Souvent décrite comme une recherche d'un "maximal agreement" (par exemple, ten Berge et Knol (1984)), sa justification devenant claire dans les paragraphes suivants, la méthode utilisée pour ramener le tableau Y à un repère commun à X est celle de la rotation procruste. Sans revenir sur la définition de cette rotation, nous rappelons ici le calcul du système orthonormé $\{v_j\}$ de E_Y superposé par la rotation au système orthonormé $\{u_j\}$ de E_X, j=1...r, quand X est choisi pour référence, pour cible. En adoptant les notations de Golub et Van Loan (1983), cette rotation s'écrit

$$R = \sum_{1,r} v_j u'_j = V_{yx} (V_{xy} V_{yx})^{-1/2} \qquad r \text{ étant le rang de } V_{xy}.$$

Ce sont ces vecteurs u_j et v_j qui interviennent dans la décomposition en valeurs singulières de V_{xy} : $V_{xy} = \sum_{1,r} \lambda_j u_j v_j'$, décomposition calculée en général à partir des équations :

$$V_{xy}V_{yx}\, u_j = \lambda_j^2 u_j , \quad \|u_j\|=1 \quad \text{et} \quad \lambda_j v_j = V_{yx} u_j , \quad v_j \neq 0$$

ou bien encore à partir des équations :

$$V_{yx}V_{xy}\, v_j = \lambda_j^2 v_j , \quad \|v_j\|=1 \quad \text{et} \quad \lambda_j u_j = V_{xy} v_j , \quad u_j \neq 0$$

Nous rangeons les valeurs singulières λ_j strictement positives par ordre décroissant, j=1...r.

Le problème de la rotation procruste est à l'origine celui de la recherche globale de R avec $R'R=I_p$ et l'ordre des valeurs propres n'est pas important. Ici la rotation est partielle et l'ordre de ces valeurs importe pour la suite.
Dans E_X, le repère commun adopté pour chacun des deux tableaux X et YR est le repère $\{u_j\}$, r-dimensionné. La comparaison des deux tableaux initiaux ainsi préparés peut alors commencer.

3.2 Analyse interbatterie de Tucker

L'analyse interbatterie de Tucker a été proposée pour décomposer l'information commune à deux batteries différentes de tests réalisés sur de mêmes individus. Nous soumettrons à cette analyse les triplets statistiques (X,I,D) et (YR,I,D), pouvant être considérés comme relatifs à une même batterie de tests effectués à deux reprises. Mais dans un premier temps, pour le rappel ci-dessous, le deuxième triplet considéré est (Y,I,D).

3.2.1 Définition de l'analyse interbatterie :

Dans cette analyse factorielle, on définit des couples successifs de variables principales, en réalisant un compromis entre la notion de variance élevée de chacune des deux variables constituant un couple pour offrir de bons résumés successifs, et la notion de corrélation élevée entre ces deux variables pour qu'elles présentent une grande communauté.

Ainsi, à l'étape j, le couple (ξ_j,η_j) obtenu est solution de

$$\max_{\xi,\eta} \operatorname{cov}(\xi,\eta) = \max_{\xi,\eta} \xi' D \eta$$

sous les contraintes de normes des vecteurs $u_j \in E_X$ et $v_j \in E_Y$, sachant que : $\xi_j = Xu_j$ et que $\eta_j = Yv_j$.
De telles contraintes de norme sont introduites en A.C.P. d'un tableau et ont ici la même justification, en permettant de garantir une certaine homogénéité de la variance de chacune des variables principales, avec l'ensemble des variances des variables qui lui sont corrélées dans ce tableau.

Lors du calcul d'un j-ème couple, la novation est introduite en imposant les contraintes d'orthogonalités de u_j et de v_j avec les vecteurs des systèmes respectifs $\{u_i\}$ et $\{v_i\}$, i=1,...,j-1. Comme en A.C.P., chacun des systèmes de variables $\{\xi_j\}$ et $\{\eta_j\}$ correspond ainsi, par leurs variances, à un début de décomposition des variances totales respectives des tableaux X et Y, même si ces systèmes ne sont pas formés de variables non corrélées.
En fait, une fois circonscrite une part essentielle de communité entre deux entités, s'il faut extraire une nouvelle part de communité, nous trouvons plus judicieux de nous intéresser d'abord à ce que l'une des entités peut avoir de nouvellement définissable par rapport à l'autre,

que de nous intéresser à une nouveauté définie par rapport à elle-même. Or c'est bien ce qui est vérifié, puisque Tucker montre que l' orthogonalité de chacun des deux systèmes $\{u_j\}$ et $\{v_j\}$ implique que ξ_i et η_h sont non corrélées pour tout i différent de h.

Les r couples (ξ_j,η_j) successifs obtenus dans cette analyse sont les solutions associées aux r optima successifs $\lambda_j = \text{cov}(\xi_j,\eta_j)$ et aux r couples (u_j,v_j), qui ont été définis en 3.1.2.

3.2.2 Analyse interbatterie de X et YR

Nous soumettons à l'analyse précédente les triplets (X,I,D) et (YR,I,D), ce dernier ayant été conditionné pour réaliser la comparaison dans de bonnes conditions. On a alors les résultats suivants:

Propriété 1

La matrice des intercovariances de X et de YR est symétrique, et sa décomposition en valeurs singulières est décomposition spectrale de $(V_{xy}V_{yx})^{1/2}$.

En effet, $X'D(YR) = V_{xy}R = (\sum_{1,r} \lambda_j u_j v_j')(\sum_{1,r} v_j u_j') = \sum_{1,r} \lambda_j u_j u_j'$

ou bien encore $V_{xy}R = V_{xy}[V_{yx}(V_{xy}V_{yx})^{-1/2}] = (V_{xy}V_{yx})^{1/2}$

En conséquence, $\xi_j = Xu_j$, $\eta_j = YRu_j$ et un couple (ξ_j,η_j) obtenu précédemment dans l'analyse interbatterie des tableaux X et Y est, dans cette analyse, associé à une <u>seule</u> forme linéaire u_j.

Au départ, une variable en i-ème position dans le tableau X est appariée avec celle de Y, un même sens étant rattaché à ces deux variables. En reprenant notre exemple, ces deux variables désignent la même matière. Une rotation, telle la rotation procruste, qui correspond partiellement à un changement de base, ne modifie pas l'appariement : la i-ème colonne de YR désigne la même matière que la i-ème colonne de X. Par suite, les deux variables ξ_j et η_j, associées à la même combinaison linéaire u_j des variables de leurs tableaux respectifs , sont, bien que de corrélation différente de 1, deux expressions de la même variable synthétique, deux supports associés à une signification unique.

Propriété 2

Quelque soit j, j = 1...r, quelque soit i, i = 1...p, la covariance de η_j *avec la variable i-ème colonne de X est égale à la covariance de* ξ_j *avec la variable i-ème colonne de YR.*

En effet, $X'D\eta_j = X'DYRu_j = \lambda_j u_j$ et $(YR)'D\xi_j = R'(Y'DX)u_j = \lambda_j u_j$

Donc $X'D\eta_j = (YR)'D\xi_j$

Les coordonnées des variables de X selon le repère $\{\xi_j\}$ sont ainsi celles de YR selon le repère$\{\eta_j\}$.

En conclusion :

Les covariances $cov(\xi_j,\eta_j)$ représentent une décomposition de la mesure d'association $tr(V_{xy}V_{yx})^{1/2}$ des deux tableaux . Les propriétés 1 et 2 nous donnent des arguments pour prendre les couples (ξ_j,η_j) comme supports de la communité des deux tableaux et montrent combien la covariance semble adaptée à la mesure de cette communité. Dans notre contexte, nous nommerons cette analyse *"analyse de communité"*, préférant réserver le terme de ressemblance à une comparaison entre nuages de points et celui de communauté au contexte analyse canonique classique. Mais cette analyse ne sera totalement précisée qu'au § 4.

3.3 Communité, redondance et indices d'association

La mesure totale de communité $tr(V_{xy}V_{yx})^{1/2}$, ne dépend pas des rotations effectuées sur l'un des tableaux, mais dépend de l'échelle de mesure adoptée. Un indice qui lui est associable, en permettant une évaluation plus intrinsèque, est l'indice de Lingoes et Schönemann, introduit dans un contexte procrustéen et s'écrivant

$$cor(X,Y) = tr(V_{xy}V_{yx})^{1/2} / (trV_x\ trV_y)^{1/2}$$

Notons que d'autres formes intéressantes, "orientation-independent or spectra-independent", sont considérées pour cette indice par Ramsay, ten Berge et Styan (1984) et que les optimisations que nous proposerons par la suite pour cet indice seront de nature différente des leurs, bien que parfois non sans analogie.

Ici, la variance totale du tableau X, soit trV_x, pourrait être interprétée comme la mesure de communité totale de X avec lui même. Mais ce rapport avec la variance est surtout intéressant pour proposer des indices dérivés de l'indice cor(X,Y), associés à une nouvelle manière d'aborder la redondance entre tableaux (l'emploi de ce terme trouvant sa justification § 6), ou à une autre mesure de communité, comme nous le montrons ci-dessous.

Indices de redondance :

Soient $\{u_j\}$, $j = r+1,...,p$ un système de p-r vecteurs complétant le système $\{u_j\}$, $j = 1...r$ pour former une base orthonormée de E_x, et $\{v_j\}$, $j = r+1,...,p$, celui complétant le système $\{v_j\}$, $j = 1...r$, pour former une base orthonormée de E_y.

En posant $\xi_j = Xu_j$ et $\eta_j = Yv_j$ pour tout $j = 1,..., p$, les variances totales de chacun des deux tableaux peuvent se décomposer :

$$trV_X = \sum_{1,p} var(\xi_j) \qquad trV_Y = \sum_{1,p} var(\eta_j)$$

Or $\sum_{r+1,p} var(\eta_j)$ correspond à la part de variance totale ne présentant aucune communité avec les variables du tableau X, la part de variance impliquée dans cette communité étant égale à $\sum_{1,r} var(\eta_j)$.

Cette dernière mesure est celle d'une redondance de variance de Y par rapport à X (cf. § 6.3). Comme l'indice cor(X,Y) peut s'écrire

$$\sum_{1,r} cov(\xi_j,\eta_j) / (\sum_{1,p} var(\xi_j).\sum_{1,p} var(\eta_j))^{1/2},$$

l'indice suivant apparait comme un indice de *redondance de Y par rapport à X* :

$$\sum_{1,r} cov(\xi_j,\eta_j) / (\sum_{1,p} var(\xi_j) \sum_{1,r} var(\eta_j))^{1/2} = cor(X,YR)$$

De même cor(XR',Y) pourra être envisagé pour mesurer la redondance de X par rapport à Y.

Indices de communité :

Dans l'expression $\sum_{1,r} cov(\xi_j,\eta_j) / (\sum_{1,r} var(\xi_j).\sum_{1,r} var(\eta_j))$, n'apparaît ni la part de variance non redondante de X par rapport à Y, ni celle de Y par rapport à X. C'est une manière de mesurer , autrement qu'avec cor(X,Y), la communité entre X et Y. Mais nous serons amenés à proposer un autre indice que celui-ci par la suite.

4 - ANALYSE DE COMMUNITE

L'étude de ressemblances est inséparable de l'étude de différences. Mais il y a des différences "superficielles", comme pourrait l'être celle relative à deux objets identiques, mais de tailles différentes. Les traits les plus caractéristiques de la ressemblance de deux objets sont d'autant plus perceptibles que l'on a superposé ces deux objets : par l'imagination on cherche à les confondre (même centre de gravité, repère commun). On poursuit ici cette démarche, en effectuant des changements d'échelle. Il est possible d'en proposer pour Y, accentuant la superposition à X, sans modifier fondamentalement les résultats de l'analyse de ressemblance menée jusqu'à présent. On pourra alors observer la ressemblance profonde entre les deux juges: la façon analogue dont ils classent les élèves les uns par rapport aux autres, dans les matières correspondantes.

Ainsi η_j devrait avoir la même variance que ξ_j puisque ces deux variables sont deux supports associés à une signification unique. Cependant, comme elles ne sont pas corrélées, il est plus convenable d'adopter pour η_j une variance correspondant à la saturation de η_j par ξ_j, c'est à dire égale à $\rho_j^2 var\xi_j$ avec $\rho_j=\rho(\xi_j,\eta_j)$. On accorde ainsi la variance de η_j à celle de ξ_j . La variable synthétique η_j n'est pas corrélée avec les autres variables synthétiques ξ_i, $i \neq j$. Pour

des j différents, il n'y a donc pas d'interférences entre les différents ajustements (ce qui ne pouvait être conçu en agissant sur les variables du tableau Y).
Un autre avantage est d'obtenir des mesures de communité plus intrinsèques, indépendantes du choix des échelles adoptées pour le tableau Y, puisque ces mesures valent alors $\rho_j^2 \, var\xi_j$.

Un tel changement d'échelles avait été considéré, dans un contexte d'ajustement entre nuages d'individus, comme correspondant à la transformation par A du tableau YR, Lafosse (1985):

$$A = \sum_{1,r} a_j u_j u_j' \quad \text{avec} \quad a_j = cov(\xi_j,\eta_j) / var(\eta_j)$$

soit encore à la transformation par A_1 du tableau Y: $A_1 = RA = \sum_{1,r} a_j \, v_j u_j'$.

Les couples de l'analyse interbatterie de (X,YA_1) sont alors les couples $(\xi_j,a_j\eta_j)$, classés selon les valeurs $\rho_j^2 \, var\xi_j = cov(\xi_j,a_j\eta_j)$.

Finalement, la démarche que nous préconisons définissant **l'analyse de communité** est la suivante:

a) effectuer éventuellement un changement d'échelle initial décrit en 3.1.1, en nommant X le tableau pris pour référence .

b) calculer les couples (ξ_j,η_j) à partir de la décomposition en valeurs singulières de V_{xy} .

c) faire comme si ce calcul avait été fait non à partir de X et Y mais à partir de X et YA_1 , et classer ces couples selon les valeurs $\rho_j^2 \, var\xi_j$, qui sont alors les mesures de communité adoptées en même temps que variances des nouvelles variables synthétiques respectives $a_j\eta_j$.

d) effectuer les représentations graphiques des individus ou des variables, en sachant que la ressemblance se situe entre parts redondantes, c'est à dire entre le tableau projeté $X_1 = X\left(\sum_{1,r} u_j u_j'\right)$ et le tableau superposé YA_1.

Dans cette procédure, il n'est pas forcément utile d'exprimer la rotation R et le changement d'échelle A préconisés, sauf si on veut confirmer le résultat des lectures sur les graphiques en revenant aux tableaux X_1 et YA_1. Tout se passe comme si on avait superposé par l'imagination les deux tableaux pour pouvoir mieux les comparer.

On réalise les représentations graphiques des variables des tableaux en se servant des premières variables synthétiques pour bâtir des repères. En se référant à la propriété 2 du § 3.2.2, on peut représenter les variables du tableau X dans un premier repère normé non orthogonal (η_1,η_2) , si $\rho(\eta_1,\eta_2)$ est voisin de 1; les coordonnées sont les composantes de

$(\rho_1^2 \text{var}\xi_1)^{1/2} u_1$ et de $(\rho_2^2 \text{var}\xi_2)^{1/2} u_2$. Ou bien utiliser un repère compromis orthonormé (ξ_1,η_2), ou (η_1,ξ_2), pour représenter aussi bien les variables réduites de X que celles de YA_1. Ainsi, par exemple, nous pourrons observer quelles sont les matières les plus impliquées quant à une façon analogue pour les deux juges de classer les élèves les uns par rapport aux autres. Dans le même temps, une représentation des élèves peut être proposée dans le repère orthonormé (u_1,u_2), les coordonnées étant composantes de ξ_j et $a_j\eta_j$, j=1,2. A un axe j est associable la mesure $\rho_j^2 \text{var}\xi_j$ désignant l'importance du résumé fourni par cet axe, tant du point de vue de l'inertie des points issus de YA_1, que du point de vue communité. On relèvera notamment certains écarts importants entre les deux représentations d'un même élève issues des deux juges. L'orientation (direction et sens) du segment passant par les deux points appariés permet une interprétation en terme de variables. Pour alléger la visualisation, on pourrait ne représenter que les écarts, en ramenant à l'origine les représentations de tous les élèves issues d'un juge.

5.- INDICE DE LIAISON ET STATISTIQUE PROCRUSTE

L'expression de l'indice d'association **cor** liée à l'étude de communité précédente devient, aprés simplification:

$$\text{cor}(X,YA_1) = (\Sigma_{1,r}\, \rho_j^2 \text{var}(\xi_j) \,/\, \text{tr}V_X)^{1/2}$$

L'expression de la statistique procruste (Sibson (1978)) devient quant à elle:

$$D^2(X,YA_1) = \text{tr}V_X\,(1 - \text{cor}^2(X,YA_1))$$

Ce lien direct entre l'indice et la distance des carrés des écarts entre individus appariés augmente fortement la qualité attribuable à cet indice. Ce lien permet de décomposer la statistique selon les nouveaux couples traduisant une communité de moins en moins élevée. Il est remarquable que cette relation ne puisse s'introduire qu'après la transformation par A_1 du tableau Y (si on prend X pour référence). Par ailleurs, nous n'avions su proposer d'indice généralisant la mesure d'association entre plus de deux tableaux qu'après avoir fait subir aux tableaux un traitement analogue tout en prenant l'un des tableaux pour référence (Lafosse (1986)).

Notons que l'on pourrait préférer $\text{cor}^2(X_1,YA_1)$ à $\text{cor}^2(X,YA_1)$, pour évaluer l'intensité de la ressemblance entre parts redondantes. Alors on a :

$$D^2(X_1,YA_1) = (\Sigma_{1,r}\, \text{var}\xi_j)\,(1 - \text{cor}^2(X_1,YA_1))$$

Remarque :

Soient $b = (b_1,b_2,\ldots,b_r) \in R^r$ et c une constante arbitraire positive fixée. Considérons le problème :

$$\max_{b} \sum_{1,r} \text{cov}(\xi_j, b_j\eta_j)$$

sous la contrainte que $\sum_{1,r} \text{var}(b_j\eta_j)$ reste fixée à la valeur c.

On se ramène à un problème où la contrainte est celle d'appartenance à une sphéroïde, si on prend pour inconnues $y_j = b_j (\text{var}\eta_j)^{1/2}$; la solution est alors obtenue rapidement : $b_j = ka_j$, où k est une constante et les a_j sont les valeurs définies au § 4.

Comme l'indice cor est indépendant d'une dilatation globale de valeur k effectuée sur un tableau, la solution correspond à une optimisation de l'indice, puisqu'on peut augmenter la valeur du numérateur sans changer celle du dénominateur, en prenant k de sorte que c soit égale à la variance totale de YR.

D'où : $\text{cor}(X,YA_1) \geq \text{cor}(X,YR)$.

On aurait donc pu introduire le changement d'échelles A, comme un changement visant à optimiser l'indice d'association, quand X est pris pour référence.

6 - UNE ANALYSE DE DIFFERENCE

6.1 Introduction

Nous voulons prolonger l'étude de communité précédente, en analysant l'influence de la contrainte d'appariement des variables de X et de Y. Cette contrainte produisait l'appariement de ξ_j avec η_j, deux supports d'une même signification, pour chaque j. Pour analyser l'influence de la contrainte d'appariement, on se propose de réaliser une étude de communité en supprimant cette contrainte. Ce ne sera donc plus vraiment une étude de communité, et même, par comparaison avec les résultats de l'analyse de communité proprement dite, ce sera une façon d'analyser une différence entre tableaux. Pour notre exemple, il s'agirait cette fois d'étudier si la façon dont l'un des deux juges positionne les élèves les uns par rapport aux autres, se retrouve sensiblement mieux chez l'autre juge quand on n'impose plus que la coincidence se produise

forcément matières par matières. Cela peut avoir son intérêt si on ne s'intéresse qu'au classement global des élèves, toutes matières confondues. Par ailleurs, le domaine des applications peut être plus vaste que le nôtre , l'ajustement considéré pouvant être appliqué à des tableaux moins particuliers, avec un nombre de variables différent, seuls les individus restant appariés. Par exemple, les tableaux pourraient correspondre à deux codages différents d'une même information obtenue sur de mêmes individus.

Les deux triplets soumis à l'analyse sont (X,I,D) et (Z I,D), où $Z = YA_1$. Pour augmenter la communité que peut présenter le tableau Z avec le tableau X, par suppression de la contrainte d'appariement des variables, nous nous proposons d'envisager un modèle d'ajustement linéaire du tableau Z au tableau X, s'écrivant: X=ZM+E , où M est une matrice carrée d'ordre p, traduisant la modélisation linéaire de la différence et E la matrice nxp des résidus. Cet ajustement est conduit pas à pas de façon à optimiser l'indice **cor** entre X et Z, dont nous avons vanté les mérites et qui va servir de point de départ à notre démarche :

$$cor^2(X,Z) = \sum_{1,r} \rho_j^2 var\xi_j \,/\, trV_X$$

Une transformation du tableau Z qui définirait un modèle d'ajustement, sur la base d'une optimisation de la valeur de cet indice, ne peut être conçue qu'à partir d'une optimisation du numérateur. En supposant les variables ξ_j fixées, un meilleur choix consisterait à remplacer chacune des variables η_j par la projection D-orthogonale de ξ_j sur F_Z, l'espace des variables issu du tableau Z. Mais cela change toute la problématique qui a conduit à la définition des variables ξ_j elles-mêmes. D'où l'analyse suivante, présentée comme une recherche de variables synthétiques issues de X.

6.2 Analyse explicative d'un tableau par un autre

Les deux triplets statistiques soumis à l'analyse sont, pour le moment, (X,I,D) et (Y,I,D) et seuls les individus pourraient être appariés ici.

6.2.1 Première composante

Nous recherchons une première variable principale α_1 dans F_X, correspondant à la meilleure explication que l'on puisse trouver dans X par les variables de Y, selon le critère :

$$\max_{\alpha,\beta} \rho^2 \text{ var } \alpha$$

où $\beta \in F_Y$ *, où* ρ *est coefficient de corrélation linéaire entre* α et β, *et avec la contrainte d'homogénéité sur la variance de* α *:* $e'e = 1$ $(\alpha = Xe\ ,\ e \in E_X)$.

Notons tout de suite que β_1 ne peut être que la projection D-orthogonale de α_1 sur F_Y.

Sous les contraintes $e'e = 1$ et $\beta'D\beta = 1$, le problème peut encore s'écrire :

$$\max_{\alpha,\beta} \text{cov}(\alpha,\beta)$$

Ayant supposé injectives les applications X et Y, on peut finalement l'exprimer sous la forme :

$$\max_{\alpha,\beta} \text{cov}(\alpha,\beta)$$

sous les contraintes $e'e = 1$, $f'f = 1$, sachant que $\alpha = Xe$ et $\beta = (Y\ V_y^{-1/2})\ f$.

En effet, on a ici $\{\beta'D\ \beta = 1\} \Leftrightarrow \{f'f = 1\}$. L'espace des variables n'est pas modifié en substituant au tableau Y le tableau $YV_y^{-1/2}$. Tout se passe alors comme s'il s'agissait d'effectuer l'analyse de Tucker des triplets statistiques (X,I,D) et $(Y\ V_y^{-1/2},I,D)$.
On connaît donc la solution :

e_1 est le premier vecteur propre de $V_{xy}\ V_y^{-1}\ V_{yx}$, associé à la valeur propre: $\mu_1^2 = \rho_1^2$ var α_1.

6.2.2. Composantes successives

On reprend, à chaque étape, le même critère à optimiser avec la contrainte d'homogénéité qui l'accompagne. Les contraintes de succession sont choisies, quant à elles, de façon à correspondre à une décomposition de la variance totale de X, soit : $e_i'\ e_j = 0$, $i \neq j$.

Ainsi, à la deuxième étape, la solution est une variable α_2 associée à un couple (α_2,β_2) et aussi au couple (e_2,f_2) par les relations $\alpha_2 = X e_2$, $\beta_2 = YV_y^{-1/2} f_2$ et par le fait que β_2 est projection orthogonale sur F_y de α_2 : $\mu \beta_2 = YV_y^{-1}Y'D \alpha_2$.

De plus e_1 vérifie : $V_{xy} V_y^{-1} V_{yx} e_1 = \mu_1^2 e_1$.

Donc on a : $\beta_1' D\beta_2 = e_1' X'D \beta_2$

$= \mu^{-1} e_1' X'D YV_y^{-1}Y'D Xe_2$

$= \mu^{-1} e_1' V_{xy} V_y^{-1} V_{yx} e_2$

$= \mu^{-1} \mu_1^2 e_1' e_2$

$= 0$

Par suite, $f_1'f_2 = 0$. On retrouve de la sorte les deux contraintes de succession relatives à une analyse interbatterie de deux triplets (X,I,D) et $(YV_y^{-1/2},I,D)$.

On en déduit finalement que e_2 est le deuxième vecteur propre de $V_{xy} V_y^{-1} V_{yx}$; d'où α_2.

Ainsi de suite pour définir toutes les variables α_j associées respectivement aux couples (α_j,β_j), aux couples (e_j,f_j) et aux valeurs propres μ_j^2. Comme le rang de $V_{xy} V_y^{-1/2}$ est aussi celui de V_{xy} , le nombre de couples successifs est encore égal à r. Notons que le système $\{\beta_j\}$ est cette fois formé de variables non corrélées, ce qui n'est pas le cas pour $\{\alpha_j\}$.

Plus connue que cette analyse, l'A.C.P.V.I.— analyse en composantes principales de variables instrumentales (Rao (1964), synthèse de Sabatier (1983))— pose le problème de la recherche de variables dans F_y explicatives de celles de X. La solution se présente parfois comme l'A.C.P. du triplet statistisque $(YV_y^{-1}Y'DX, I, D)$, où le tableau correspond à la projection des variables de X sur F_y. La forme quadratique associée à cette A.C.P. est encore $V_{xy} V_y^{-1} V_{yx}$, et les variables explicatives obtenues sont les variables β_j. Les valeurs propres apparaissent comme des mesures d'intensité d'explication apportée par les variables.

En fait, les variables α_j ont été introduites pour la première fois par Johansson (1981) dans une analyse présentée comme une extention possible de l'analyse de redondance de Wollenberg (1977). Les couples (e_j,f_j) y étaient définis à partir de contraintes d'orthogonalité recherchées pour se rapprocher d'une analyse canonique. Cependant, la présentation donnée par nous pour définir cette analyse rend celle-ci plus autonome, moins reliée à l'obtention des variables explicatives, non introduite à partir d'une recherche de couples, même si ces couples sont finalement encore obtenus (par le biais de l'analyse interbatterie, non considérée par Johansson). En particulier, l'introduction des contraintes est différente. De plus, la façon de poser le problème à partir de l'indice **cor** nous appartient, et cela nous ramène à notre propos initial.

6.2.3 Ajustement de Y à X

Il s'agit de procéder pas à pas à un ajustement du tableau Y au tableau X et les couples (α_j,β_j) vont nous permettre d'y arriver : les variables synthétiques α_j sont apparues comme les mieux expliquées, les variables β_j les plus explicatives et de plus, la correspondance entre α_j et β_j et la non corrélation des β_j, nous montrent que l'explication trouvée α_j est toute entière apportée par β_j.

Un ajustement peut donc s'envisager en utilisant ces couples pour supports, en s'inspirant de la superposition de deux tableaux précédemment décrite en introduction à l'analyse de communité proprement dite. Ici la rotation à considérer est celle qui permet de superposer le système orthonormé $\{f_j\}$ au système orthonormé $\{e_j\}$ et le changement d'échelle est celui des affinités orthogonales opérées dans les directions communes ainsi définies, affinités qui correspondent aux régressions linéaires simples de α_j en β_j. La transformation qui effectue cela simultanément est donc la suivante, opérée sur le tableau $YV_y^{-1/2}$:

$$A_2 = \sum_{1,r} (\rho_j^2 \, \mathrm{var}\alpha_j)^{1/2} \; f_j \, e_j'$$

Or cette expression n'est autre qu'une décomposition en valeurs singulières de $V_y^{-1/2} V_{yx}$, ce qui mène finalement à la transformation du tableau Y par $L = V_y^{-1} V_{yx}$ et au modèle d'ajustement de Lebart, Morineau, Fénelon (1979) considéré dans leur analyse des covariances partielles :

$$X = YL + E$$

L'indice cor entre tableau cible et tableau ajusté s'exprime cette fois :

$$cor(X,YL) = [tr\ (V_{xy}\ V_y^{-1}\ V_{yx}) / tr\ V_x\ ]^{1/2}$$

et $cor^2(X,YL)$ est l'indice de redondance de Y par rapport à X de Stewart et Love (1968).

Cet indice apparaît donc aussi comme un indice de communité entre tableaux ajustés.

L'analyse de Wollenberg consistait à décomposer cet indice selon les mesures $\rho_j^2\, var\alpha_j$, le numérateur de cor^2 en étant la somme.
Des développements récents ont été apportés par Lazrac et Cléroux (1988) sur la mesure entre vecteurs aléatoires par l'indice de Stewart et Love (article [1]) et à l'usage de cette indice en sélection de variables (article [2]).

6.3 Analyse de la différence

L'analyse de la communité des deux tableaux X et Y s'est effectuée en fait, après traitement, sur les tableaux X et Z. L'analyse de différence qui s'ensuit porte sur une différence relevée entre les triplets (X,I,D) et (Z,I,D). Il s'agit d'observer l'effet produit, quand on supprime l'appariement des variables, sur la façon dont les variables disposent les individus les uns par rapport aux autres. Pour cela, on compare les résultats de l'analyse de communité avec ceux d'une analyse menée sans cette contrainte d'appariement.

Dans l'ajustement précédemment décrit entre X et Y, tout se passe comme s'il s'agissait de réaliser une analyse de communité des triplets (X,I,D) et $(YV_y^{1/2}, I, D)$, soit encore une analyse entre (X,I,D) et (Y,V^{1},D). C'est donc le même espace de variables F_y qui est considéré aussi bien dans l'analyse de (X,I,D) et (Y,I,D) que dans celle de (X,I,D) et $(YV_y^{-1/2},I,D)$, car un changement de métrique dans l'un des espaces des individus ne change pas celui des variables. L'image par X'D de cet espace commun nous donne ainsi le même sous-espace image dans E_X.

Par conséquent, dans la décomposition en valeurs singulières de V_{xy}, soit $\Sigma_{1,r}\, \lambda_j\, u_j\, v_j'$, et dans celle de $V_{xy}V_y^{1/2}$, soit $\Sigma_{1,r}\, \mu_j\, e_j\, f_j'$, les deux systèmes orthonormés $\{u_j\}$ et $\{e_j\}$ engendrent ce même sous-espace image.
Par suite, les variances totales redondantes sont les mêmes :

$$\Sigma_{1,r}\ var\alpha_j = \Sigma_{1,r}\ var\xi_j$$

Cela justifie à nos yeux l'introduction de la notion de redondance de variance considérée en analyse de communité à partir de l'indice cor (§ 3.3).
Par ailleurs, le fait de substituer le tableau YA_1 au tableau Y en analyse de communité, ou de substituer le tableau YL au tableau $YV_y^{-1/2}$ en analyse de la différence, ne change rien à ces analyses si elles sont considérées en temps qu'analyses factorielles. C'est en effet les mêmes couples principaux qui sont obtenus et classés de la même manière.
C'est dire aussi que l'analyse explicative de (X,I,D) par (Z,I,D) donne le même résultat que celle de (X,I,D) par (Y,I,D), ne dépendant pas des rotations ou des changements d'échelle considérés sur le tableau explicatif. Plus précisément, dans les modèles explicatifs associés, les résidus obtenus sont identiques :

$$X = ZM + E$$
$$X = YL + E$$

Ce résultat pourrait également être immédiatement déduit des travaux de Tyler (1982), qui s'est intéressé aux transformations laissant invariante la redondance mesurée par l'indice de Stewart et Love.

Nous en déduisons que : $cor(X,ZM) = [tr (V_{xy} V_y^{-1} V_{yx}) / tr V_x]^{1/2}$
indice que l'on peut associer à la statistique procruste :

$$D^2(X,ZM) = tr(V_x) [1 - cor^2(X,ZM)]$$

Notons ici que le travail d'ajustement proprement dit s'est effectué entre X_1 et Z, plutôt qu'entre X et Z. On peut donc juger intéressant d'écrire ces formules (et les mesures ci-dessous) en substituant à la variance totale de X celle redondante de X_1.

$[cor^2(X,ZM) - cor^2(X,Z)]$, ou bien $[D^2(X,Z) - D^2(X,ZM)]$, sont des mesures évaluant l'importance de la traduction par M de la différence.

$1 - cor^2(X,ZM)$ ou bien $D^2(X,ZM)$ sont des mesures évaluant l'importance des résidus.
Ces mesures sont décomposables selon les résumés successifs précédemment considérés.

Remarques :

Il est important de noter que le programme permettant les calculs et les représentations graphiques en analyse de communité peut être utilisé tel quel en analyse de différence, par exemple pour décrire les résidus E. Il suffit de substituer au tableau Y le tableau $YV_y^{-1/2}$

(en prenant garde toutefois à ne pas effectuer sur $YV_y^{-1/2}$ le changement d'échelle initial éventuel préconisé en 3.1.1).

Une analyse explicative de X par Y, où Y n'intervient que par l'intermédiaire de l'espace engendré F_y, peut inquiéter quant à la validité de certaines composantes qui y sont définies, si certaines dimensions de F_y sont relatives à un "bruit" plutôt qu'à un "signal". Il peut donc être souhaitable, avant de faire le calcul, de substituer au tableau Y un tableau reconstitué à partir des premières composantes principales de l'A.C.P. du triplet (Y,I,D). Mais ce problème peut ne pas être forcément à envisager , dans le cas de p mesures répétées 2 fois : les p-r dimensions perdues dans l'analyse de communité de X et Y, en remplaçant Y par Z, seront justement interprétés comme celles associables à un bruit, les r autres conservées étant relatives au signal.

Bien que défini dans un contexte non procustéen, on aurait pu envisager d'utiliser l'indice RV de Robert et Escoufier (1976) au lieu de l'indice cor. Mais l'inégalité $cor(X,YA_1) \geq cor(X,YR)$, que nous avons utilisée comme un des justificatifs à notre démarche, ne semble pas devoir être toujours vérifiée quand on remplace cor par RV.

Le lien entre analyse de Tucker et analyse canonique classique a été étudié par Momirovic et Dobric (1985) : en particulier, l'analyse canonique de (X,Y) peut se présenter comme une analyse de Tucker des tableaux $XV_x^{-1/2}$ et $YV_y^{-1/2}$. Pour nous, l'information redondante d'un tableau par rapport à un autre peut se décrire à travers une analyse de communité, une ACPVI ou une analyse canonique : elle est contenue dans un même espace vectoriel engendré.

7 - ETUDE D'UN EXEMPLE

14 élèves ont été notés par deux juges, pour donner les tableaux X et Y.
Les 5 colonnes de chaque tableau correspondent aux différentes épreuves subies, qui sont, dans l'ordre, Math., Physique, Chimie, Histoire, Lettres.
On a décidé de réduire les variables du tableau X. D'où l'usage de la même métrique dans R^5 relativement à Y. Les variables-matières sont centrées. La comparaison des deux tableaux ainsi préparés et donnés par la suite peut commencer.

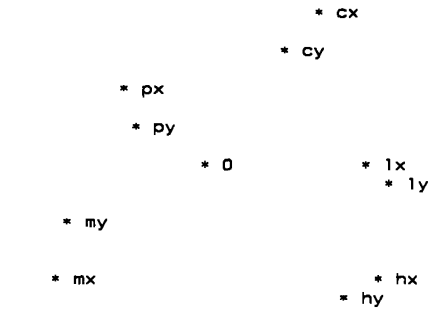
1 Comparaison des deux tableaux, avec contrainte d'appariement des variables.

Le rang de V_{xy} est égal à 4. Après superposition des deux tableaux — visant notamment à rapprocher les échelles utilisées par le juge Y de celles utilisées par le juge X en substituant à Y le tableau YA_1 — nous obtenons 4 parts de communité associées aux 4 premiers couples traduisant au mieux la communité des deux tableaux : 1.7866 1.1669 .4185 .0052
Comme ces mesures sont aussi mesures de l'importance du résumé fourni, les deux premiers axes associés à 87% de la variance totale de YA_1 permettent d'aborder l'essentiel de l'information commune aux deux tableaux.
cor2 valant .6754, on peut estimer la valeur de cet indice traduisant la communité globale de X et YA_1 suffisamment éloigné de 1 et trouver la différence entre les deux juges plutôt significative (on regrette ici de ne pas connaître la loi d'une statistique associée à cet indice, qui permettrait de faire dépendre notre appréciation du nombre d'élèves). D'après le graphe 1, toutes les matières semblent contribuer également à la différence entre les deux juges, vus les écarts de même ordre observables entre les deux représentations données pour chaque matière.

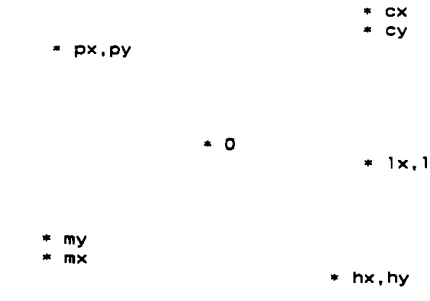
2 Comparaison sans contrainte d'appariement des matières

Les nouvelles mesures de communité valent cette fois : 2.3980 1.4959 .4562 .0201 , et donc l'essentiel de la communité est accessible à partir des deux premiers couples.
Le calcul s'est effectué entre X et $YV_y^{-1}V_{yx}$. Le travail de superposition est donc réalisé.
cor2 vaut .8740 . Cette fois, la loi de la statistique est connue (sous hypothèse de normalité, Lazraq et Cléroux [1]) et un test pourrait conduire à admettre que l'indice n'est pas significativement différent de 1, que les deux tableaux ne diffèrent plus significativement.
Nous pouvons dire que les deux juges notent différemment, si on les compare matière par matière, mais que globalement, toutes matières confondues, le classement des élèves les uns par rapport aux autres est à peu près le même.
D'après le graphe 2, s'il fallait cependant désigner deux matières induisant encore une différence entre les deux tableaux, malgré la suppression de la contrainte d'appariement, nous proposerions math. et chimie.
L'étude succinte qui vient d'être faite peut être complétée par les représentations graphiques des élèves, le jeu de dualité, le retour aux tableaux considérés...

1.9570	-1.1994	-1.7568	1.5006	-.9743	
2.3839	-.2144	-1.7933	-.7702	-1.2620	
.2877	3.0536	-1.2041	-1.1161	-.6289	
-.8178	.1101	1.6115	-2.2440	-1.1469	
-.8123	-.8054	1.2465	-.3792	.2343	
-.7905	-.7359	.8607	.4479	.7523	
-.7193	-.6663	.4905	1.0796	1.3278	tableau X
-.6865	-.5852	.3549	1.2450	1.4429	
-.5934	-.4114	.3549	.8991	.9249	
-.4894	-.2144	.3236	.4329	.9825	
-.3636	.0521	.2246	.1171	.6372	
-.2213	.3071	.2454	-.0784	-1.6073	
.0360	.5273	-.1717	-.2890	.0041	
.8296	.7822	-.7869	-.8454	-.6865	
1.1536	-.8551	-.8465	.6982	-.2919	
1.7173	-.4263	-.8674	-1.2869	-.5221	
.2451	1.0918	-.4190	-.6402	-.0041	
-.7455	.5240	1.4164	-2.3396	-.5221	
-.9644	-.2293	1.3225	-.4598	-.6372	
-.8495	.2922	.0399	-1.5726	1.0894	
-.9042	-.2641	.3683	2.0968	-.3494	tableau Y
-.8550	-.2177	.2536	2.1269	.0534	
-.7017	-.0670	.2432	1.4953	.2261	
-.5102	.0836	.1441	1.1043	.0534	
-.3405	.1879	.0190	.8035	.2836	
.0426	.2806	-.2417	.3373	.4563	
.6665	.0952	-.5181	-.3094	.3412	
2.0457	-.4958	-.9143	-2.0539	-.1768	



Graphe 1 . Representation des variables (appariement)



graphe 2 . Representation des variables (sans appariement)

REFERENCES

CAILLIEZ, F., PAGES, J.P. (1976)

Introduction à l'analyse des données. SMASH,9, Rue Durban - 75016 PARIS

CAZES, P. (1975)

Application de l'analyse des données au traitement de problèmes géologiques.

Thèse de 3-ème cycle. PARIS

GOLUB,G.H. et C.F. VAN LOAN (1983)

Matrix computations. The John Hopkins University Press.

JOHANSSON, J.K. (1981)

An extension of Wollenberg's redundancy analysis. Psychometrika, vol. 46, n° 1, pp.93-105

LAFOSSE, R. (1985)

Une nouvelle analyse procustéenne de deux tableaux, appariement typique et atypique de deux nuages. 4 th. Int. Symp. data Analysis and Informatics, Tome II, 1-8, INRIA, pp. 407-414

LAFOSSE, R. (1986)

Mètriques et analyses factorielles de deux tableaux ou plus.

Statistique et Analyse de Données. vol.11 n°3, pp. 51-75

LAZRAQ, A. , CLEROUX, R. (1988) [1]

Etude comparative de différentes mesures de liaison entre deux vecteurs aléatoires.

Statistique et analyse des données vol. 13 n°1, pp. 15-38

LAZRAQ, A. , CLEROUX, R. (1988) [2]

Un algorithme pas à pas de sélection de variables en régression linéaire multivariée.

Statistique et analyse des données vol. 13 n°1, pp. 39-58

LEBART, L., MORINEAU, A., FENELON, J.P. (1979)

Traitement des données statistiques : méthodes et programme. Dunod éditeur.

LINGOES, J.C., SHÖNEMANN, P.H. (1974)

Alternative measures of fit for the Shönemann-Carroll matric fitting algorithm.

Psychometrika, Vol. 39, n° 4. pp.423-429

MOMROVIC, K., DOBRIC, V. (1985)

Some relations between canonical and quasicanonical analyses.

4 th. Int. Symp. Data Analysis and Informatics -INRIA. poster pp. 101-105

RAO, C.R. (1964)

The use and the interpretation of component analysis in applied research.

Sankya, ser.A, 26.pp. 329-358

RAMSAY, J.O. , TEN BERGE, J. , STYAN, G.P.H. (1984)

Matrix correlation. Psychometrika, vol. 49, n°3, pp. 403-423

ROBERT, P. ESCOUFIER, Y. (1976)

A unifying tool for linear multivariate statistical methods : the RV coefficient.

Applied Statistics, Vol. 25, n° 3. pp.257-265

SABATIER, R. (1983)

Approximations d'un tableau de données. Application à la reconstruction des paléoclimats.

Thèse de 3ème cycle, Université MONTPELLIER.

SIBSON, R. (1983)

Studies in the Roustness of control of Multidimensional Scaling : Prorustes Statistics.

J. R. Statist. Soc. B, Vol. 40, n°2, pp. 234-239

SCHÖNEMANN, P.H., CARROLL, R.M. (1970)

Fitting one matrix to another under choice of control dilatation and a ridig motion.

Psychometrika, Vol. 35, n° 2, pp. 245-257

STEWART, D., LOVE, W. (1968)

A general canonical correlation index. Psychological Bull., Vol. 70, pp. 160-163.

TEN BERGE J.M.F., KNOL D.L. (1984)

Orthogonal rotations to maximal agreement for two or more matrices of different column orders.

Psychometrika, vol. 49, pp. 49-55

TUCKER, L.R. (1958)

An inter-battery method of factor analysis. Psychometrika, Vol. 23, n° 2.

TYLER, D. (1982)

On the optimality of the simultaneous redundancy transformations.

Psychometrika, Vol. 47, n° 1, pp. 77-87

WOLLENBERG, A. (1977)

Redundancy analysis. An alternative for canonical correlation analysis.

Psychometrika, Vol. 42, n° 2, pp. 207-221