

SPAD•TM

VERSION **5.0**

Analyse de tableaux multiples



CISIA-CERESTA - 261 rue de Paris - 93556 MONTREUIL Cedex

tel : +33 (0)1 55 82 15 15 - Fax : +33 (0)1 43 63 21 00

e-mail : ciaia@fr.inter.net - Web : <http://www.cisia.com>

SPAD•TM

Option Tableaux Multiples

**MANUEL DE
REFERENCE**

SPAD•TM ®

Manuel de référence

*Le **logiciel** décrit dans le manuel est diffusé dans le cadre d'un accord de licence d'utilisation et de non divulgation, et ne peut être utilisé ou copié qu'en conformité avec les stipulations de l'accord. Toute copie du programme sur cassette, disque ou autre support à des fins autres que l'usage personnel du programme par le licencié est interdite par la loi. Les informations figurant dans ce **manuel** sont sujettes à révision sans préavis et ne présentent aucun engagement de la part du CISIA.*

© Copyright CISIA•CERESTA 1993, 2001
ISBN 2-906711-20-9

Centre International de Statistique et d'Informatique Appliquées
261 rue de Paris, 93556 Montreuil Cedex (France)
Tel : 01 55 82 15 15 – Fax : 01 43 63 21 00
e-mail : cisia@fr.inter.net – Web : <http://www.cisia.com/>

SOMMAIRE

Procédure AFMUL 6

1.	PRÉSENTATION	6
2.	INSTRUCTIONS DE COMMANDE.....	7
3.	PRÉSENTATION DÉTAILLÉE DES PARAMÈTRES	8
4.	DÉFINITION DES GROUPES DE VARIABLES	10
5.	EXEMPLE DE COMMANDE	11
6.	FICHIERS NÉCESSAIRES À L'EXÉCUTION.....	12

PROCEDURE AFMUL La méthode 13

1.	TABLEAU ANALYSÉ ; NOTATIONS ; EXEMPLES.....	13
2.	TECHNIQUES CLASSIQUES OU AFM ?	14
3.	LA PONDÉRATION DES GROUPES DE VARIABLES	14
4.	L'AFM EST UNE ANALYSE FACTORIELLE	15
5.	REPRÉSENTATION SUPERPOSÉE DES <i>J</i> NUAGES D'INDIVIDUS.	16
6.	L'AFM EST UNE ANALYSE MULTICANONIQUE.....	18
7.	ETUDE GLOBALE DES GROUPES	19
8.	DEUX MESURES DE LIAISON ENTRE DEUX GROUPES DE VARIABLES	20
9.	REPRÉSENTATION DES GROUPES.....	20
10.	LIENS ENTRE L'ANALYSE GLOBALE ET LES ANALYSES SÉPARÉES DES GROUPES	21
11.	BIBLIOGRAPHIE SOMMAIRE.....	22

Exemple 1 : Analyse Factorielle Multiple..... 23

1.	MATRICE DES CORRÉLATIONS ET/OU DES COVARIANCES ENTRE COLONNES ACTIVES.....	24
2.	DIAGRAMMES DES INERTIES (= VALEURS PROPRES) DES ANALYSES SÉPARÉES	25
3.	MATRICE DES CORRÉLATIONS ENTRE FACTEURS PARTIELS.....	27
4.	DIAGRAMME DES INERTIES (= VALEURS PROPRES) DE L'ANALYSE GLOBALE	28
5.	MATRICES DES LIAISONS ENTRE GROUPES	28
6.	AIDES À L'INTERPRÉTATION DES GROUPES	30
7.	CORRÉLATIONS ENTRE LES VARIABLES CANONIQUE ET LES VARIABLES GÉNÉRALES.....	31
8.	RAPPORTS (INERTIE INTER)/(INERTIE TOTALE).....	32
9.	AIDE À L'INTERPRÉTATION DES INDIVIDUS ACTIFS.....	33
10.	INDIVIDUS AYANT LES PLUS FORTES CONTRIBUTIONS.....	35
11.	INDIVIDUS AYANT LES PLUS FORTES ET LES PLUS FAIBLES INERTIES INTRA	36
12.	INDIVIDUS PARTIELS AYANT LES PLUS FORTES INERTIES INTRA	37
13.	AIDES À L'INTERPRÉTATION DES MODALITÉS (CENTRES DE GRAVITÉ).....	37
14.	CENTRES DE GRAVITÉ AYANT LES PLUS FORTES CONTRIBUTIONS	40
15.	CENTRES DE GRAVITÉ AYANT LES PLUS FORTES ET LES PLUS FAIBLES INERTIES INTRA.....	41
16.	CENTRES DE GRAVITÉ PARTIELS AYANT LES PLUS FORTES INERTIES INTRA.....	42
17.	AIDES À L'INTERPRÉTATION DES COLONNES.....	43
18.	AIDES À L'INTERPRÉTATION DES AXES PARTIELS	45
19.	REPRÉSENTATION GRAPHIQUE DES VARIABLES CONTINUES	47
20.	REPRÉSENTATION GRAPHIQUE DES INDIVIDUS MOYENS	48
21.	REPRÉSENTATION SUPERPOSÉE DES NUAGES (D'INDIVIDUS) PARTIELS	49
22.	REPRÉSENTATION DES VARIABLES NOMINALES.....	50
23.	REPRÉSENTATION DES GROUPES.....	51

Procédure STTIS 52

1.	PRÉSENTATION	52
2.	INSTRUCTIONS DE COMMANDE.....	55
3.	PRÉSENTATION DÉTAILLÉE DES PARAMÈTRES	56
4.	EXEMPLES DE COMMANDE	60
5.	FICHIERS NÉCESSAIRES À L'EXÉCUTION.....	62

PROCEDURE STTIS La méthode 63

1.	GÉOMÉTRIE D'UN NUAGE DE POINTS	63
2.	FAUT-IL COMPARER DES VECTEURS NORMÉS OU NON NORMÉS ?	67
3.	EXEMPLES DE CALCUL DE PRODUITS SCALAIRES	69
4.	PROCÉDURE STTIS, OPTION N°1	71
5.	PROCÉDURE STTIS, OPTION N°2	75
6.	PROCÉDURE STTIS, OPTION N°3	81
7.	RÉFÉRENCES.....	84

Exemple 2 : Analyse Conjointe de Tableaux - Méthode STATIS.... 85

1.	SORTIES DE L'OPTION N°1 : COMPARAISON DES TABLEAUX NON NORMÉS.....	86
2.	SORTIES DE L'OPTION N°2 : COMPARAISON DES NUAGES DE LIGNES	91
3.	SORTIES DE L'OPTION N°3 : COMPARAISON DES MATRICES DE CORRÉLATIONS.....	109

1. Présentation

1.1 Objet

Cette procédure effectue l'analyse factorielle d'un fichier dans lequel un ensemble d'individus est caractérisé par plusieurs groupes de variables. Les variables peuvent être nominales ou continues.

Le premier intérêt de la méthode est de réaliser une analyse factorielle dans laquelle l'influence des différents groupes de variables est équilibrée a priori. Cet équilibre est nécessaire car les groupes de variables diffèrent toujours selon le nombre de variables et/ou selon la structure de ces variables (i.e. leurs interrelations). En particulier, ceci permet de faire intervenir, en tant qu'éléments actifs, simultanément des groupes de variables nominales et des groupes de variables continues.

La procédure crée, à partir des fichiers NDIC et NDON (créés par la procédure SELEC), le fichier NGUS des coordonnées factorielles.

1.2 Edition

La procédure fournit, dans un premier temps, le tri-à-plat des variables nominales et les statistiques sommaires des variables continues; elle édite, pour chaque groupe de variables, l'histogramme des valeurs propres issues de l'analyse factorielle effectuée sur ce groupe de variables. Elle édite aussi, pour chaque groupe de variables continues, la matrice des corrélations (ou des covariances selon la normalisation). On édite ensuite la matrice des corrélations entre facteurs partiels (facteurs des analyses séparées).

On trouve ensuite les résultats de l'AFM proprement dite, analyse factorielle globale de l'ensemble des variables : on obtient , l'histogramme des valeurs propres, coefficients **Lg** et **RV** de liaison entre groupes, les corrélations entre les variables canoniques et les facteurs de l'analyse globale, les rapports inertie inter/inertie totale, les contributions et cosinus carrés de tous les groupes (actifs et illustratifs). Si l'édition en est demandée, le listage comporte les coordonnées, contributions et cosinus carrés des individus et individus partiels associés (actifs et éventuellement illustratifs).

Pour les individus et les modalités, on trouve les tableaux triés des éléments ayant les plus fortes contributions, les éléments ayant les plus fortes et les plus faibles inerties intra, les éléments partiels ayant les plus fortes inerties intra.

De plus, on édite les coordonnées et valeurs-test de toutes les modalités et de toutes les modalités partielles.

Le listage suivant fournit les coordonnées, les contributions, les cosinus carrés des variables continues actives et illustratives, les corrélations entre les axes et les variables non réduites (actives et illustratives). On obtient enfin les coordonnées, les contributions et les cosinus carrés des axes partiels des analyses (actifs et illustratifs).

1.3 Paramètres

La procédure comporte deux types de paramètres :

- **Les paramètres de fonctionnement** contrôlent les calculs effectués : NXGRP indique le nombre maximal de groupes. NAXE est le nombre d'axes retenus pour l'analyse partielle. NFACT fixe le nombre d'axes qui seront retenus pour l'analyse finale tandis que PCMIN indique le seuil d'apurement des modalités.
- **Les paramètres d'édition** contrôlent l'édition des valeurs propres et le nombre d'axes édités (NAXED) pour les analyses partielles et globales.

2. Instructions de commande

Les valeurs par défaut sont indiquées entre parenthèses, à la suite du nom du paramètre.

- (1) PROC AFMUL analyses factorielles multiples
- (2) titre de l'analyse
- (3) NXGRP (15) nombre maximal de groupes de variables.
 NAXE (3) nombre d'axes partiels retenus
 TOT : tous les axes des analyses partielles.
 NAXED(5) nombre d'axes pour lesquels seront édités les coordonnées.
 0 ou NON : pas d'édition.
 TOT : tous les axes sont édités.
 NFACT (5) nombre d'axes retenus de l'analyse globale.
 TOT : tous les axes sont conservés.
 PCMIN (2.) seuil d'apurement des modalités (en pourcentage).
 LEDIN (0 ou NON) édition des coordonnées des individus et des individus partiels.
 0 ou NON : pas d'édition.
 1 ou ACT : édition pour les individus actifs.
 2 ou TOT : édition pour tous les individus.
- (4) listes de commande des analyses partielles.
 Dans chaque liste, on indique le type de l'analyse partielle et on définit les variables " colonnes " sur lesquelles l'analyse sera effectuée.
- (5) FIN fin des listes de commandes d'analyses partielles.

3. Présentation détaillée des paramètres

NXGRP

nombre maximal de groupes de variables

- *valeurs possibles* : toutes les valeurs plus grandes que 1
- valeur par défaut : 15

NAXE

nombre d'axes partiels retenus pour les analyses partielles

- *valeurs possibles* : de 1 à nombre des variables à analyser pour les groupes de variables continues (ACP).
de 1 à (nombre des modalités - nombre de questions) pour les groupes de variables nominales (ACM).
- valeur par défaut : 3

Ce paramètre détermine le nombre d'axes partiels (=axes des analyses séparées des groupes) qui seront sauvegardés et traités comme des variables illustratives. Il indique le nombre d'axes partiels pour chaque groupe. Pour un groupe de variables continues, la valeur maximale de ce paramètre est égale au nombre des variables continues; pour un groupe des variables nominales, sa valeur est égale au nombre de modalités moins le nombre de questions dans ce groupe.

NAXED

nombre d'axes pour lesquels seront éditées les coordonnées

- *valeurs possibles* : de 1 à valeur de NAXE.
TOT (tous les axes sont retenus).
- valeur par défaut : 3

Lorsqu'on édite un tableau de coordonnées (pour les individus ou les variables), les coordonnées sont listées par groupe de 5. La valeur par défaut assure donc l'édition des 5 premiers axes pour tous les tableaux demandés. Si NAXED n'est pas un multiple de 5, le tableau est complété par des 0.

NFACT

nombre d'axes retenus de l'analyse globale

- *valeurs possibles* : de 1 à la somme de tous les axes possibles des analyses partielles.
- valeur par défaut : 5

Ce paramètre détermine le nombre d'axes de l'analyse globale qui seront sauvegardés, et qui pourront donc être édités. La valeur maximale de ce paramètre est la somme du nombre de variables continues actives NCACT et du nombre de modalités actives NMACT moins le nombre de questions actives NQACT. Si la valeur codée pour NFACT s'avérait de ce fait trop grande, elle serait ramenée à la valeur maximale autorisée. Dans la pratique, quelques dizaines d'axes sont généralement suffisants pour les utilisations ultérieures.

PCMIN**seuil d'apurement des modalités actives (en pourcentage)**

- valeurs possibles: réelles, de 0. à 99.
- valeur par défaut: 2.

Pour chaque question, les individus possédant une modalité de poids total faible se voient attribuer au hasard une des modalités de poids suffisant dans la question considérée (cette opération d'apurement permet de conserver au tableau des données son caractère "disjonctif complet").

Le paramètre PCMIN fixe le pourcentage du poids total des individus actifs en dessous duquel une modalité est considérée de poids trop faible. Si tous les individus ont le poids 1, PCMIN est le pourcentage de l'effectif des individus actifs au dessous duquel une modalité se trouvera ventilée.

Si toutes les modalités d'une question (ou toutes sauf une) ont un poids trop faible, la question elle-même sera passée en illustrative pour le calcul des axes.

La valeur par défaut (2 %) convient pour la plupart des analyses. Si le paramètre est mis à 0, seules les modalités de poids nul seront éliminées. Dans le cas où aucun individu actif n'a un poids nul, ceci revient à ignorer les modalités d'effectif nul.

LEDIN**édition des coordonnées des individus**

- *valeurs possibles* : 0 ou NON (pas d'édition)
1 ou ACT (édition pour les individus actifs)
2 ou TOT (édition pour tous les individus)

- *valeur par défaut* : 0 ou NON

L'édition fournit les coordonnées pour chaque individu, suivies des contributions à l'inertie (nulles pour les individus illustratifs) et des cosinus carrés.

4. Définition des groupes de variables

Après la liste de paramètres, il faut définir la liste de des groupes de variables ou, ce qui revient au même, la liste des analyses partielles. Chaque liste est constituée par deux mots-clés, une liste des variables et de paramètres optionnels.

La forme générale suit le schéma suivant :

mot-clé 1 mot-clé 2 liste des variables du groupe
paramètres optionnels

mot-clé 1 spécifie le mode des variables dans l'analyse globale. Il peut valoir GACT, GILL.

GACT : les variables du groupe sont actives.

GILL : les variables du groupe sont illustratives.

mot-clé 2 spécifie le type de l'analyse partielle. Il peut valoir : NOMI, COVA, CORR

NOMI : ACM sur variables nominales.

COVA: ACP non normée sur variables continues.

CORR: ACP normée sur variables continues.

liste des variables du groupe est une liste donnant les numéros d'origine des variables choisies. Chaque numéro est précédé de la lettre 'V'. Le séparateur peut être le blanc ou la virgule. A l'intérieur du groupe, les variables doivent être de même type.

Paramètres optionnels

Dans chaque analyse partielle, on peut introduire des paramètres optionnels qui contrôlent le nombre de axes partiels (NAXE), le nombre de valeurs propres à éditer (NAXED) et la métrique utilisée dans cette analyse (PERM).

- 1) NAXE : Le nombre d'axes partiels à conserver entant que variable illustrative dans l'analyse globale.
- 2) NAXED: le nombre de valeurs propres à éditer (NAXED).
- 3) PERM : pondération des variables dans le groupe (valeurs réelles).
Il doit y avoir autant de valeurs que de variables du groupe. Par exemple,

GACT NOMI V3, V10--V15 NAXE = 6 >
PERM = 0.1, 30., 10.5, 4., 51., 65.

Dans une analyse partielle, si la valeur du paramètre NAXE (ou NAXED) n'est pas donnée, elle vaut la valeur du paramètre NAXE (ou NAXED) pour tous les groupes; si la pondération n'est pas définie, les variables sont équipondérées (chaque variable a un poids de 1 par défaut).

5. Exemple de commande

Un exemple est présenté ci-dessous. Dans la procédure AFMUL, la commande de la procédure SELEC qui doit précéder AFMUL est donnée pour mémoire.

```

-----+-----1-----+-----2-----+-----3-----+-----4-----+-----5-----+-----6-----+-----7-----+
PROC SELEC
Selection des individus et des variables
LEDIT = COURT, LESTI = NON, LZERO = REC, LSELI = TOT, IMASS = UNIF
NOMI ILL 1,2
CONT ACT 3--13
FIN
:=====
PROC AFMUL
Analyse Factorielle Multiple
NXGRP=5, NFACT=5, NAXED=5, NAXE=5, PCMIN=2., LEDIN=NON
GACT CORR V8--V10 V13 NAXE=2 NAXED=2
GACT COVA V3--V7, V11,V12 NAXE=3 NAXED=2
GILL NOMI V1,V2
FIN
-----+-----1-----+-----2-----+-----3-----+-----4-----+-----5-----+-----6-----+-----7-----+

```

Dans cet exemple, l'utilisateur veut effectuer une analyse sur un mélange des variables de différentes natures. Les groupes de variables sont définies au travers de la liste des analyses partielles. Chaque élément de la liste comporte le mode, le type de l'analyse partielle (qui implique le nature de variables dans le groupe), la liste des "variables" constituant le groupe à analyser et enfin le nombre d'axes retenus et édités . Par exemple, la première instruction

GACT CORR V8--V10 V13 NAXE=3 NAXED=2

précise que cette analyse partielle est une ACP normée (mot-clé CORR) sur un groupe des variables continues constitué par les variables 8 à 10 et la variable 13. Elles sont actives dans l'analyse globale (mot-clé GACT). Enfin l'utilisateur demande de retenir 3 axes factoriels de cette analyse partielle (ces axes joueront le rôle de variables illustratives dans l'analyse globale) et édite les deux premières valeurs propres.

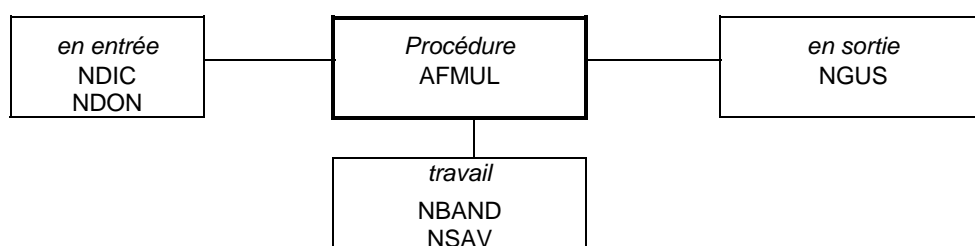
Pour la troisième instruction :

GILL NOMI V1,V2

Il s'agit d'une ACM (mot-clé NOMI) sur le groupe des variables nominales 1 et 2. On retient trois (valeur par défaut) axes et édite 5 (valeur par défaut) valeurs propres. Les variables sont prises en compte comme variables illustratives dans l'analyse globale (mot-clé GILL).

6. Fichiers nécessaires à l'exécution

- en lecture NDIC (dictionnaire utile)
NDON (données utiles)
- en écriture NGUS (coordonnées factorielles)
- de travail NSAV, NBAND



PROCEDURE AFMUL

La méthode

1. Tableau analysé ; notations ; exemples

L'AFM (Analyse Factorielle Multiple) traite des tableaux dans lesquels un ensemble d'individus est décrit par plusieurs groupes de variables. Au sein d'un même groupe, les variables doivent être de même type (continu ou nominal) mais d'un groupe à l'autre, même pour les groupes actifs, les variables peuvent être de types différents.

groupes	1		j			J	
variables	1	K_1	1	k	K_j	1	K_J
individus	1			\vdots			
	i			x_{ik}			
	I						

Tableau des données. x_{ik} : valeur de la variable k pour l'individu i . Si k est continue x_{ik} est un nombre ; si k est nominale x_{ik} est un numéro de modalité. Un groupe est désigné indifféremment par son numéro (j) ou son nombre de variables (K_j).

Exemples

Enquête. Les individus sont les enquêtés ; les variables sont les questions. Ces questions sont regroupées en thèmes qui constituent chacun un groupe.

Analyse sensorielle. Les individus sont les produits. Un premier groupe de variables comprend des descripteurs sensoriels ; un second groupe comprend des résultats d'analyses chimiques.

Ecologie. Les individus sont des points d'observation. Un premier groupe de variables décrit les caractéristiques du sol ; un second groupe décrit la flore.

Données temporelles. Un même ensemble d'individus a été observé à plusieurs dates. A chaque date, on mesure plusieurs variables. Dans un tel cas, on peut définir les groupes de variables de deux façons différentes : un groupe peut rassembler les variables relatives à une même date ou rassembler, lorsque les variables observées sont les mêmes d'une date à l'autre, les différentes dates pour une même variable.

2. Techniques classiques ou AFM ?

Face à ce type de données, la pratique usuelle en analyse factorielle consiste à introduire un seul groupe en actif, les autres étant introduits en illustratif. Ainsi, on obtient une certaine homogénéité des variables actives, caractéristique qui conduit naturellement à une problématique simple : on recherche d'abord les principaux facteurs de variabilité des individus en fonction d'un thème précis et, dans un deuxième temps, on relie chacune des variables illustratives à ces principaux facteurs de variabilité.

Cette méthodologie de base est excellente. Plusieurs remarques incitent à la compléter.

- Dans cette démarche, la seule structure multidimensionnelle véritablement prise en compte est celle des variables actives ; les variables illustratives interviennent indépendamment les unes des autres.
- La notion d'homogénéité est relative. Souvent, dans un questionnaire, les questions peuvent être regroupées en thèmes à l'intérieur desquels on peut distinguer des sous-thèmes. Quel niveau choisir pour définir les variables actives ?
- Selon un autre point de vue, on peut vouloir introduire en actif plusieurs thèmes simultanément, c'est-à-dire les prendre simultanément en compte dans la définition de la ressemblance entre individus. Nous détaillons dans la suite en quoi introduire plusieurs thèmes en actif est différent d'introduire en actif toutes les variables sans distinction de thèmes.

Conclusion. En un certain sens, le prolongement naturel de la méthodologie usuelle consiste à prendre en compte la structure de groupe de variables dans les analyses. Le fait de distinguer des groupes dans un ensemble de variables d'une part pose un problème technique d'équilibre entre les groupes et d'autre part enrichit la problématique. Nous examinons successivement ces deux points.

3. La pondération des groupes de variables

Plusieurs arguments incitent à pondérer les groupes de variables. Le plus simple est le suivant. Si l'on introduit tous les groupes de variables en actif sans se préoccuper d'équilibrer leur influence, on peut imaginer qu'un groupe contribue de façon majeure à la construction des premiers axes. Auquel cas, on pense réaliser une analyse fondée sur tous les groupes mais, en pratique, on examine les résultats de l'analyse d'un seul d'entre eux.

Ainsi, implicitement, introduire simultanément en actif plusieurs groupes de variables nécessite d'équilibrer l'influence de ces groupes. Cette influence dépend directement de la structure du groupe au sens de la répartition de l'inertie (des nuages qu'il induit) dans les différentes directions de l'espace. Par exemple, si un groupe présente une très forte inertie dans une direction donnée, cette inertie va influencer fortement le premier axe de l'analyse globale.

D'où l'idée de normaliser l'inertie axiale maximum de chacun des groupes. Techniquement, cela revient à accorder à chaque variable du groupe j un poids égal à $1/\lambda_j^j$ en appelant λ_j^j la première valeur propre de l'analyse factorielle séparée du groupe j . Ces analyses séparées sont souvent appelées « partielles » par la suite.

Cette pondération, qui possède des propriétés remarquables qui apparaissent par la suite, s'interprète directement de façon simple : si l'on considère les deux nuages (celui des individus et celui des variables) induits par chaque groupe de variables, la pondération de l'AFM normalise ces deux nuages en rendant leur inertie axiale maximum égale à 1.

Cette pondération n'équilibre pas l'inertie totale des groupes. Ainsi, un groupe de forte dimensionalité aura une influence globale dans l'analyse importante en ce sens que ce groupe contribuera à la construction de nombreux axes. En revanche, ce groupe fortement multidimensionnel n'a aucune raison particulière d'infléchir de façon privilégiée les premiers axes. Corrélativement, un groupe unidimensionnel ne peut contribuer fortement à la construction que d'un seul axe mais n'a aucune raison particulière de ne pas infléchir le premier axe.

Remarque : une mesure de la dimensionalité (N_g)

La pondération présentée ci-dessus suggère un indice de dimensionalité d'un tableau. Appliqué au tableau engendré par le groupe K_j , cet indice s'écrit :

$$[N_g(K_j)]^2 = \sum_s \left(\frac{\lambda_s^j}{\lambda_1^j} \right)^2$$

$(N_g)^2$ est la somme des carrés des valeurs propres de l'analyse factorielle séparée du groupe en utilisant la pondération de l'AFM. Cet indicateur est d'autant plus grand que le tableau possède un grand nombre de facteurs d'importance comparable au premier d'entre eux. Si le groupe ne comporte que des variables continues réduites non corrélées deux à deux, cet indicateur est égal au nombre de variables. Il peut s'interpréter comme un « nombre de dimensions », chacune étant « pondérée » par son inertie. L'élévation au carré se justifie pleinement de façon indirecte mais s'interprète directement aussi : elle renforce le rôle des directions de plus grande inertie.

4. L'AFM est une analyse factorielle

Le cœur de la méthode est une analyse factorielle de l'ensemble des groupes actifs (dite analyse globale; le nuage des individus associé à cette analyse est dit global ou encore « moyen »). Pour les variables continues, l'AFM se comporte comme une ACP (les variables étant pondérées) ; pour les variables nominales, l'AFM se comporte comme une ACM (les variables étant pondérées). C'est la présence des poids des variables, qui équilibrent les inerties axiales maximums des groupes, qui rend possible la présence simultanée de variables continues et nominales parmi les éléments actifs. Elle rend aussi possible la présence simultanée de variables continues centrées réduites et de variables continues non réduites.

L'objectif est ici de mettre en évidence les principaux facteurs de variabilité des individus, ces derniers étant décrits de façon équilibrée par plusieurs groupes de variables (ceux introduits en actifs).

Selon ce point de vue, l'AFM fournit les résultats classiques des analyses factorielles, soit axe par axe :

- les coordonnées, contributions et cosinus carrés des individus ;
- les coefficients de corrélation entre les variables continues et les facteurs ;
- pour chaque modalité des variables nominales, la coordonnée (et la valeur-test associée) du centre de gravité des individus présentant cette modalité.

Remarque sur les modalités en AFM

En AFM, les modalités sont représentées exactement par les centres de gravité associés. Sous le terme de contribution, le logiciel fournit, pour chaque modalité, le rapport entre l'inertie de ce centre de gravité et l'inertie totale. Pour les variables (nominales) actives, cet indicateur n'est que proportionnel axe par axe à la notion usuelle de contribution utilisée en ACM (ainsi sa somme pour tous les éléments actifs ne vaut pas 1). Cet indicateur possède la propriété suivante : sa somme pour toutes les modalités d'une même variable et pour l'axe de rang s est égale au rapport de corrélation entre cette variable et le facteur de rang s . Il est ainsi aisé de détecter, pour chaque facteur, les variables nominales qui lui sont liées. On calcule ce rapport de corrélation pour toutes les variables nominales, actives et illustratives. Chaque rapport de corrélation est complété par une valeur-test, celle qui est associée au test F de l'analyse de variance « expliquant » le facteur en fonction de la variable nominale.

5. Représentation superposée des J nuages d'individus.

A chaque groupe de variables j , on associe un nuage d'individus N_j^j , nuage étudié dans l'analyse factorielle séparée du groupe j . Ce nuage est dit « partiel » ; il contient les individus « partiels », notés j^i (individu i vu par le groupe j).

On peut souhaiter examiner s'il existe des structures communes à ces J nuages N_j^j , c'est-à-dire des parentés, d'un nuage à l'autre, entre les proximités entre points homologues.

Pour cela, on cherche une représentation superposée des N_j^j qui :

- représente bien chacun des N_j^j ;
- mette en avant les parentés précitées, c'est-à-dire fasse apparaître proches, autant que faire se peut, les points homologues.

Cette problématique est proche de celle de l'analyse procustéenne. L'AFM fournit une telle représentation superposée en projetant, sur les axes issus de l'analyse globale, les N_j^j à la manière d'éléments illustratifs (en fait il ne s'agit pas exactement d'éléments illustratifs puisque leurs données ont contribué à la construction des axes ; cette remarque est importante puisque cette représentation n'a de sens que pour les N_j^j correspondant aux groupes actifs).

5.1 Un indicateur de ressemblance globale entre les représentations axiales des N_i^j

Concrètement, lorsque les différents groupes induisent la même structure sur les individus, les points homologues sont proches. Cette propriété globale s'évalue au moyen de l'indicateur dont la construction est décrite ci-après.

On considère l'ensemble des points des J nuages N_i^j et la partition de ces $I \times J$ points en I classes, partition selon laquelle on place dans la même classe i les J points j homologues de l'individu i . Lorsqu'un axe fait apparaître une structure commune aux différents groupes, les points j relatifs au même individu i sont proches les uns des autres et cette partition présente, le long de l'axe, une faible inertie intra-classe. D'où l'idée de calculer axe par axe le rapport *inertie inter* / *inertie totale* associé à cette partition. Ce rapport est d'autant plus proche de 1 que l'axe met en évidence une structure commune aux différents groupes.

Attention : ce rapport ne décroît pas nécessairement avec le rang de l'axe puisque ce n'est pas le critère optimisé par l'AFM.

5.2 Examen détaillé des représentations axiales des N_i^j

Plus localement, on examine pour chaque individu i (i.e. chaque sous-nuage des j associés au même individu i) sa part d'inertie intra axe par axe. En sélectionnant les projections des individus ayant les plus faibles et les plus fortes inerties intra, on repère ainsi ; pour l'axe s :

- les individus dont les points homologues sont très proches (faible inertie intra) ; ils illustrent bien la structure commune représentée par l'axe s ;
- les individus dont les points homologues sont éloignés les uns des autres (forte inertie intra) ; ils constituent les « exceptions » à la structure commune.

Dans le détail, on peut examiner l'éloignement de chaque point j par rapport au point moyen de l'individu i . Cet examen se fait visuellement ou en sélectionnant les projections des j qui ont les plus fortes parts d'inertie intra. Ceci permet de détecter les individus qui s'écartent non plus globalement de la structure commune représentée par un axe mais du point de vue d'un ou de quelques groupes seulement.

5.3 Cas des modalités

En analyse factorielle, lorsque les individus sont nombreux, cas des enquêtes par exemple, ils ne sont pas examinés directement mais au travers des modalités des variables nominales, que celles-ci soient actives ou illustratives. Ainsi :

- en ACP, on représente chaque modalité par le centre de gravité des individus qui la possèdent ;
- en ACM, les points représentant les modalités ne coïncident avec ces centres de gravité qu'à un coefficient près variant axe par axe (application de la fameuse propriété barycentrique de l'AFC au TDC).

En AFM, on représente toujours les modalités par les centres de gravité des individus. Ceci permet de traiter les modalités comme des individus. En particulier, on peut représenter chaque modalité (par exemple la catégorie *étudiant*) par un point global (les cdg des étudiants) et un point partiel par groupe (par exemple le cdg des points partiels des étudiants pour le groupe j).

6. L'AFM est une analyse multicanonique

Lorsque l'on confronte plusieurs groupes de variables, on cherche à répondre à la question très générale : quels sont les facteurs communs aux différents groupes de variables ?

Dans le cas simple de deux groupes, ce problème est celui de l'analyse canonique. Lorsque l'on s'intéresse à plus de deux groupes, on parle d'analyse multicanonique. Plusieurs méthodes d'analyse multicanonique existent. La plus utilisée est celle due à J.D. CARROLL, dans laquelle on procède en deux temps :

- recherche d'une suite de variables, orthogonales entre elles, les plus liées à l'ensemble des groupes de variables (variables générales) ;
- recherche, pour chaque variable générale s et dans chaque groupe j , de la combinaison linéaire des variables du groupe j la plus liée à la variable générale s (variable canonique notée plus loin F_s^j).

L'AFM peut être interprétée selon cette grille d'analyse.

Pour cela, on définit d'abord un indice de liaison entre une variable continue z et un groupe de variables $K_j = \{v_k, k=1, K_j\}$ par :

$$L_g(z, \{v_k, k=1, K_j\}) = L_g(z, K_j) = \text{inertie projetée des } v_k \text{ sur } z.$$

Pour un groupe nominal, les v_k sont les indicatrices, codées et pondérées comme en ACM. Lorsque les variables sont continues centrées réduites et de poids p_k :

$$L_g(z, K_j) = \sum_k p_k [r(z, v_k)]^2$$

Cet indice est celui utilisé en ACP lorsque l'on dit que la première composante principale est la combinaison linéaire des variables la plus liée à l'ensemble des variables.

Lorsque $L_g(z, K_j)=0$, la variable z est non corrélée à chacune des variables du groupe K_j .

Du fait de la pondération de l'AFM, $L_g(z, K_j) \leq 1$ avec $L_g(z, K_j)=1$ si z est la première composante principale de K_j .

Le premier facteur de l'AFM rend maximum l'inertie projetée de l'ensemble des variables actives, soit :

$$\sum_j L_g(z, K_j) \text{ maximum}$$

En ce sens, les facteurs de l'AFM peuvent être vus comme les variables générales d'une analyse multicanonique (dans la méthode de Carroll, la liaison entre une variable et un groupe est définie par le coefficient de corrélation multiple).

Toujours selon cette grille d'analyse, en AFM, les facteurs des nuages partiels (projection des N_j^i sur les variables générales) peuvent être interprétés comme des variables canoniques.

6.1 En pratique

En résumé, en AFM, les facteurs de l'analyse globale (notés F_s) constituent les facteurs communs et les facteurs des nuages partiels (notés F_s^j) constituent les représentations des facteurs communs dans les groupes. Pour juger du caractère véritablement commun (aux différents groupes) des facteurs de l'analyse globale, on calcule, pour chaque groupe j et chaque facteur s , le coefficient de corrélation entre la variable générale F_s et la variable canonique F_s^j . Lorsque ce coefficient (toujours positif) est élevé, alors la structure mise en évidence par la variable générale de rang s « existe » dans le groupe j . Sinon elle ne s'y trouve pas. Le bilan de ces coefficients de corrélation se fait à l'aide d'un tableau dont un exemple est donné ci-dessous.

	axes					
	1	2	3	4	5	6
Groupe 1 (olfaction au repos)	.89	.96	.89	.48	.42	.27
Groupe 2 (vision)	.93	.22	.16	.22	.17	.08
Groupe 3 (olfaction après agitation)	.97	.89	.90	.57	.66	.49
Groupe 4 (gustation)	.95	.87	.30	.25	.52	.57

Exemple test : les vins de l'essai terroir de l'INRA d'Angers. Corrélations, pour les 6 premiers axes de l'AFM, entre les projections du nuage global et celles de chacun des 4 nuages associés à un seul groupe actif.

Dans cet exemple, on considère qu'il y a trois facteurs communs :

- le premier, commun aux 4 groupes ;
- le second, commun à 3 groupes seulement (groupes 1, 3 et 4) ;
- le 3^e, commun à 2 groupes seulement (groupes 1 et 3).

Remarque. Il n'existe pas actuellement de valeur-test associée à ces coefficients de corrélation ; ils se servent mutuellement de références.

Cet exemple montre comment l'AFM peut mettre en évidence des facteurs communs à tous les groupes et/ou des facteurs communs à certains d'entre eux et/ou des facteurs spécifiques de certains groupes.

7. Etude globale des groupes

On souhaite quelquefois étudier globalement les groupes de variables avec des questions du type : ces deux groupes définissent-ils des structures analogues sur les individus c'est-à-dire sur les nuages N_j^i ? On retrouve le problème de la représentation simultanée et celui de la recherche des facteurs communs sauf que cette fois on désire un critère global de ressemblance entre les groupes.

8. Deux mesures de liaison entre deux groupes de variables

Une mesure classique de ressemblance entre deux groupes de variables K_j et K_l est le coefficient RV dû à Y. Escoufier, toujours compris entre 0 et 1 et tel que :

- $RV(K_j, K_l)=0$ si chaque variable du groupe j est non corrélée avec chaque variable du groupe l ;
- $RV(K_j, K_l)=1$ si les deux nuages N_j et N_l sont homothétiques.

Cette mesure est complétée par la mesure \mathbf{L}_g , déjà présentée pour mesurer la liaison entre une variable et un groupe de variables, qui se généralise à la liaison entre deux groupes de variables $K_j=\{z_k, k=1, K_j\}$ et $K_l=\{v_t, t=1, K_l\}$ ainsi (en notant p_k le poids de la variable k) :

$$\mathbf{L}_g(K_j, K_l)=\sum_k p_k \mathbf{L}_g(z_k, K_l)$$

Soit, dans le cas particulier de variables continues centrées réduites :

$$\mathbf{L}_g(K_j, K_l)=\sum_k \sum_t p_k p_t r^2(z_k, v_t)$$

vaut 0 si chaque variable du groupe j est non corrélée avec chaque variable du groupe l . Il est d'autant plus grand que chaque variable d'un groupe est plus liée à chaque variable de l'autre groupe.

Cet indicateur n'est pas borné. $\mathbf{L}_g(K_j, K_l)$ est d'autant plus grand que K_j et K_l sont multidimensionnels **et** présentent des directions communes nombreuses et proches des directions d'inertie importante de chaque groupe. C'est en quelque sorte un « nombre de dimensions communes », chacune étant « pondérée » par son inertie.

9. Représentation des groupes

Pour chaque groupe de variables, on peut calculer la matrice des distances entre individus soit un ensemble de P nombres. A cette matrice de distances correspond, via un passage par la matrice des produits scalaires induite (passage qui joue le rôle d'une normalisation des matrices de distances), un point dans un espace à P dimensions (noté R^P). Dans cet espace, à un groupe correspond un point et à l'ensemble des groupes correspond un nuage de J points noté N_J . Dans ce nuage N_J , deux points j et l sont d'autant plus proches que les structures conférées par les deux groupes K_j et K_l sur les individus se ressemblent. D'où l'idée d'examiner une image approchée de N_J .

L'image approchée fournie par l'AFM est obtenue en projetant N_J , non pas sur ses axes principaux d'inertie (comme le fait STATIS) mais sur les vecteurs de R^P induit par les facteurs de l'analyse globale (le facteur de rang s peut être vu comme un groupe ne comportant qu'une variable ; on peut lui associer une matrice de distances entre individus et donc un vecteur de R^P).

L'intérêt immédiat de cet espace de projection est que les axes sur lesquels on projette N_j sont interprétables et surtout possèdent la même interprétation que les axes de l'analyse globale (de même que, du fait de la dualité de l'analyse factorielle, l'axe de rang s sur lequel on projette les individus et l'axe de rang s sur lequel on projette les variables ont la même interprétation).

Cette représentation possède la propriété suivante : la coordonnée du groupe j sur l'axe de rang s est égale à $L_g(z_s, K_j)$. Ainsi :

- les coordonnées des groupes sont toujours comprises entre 0 et 1 ;
- la proximité entre deux groupes le long de l'axe s signifie que ces deux groupes possèdent la structure exprimée par le facteur de rang s avec la même intensité. Autrement dit, la représentation des groupes fait apparaître les groupes « semblables » du point de vue des facteurs communs de l'analyse globale.

Cette représentation, présentée ici surtout comme une aide à l'interprétation des représentations des individus et des variables, possède aussi sa propre optimalité : les axes sur lesquels on projette N_j , compte tenu de la contrainte usuelle d'être orthogonaux entre eux et de celle d'être de rang 1 (i.e. induit par une direction de R^I), rendent maximum la somme (et non la somme des carrés) des coordonnées (somme qui est égale aux valeurs propres de l'analyse globale). Il en résulte que la contribution du groupe j à l'axe s (contribution égale à la somme des contributions « usuelles » des variables du groupe j à l'axe s) est égale à la coordonnée du groupe rapportée à la somme des coordonnées.

Cette représentation est complétée par les cosinus carrés calculés dans l'espace R^{I^2} .

10. Liens entre l'analyse globale et les analyses séparées des groupes

Il est toujours important et intéressant de relier les résultats de l'AFM avec ceux des analyses séparées des groupes.

Pour cela, on projette, comme des variables continues illustratives, les facteurs des analyses séparées (dits axes ou facteurs « partiels »).

Remarque. Il est équivalent de réaliser l'AFM sur les variables initiales ou sur les facteurs des analyses séparées (chacun étant, au sein de son groupe, « pré-affecté » d'un poids proportionnel à la valeur propre qui lui est associée). Ainsi :

- pour chaque axe partiel, le rapport entre son inertie projetée le long de l'axe s et la valeur propre associée à l'axe s s'interprète, pour un groupe actif, véritablement comme une contribution à l'axe de rang s de l'AFM ;
- l'AFM apparaît comme une méthode fournissant une représentation optimale des facteurs des analyses séparées.

Ce dernier point est important pour les applications : l'AFM est un outil très commode pour comparer les résultats d'analyses factorielles portant sur les mêmes individus.

Par exemple, si l'on veut comparer, pour un même ensemble de variables quantitatives, l'ACP normée et l'ACP non normée, on fait apparaître deux fois les variables dans la base de données, une première fois en tant que groupe continu réduit et une seconde en tant que groupe continu non réduit ; l'AFM est alors un outil optimal de visualisation des facteurs des deux analyses.

11. Bibliographie sommaire

Escofier B. & Pagès J. (1988-1990-1993) Analyses factorielles simples et multiples ; objectifs, méthodes et interprétation. 267 p, Dunod, Paris. Cet ouvrage a été traduit en Espagnol en 1992 (Presses Universitaires de Bilbao).

Pagès J. (1996). Eléments de comparaison entre l'Analyse Factorielle Multiple et la méthode STATIS. *Revue de Statistique appliquée* XLIV (4) 81-95

Pagès J. (1996). Quelques apports de l'AFM à l'analyse de données sensorielles. *J. int. Sci. Vigne Vin.* **30**, n°4

Exemple 1 : Analyse Factorielle Multiple

Ces listages sont extraits de l'exemple des vins de l'essai terroir de l'INRA d'Angers traité dans : B. Escofier et J. Pagès ; analyses factorielles simples et multiples ; Dunod.

21 vins sont décrits à l'aide de 29 variables réparties en 5 groupes : olfaction au repos (5 variables), aspect visuel (3 variables), olfaction après agitation (10 variables), gustation (9 variables), impression d'ensemble (2 variables). Les valeurs sont les moyennes des notes attribuées par 36 juges à chacun des vins pour chacune des variables : les variables sont donc continues ; elles seront centrées et réduites. L'origine des vins est décrite par 2 variables nominales : l'appellation (3 modalités) et le terroir (4 modalités). Les groupes *impression d'ensemble* et *origine* sont introduits en illustratif, les 4 autres étant actifs.

1. Matrice des corrélations et/ou des covariances entre colonnes actives

Seules les colonnes actives (variables continues, ou indicatrices pour les variables nominales) interviennent. Il n'est pas fait ici de distinction entre les variables continues et les indicatrices quand il y en a. Seules les valeurs concernant les variables continues sont facilement interprétables (covariance ou coefficient de corrélation).

MATRICE DES CORRELATIONS ENTRE COLONNES ACTIVES																	
	Rint	Rqlt	Rfru	Rflo	Répi	Vint	Vnua	Vsur	Oint	Oqlt	Ofru	Oflo	Oépi	Ovég	Ophé	Oiab	Oper
Rint	1.00																
Rqlt	0.55	1.00															
Rfru	0.40	0.77	1.00														
Rflo	0.10	0.23	0.19	1.00													
Répi	0.53	-0.11	-0.11	-0.28	1.00												
Vint	0.61	0.63	0.46	0.30	0.25	1.00											
Vnua	0.58	0.60	0.43	0.29	0.24	0.99	1.00										
Vsur	0.53	0.75	0.64	0.32	0.06	0.89	0.88	1.00									
Oint	0.86	0.43	0.29	0.24	0.47	0.62	0.58	0.58	1.00								
Oqlt	0.13	0.70	0.71	0.34	-0.40	0.61	0.60	0.78	0.20	1.00							
Ofru	0.23	0.76	0.88	0.15	-0.29	0.50	0.46	0.70	0.21	0.83	1.00						
Oflo	-0.23	0.00	-0.07	0.75	-0.30	0.00	-0.04	0.08	-0.03	0.22	0.05	1.00					
Oépi	0.50	0.16	0.13	-0.16	0.63	0.38	0.42	0.27	0.46	-0.10	-0.11	-0.46	1.00				
Ovég	0.21	-0.45	-0.43	-0.34	0.36	-0.29	-0.31	-0.46	0.20	-0.57	-0.55	-0.45	0.13	1.00			
Ophé	0.36	0.17	0.15	0.26	0.08	0.44	0.40	0.38	0.55	0.26	0.12	-0.02	0.32	0.21	1.00		
Oiab	0.61	0.81	0.73	0.42	-0.06	0.75	0.73	0.88	0.65	0.70	0.71	0.10	0.22	-0.40	0.45	1.00	
Oper	0.59	0.71	0.60	0.56	0.13	0.79	0.77	0.84	0.69	0.67	0.57	0.30	0.36	-0.38	0.47	0.85	1.00
Oqab	0.01	0.65	0.68	0.53	-0.36	0.50	0.50	0.67	0.11	0.83	0.74	0.43	-0.11	-0.83	0.09	0.66	0.67
Gatt	0.46	0.59	0.43	0.35	0.17	0.78	0.76	0.77	0.57	0.60	0.54	0.31	0.27	-0.50	0.19	0.63	0.83
Gaci	0.14	-0.08	0.10	-0.30	0.32	-0.22	-0.20	-0.22	0.00	-0.10	-0.07	-0.27	0.24	0.37	0.03	-0.25	-0.01
Gast	0.70	0.57	0.36	0.22	0.49	0.88	0.87	0.79	0.74	0.46	0.28	-0.03	0.58	-0.16	0.43	0.66	0.78
Galc	0.51	0.64	0.51	0.14	0.07	0.64	0.62	0.81	0.64	0.64	0.54	0.00	0.33	-0.27	0.36	0.78	0.75
Géqu	0.05	0.64	0.54	0.50	-0.36	0.57	0.55	0.71	0.17	0.78	0.73	0.51	-0.07	-0.82	0.18	0.67	0.69
Gvel	0.14	0.67	0.52	0.52	-0.31	0.68	0.67	0.76	0.28	0.80	0.69	0.46	-0.03	-0.73	0.19	0.70	0.74
Game	0.63	0.22	0.02	-0.12	0.54	0.43	0.42	0.48	0.61	0.04	0.08	-0.17	0.48	0.15	0.38	0.31	0.43
Gifb	0.50	0.72	0.56	0.40	0.12	0.81	0.80	0.84	0.57	0.69	0.58	0.25	0.40	-0.57	0.29	0.76	0.91
Ghar	0.30	0.67	0.54	0.52	-0.14	0.80	0.80	0.84	0.39	0.77	0.66	0.39	0.13	-0.70	0.25	0.77	0.83
	Rint	Rqlt	Rfru	Rflo	Répi	Vint	Vnua	Vsur	Oint	Oqlt	Ofru	Oflo	Oépi	Ovég	Ophé	Oiab	Oper
	Oqab	Gatt	Gaci	Gast	Galc	Géqu	Gvel	Game	Gifb	Ghar							
Oqab	1.00																
Gatt	0.64	1.00															
Gaci	-0.23	-0.11	1.00														
Gast	0.37	0.72	-0.01	1.00													
Galc	0.49	0.67	0.01	0.72	1.00												
Géqu	0.87	0.73	-0.38	0.38	0.48	1.00											
Gvel	0.86	0.83	-0.42	0.49	0.54	0.94	1.00										
Game	-0.14	0.39	0.21	0.55	0.52	0.01	0.04	1.00									
Gifb	0.70	0.93	-0.10	0.79	0.75	0.77	0.82	0.41	1.00								
Ghar	0.81	0.88	-0.37	0.63	0.62	0.91	0.95	0.17	0.91	1.00							
	Oqab	Gatt	Gaci	Gast	Galc	Géqu	Gvel	Game	Gifb	Ghar							

2. Diagrammes des inerties (= valeurs propres) des analyses séparées

RESULTATS DES ANALYSES PAR GROUPE

GROUPE 1 (ACP NORMEE)

STATISTIQUES SOMMAIRES DES VARIABLES CONTINUES

EFFECTIF TOTAL : 21 POIDS TOTAL : 21.00

NUM . IDEN - LIBELLE	EFFECTIF	POIDS	MOYENNE	ECART-TYPE	MINIMUM	MAXIMUM
3 . Rint - Int. olfactive rep.	21	21.00	3.11	0.28	2.64	3.71
4 . Rqlt - Qlt aromatique rep.	21	21.00	3.05	0.20	2.59	3.43
5 . Rfru - Note fruitée repos	21	21.00	2.71	0.20	2.38	3.15
6 . Rflo - Note florale repos	21	21.00	2.06	0.14	1.83	2.28
7 . Répi - Note épicée repos	21	21.00	1.99	0.23	1.67	2.67

MATRICE DES CORRELATIONS

	Rint	Rqlt	Rfru	Rflo	Répi
Rint	1.00				
Rqlt	0.55	1.00			
Rfru	0.40	0.77	1.00		
Rflo	0.10	0.23	0.19	1.00	
Répi	0.53	-0.11	-0.11	-0.28	1.00

VALEURS PROPRES

APERCU DE LA PRECISION DES CALCULS : TRACE AVANT DIAGONALISATION .. 5.0000

SOMME DES VALEURS PROPRES 5.0000

HISTOGRAMME DES 5 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	2.2419	44.84	44.84	*****
2	1.5164	30.33	75.17	*****
3	0.8154	16.31	91.47	*****
4	0.2819	5.64	97.11	*****
5	0.1444	2.89	100.00	*****

...

GROUPE 5 (ACM)
 APUREMENT DES MODALITES
 SEUIL (PCMIN) :
 AVANT APUREMENT : 2 QUESTIONS 7 MODALITES ASSOCIEES
 APRES : 2 QUESTIONS 7 MODALITES ASSOCIEES
 POIDS TOTAL DES INDIVIDUS ACTIFS : 21.00
 TRI-A-PLAT DES QUESTIONS

IDENT	MODALITES LIBELLE	AVANT APUREMENT EFF. POIDS	EFF. POIDS	APRES APUREMENT POIDS	HISTOGRAMME DES POIDS RELATIFS
1 . Appellation					
	Saum - Saumur	11 11.00	11	11.00	*****
	Bour - Bourgueil	6 6.00	6	6.00	*****
	Chin - Chinon	4 4.00	4	4.00	*****
2 . Terroir					
	Séqr - Séq. de référence	7 7.00	7	7.00	*****
	Mil1 - Milieu 1	7 7.00	7	7.00	*****
	Mil2 - Milieu 2	5 5.00	5	5.00	*****
	Mil4 - Milieu 4	2 2.00	2	2.00	*****

VALEURS PROPRES

APERCU DE LA PRECISION DES CALCULS : TRACE AVANT DIAGONALISATION .. 2.5000
 SOMME DES VALEURS PROPRES 2.5000

HISTOGRAMME DES 5 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.7248	28.99	28.99	*****
2	0.6388	25.55	54.55	*****
3	0.5000	20.00	74.55	*****
4	0.3612	14.45	88.99	*****
5	0.2752	11.01	100.00	*****

EDITION SOMMAIRE DES VALEURS PROPRES SUIVANTES

6 = 0.0000 7 = 0.0000

Pour les variables continues, ces analyses sont des ACP, normées (c'est le cas ici) ou non (la réduction est commandée par le type du groupe).

Dans le cas de variables nominales (ici le cinquième groupe), ces analyses sont des ACM.

3. Matrice des corrélations entre facteurs partiels

MATRICE DES CORRELATIONS ENTRE FACTEURS PARTIELS (GGFF AVEC G = GROUPE ET F = FACTEUR)

	101	102	103	104	105	201	202	203	301	302	303	304	305	401	402	403	404
101	1.00																
102	0.00	1.00															
103	0.00	0.00	1.00														
104	0.00	0.00	0.00	1.00													
105	0.00	0.00	0.00	0.00	1.00												
201	-0.73	0.09	-0.15	0.05	0.18	1.00											
202	-0.28	-0.38	0.29	-0.11	0.00	0.00	1.00										
203	0.18	-0.11	0.02	-0.01	-0.03	0.00	0.00	1.00									
301	-0.81	-0.34	-0.05	-0.11	0.15	0.80	0.41	-0.09	1.00								
302	-0.27	0.67	-0.20	0.17	-0.25	0.32	-0.28	-0.06	0.00	1.00							
303	-0.17	0.17	0.80	-0.13	-0.01	0.10	0.07	0.36	0.00	0.00	1.00						
304	0.02	0.22	-0.33	-0.19	0.49	0.08	-0.19	0.35	0.00	0.00	0.00	1.00					
305	0.29	-0.29	0.05	-0.17	0.14	-0.04	-0.12	0.18	0.00	0.00	0.00	0.00	1.00				
401	-0.71	-0.12	-0.15	0.04	0.30	0.88	0.21	-0.06	0.91	0.10	-0.04	0.23	-0.05	1.00			
402	-0.17	0.74	-0.01	0.07	-0.11	0.15	-0.03	0.02	-0.10	0.69	0.09	0.05	-0.13	0.00	1.00		
403	0.12	0.26	-0.21	0.39	-0.01	0.16	-0.05	-0.07	-0.15	0.30	-0.12	0.01	-0.14	0.00	0.00	1.00	
404	-0.15	0.01	0.05	-0.08	-0.03	0.12	-0.14	0.06	0.07	0.18	0.15	-0.03	-0.04	0.00	0.00	0.00	1.00
405	0.03	0.41	-0.42	-0.14	0.25	0.16	-0.77	0.01	-0.18	0.18	-0.13	0.40	0.03	0.00	0.00	0.00	0.00
501	-0.02	-0.35	-0.04	-0.40	0.60	-0.03	0.08	0.06	0.14	-0.53	0.11	0.56	0.15	0.13	-0.55	-0.09	0.11
502	-0.36	0.49	-0.34	0.39	0.14	0.38	-0.29	-0.19	0.03	0.58	-0.06	0.21	-0.28	0.21	0.24	0.60	0.02
503	0.55	0.11	-0.23	-0.05	-0.22	-0.42	-0.16	0.06	-0.50	-0.05	-0.25	0.04	0.00	-0.48	-0.23	0.50	0.15
504	0.05	0.44	0.02	-0.59	-0.19	-0.10	0.19	0.16	-0.17	0.27	0.25	0.16	-0.19	-0.20	0.46	0.05	0.00
505	-0.17	-0.42	-0.45	-0.23	-0.29	0.21	0.04	0.02	0.37	0.01	-0.40	-0.13	0.11	0.25	-0.37	-0.03	-0.01
601	0.55	0.48	0.03	0.03	-0.35	-0.65	-0.36	0.09	-0.87	0.34	0.08	-0.10	0.04	-0.87	0.37	0.17	0.09
602	-0.04	0.05	0.14	0.17	-0.24	0.12	-0.03	0.22	0.00	0.08	0.15	-0.41	-0.30	-0.03	-0.02	0.05	0.13
	101	102	103	104	105	201	202	203	301	302	303	304	305	401	402	403	404
	405	501	502	503	504	505	601	602									
405	1.00																
501	0.18	1.00															
502	0.41	0.00	1.00														
503	0.11	0.00	0.00	1.00													
504	-0.11	0.00	0.00	0.00	1.00												
505	0.01	0.00	0.00	0.00	0.00	1.00											
601	0.13	-0.29	0.13	0.43	0.37	-0.28	1.00										
602	0.00	-0.37	-0.06	0.08	-0.02	0.06	0.00	1.00									
	405	501	502	503	504	505	601	602									

Les facteurs partiels sont les facteurs des analyses séparées. 102 : deuxième facteur de l'analyse du groupe 1. Le nombre de facteurs par groupe est défini dans le menu Paramètres.

Les facteurs d'une même analyse sont non corrélés. Les facteurs étant définis au signe près, le signe des coefficients de corrélation ne s'interprète pas. Par exemple, les premiers facteurs des groupes 1 et 3 ont un coefficient de corrélation de .81

4. Diagramme des inerties (= valeurs propres) de l'analyse globale

Compte tenu de la pondération, la première valeur propre a pour valeur maximum le nombre de groupes actifs (ici 4). Ainsi, la valeur 3.4619 indique que le premier facteur de l'analyse globale correspond à une direction d'inertie importante de chacun des groupes.

ANALYSE GLOBALE
VALEURS PROPRES

APERCU DE LA PRECISION DES CALCULS : TRACE AVANT DIAGONALISATION .. 7.0111
SOMME DES VALEURS PROPRES 7.0111

HISTOGRAMME DES 5 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	3.4619	49.38	49.38	*****
2	1.3668	19.49	68.87	*****
3	0.6154	8.78	77.65	*****
4	0.3722	5.31	82.96	*****
5	0.2704	3.86	86.82	*****

5. Matrices des liaisons entre groupes

L'indice de liaison $L_g(j, l)$ entre 2 groupes j et l est la somme, pondérée par les poids des variables de l'AFM, des carrés des covariances (coefficients de corrélation pour les variables continues réduites) entre chaque colonne (variable continue ou indicatrice) du groupe j et chaque colonne du groupe l .

Si les groupes j et l comportent chacun une seule variable continue, L_g est le carré du coefficient de corrélation entre les deux variables. Si les groupes j et l comportent l'un une seule variable continue et l'autre une seule variable nominale, L_g est le rapport de corrélation entre les deux variables. Si les groupes j et l comportent chacun une seule variable nominale, L_g est le critère χ^2 (divisé par l'effectif total l) du tableau de contingence croisant les deux variables.

$L_g(j, l)$ vaut 0 si chaque colonne du groupe j est non corrélée avec chaque colonne du groupe l . Il est d'autant plus grand que les deux groupes ont des directions d'inertie importante en commun.

COEFFICIENTS L _g DE LIAISON ENTRE GROUPES						
	1	2	3	4	5	6
1	1.61					
2	0.55	1.00				
3	1.05	0.70	1.37			
4	0.68	0.80	0.94	1.12		
5	0.85	0.28	0.71	0.51	2.65	
6	0.47	0.43	0.82	0.81	0.33	1.01

COEFFICIENTS RV DE LIAISON ENTRE GROUPES						
	1	2	3	4	5	6
1	1.00					
2	0.44	1.00				
3	0.71	0.60	1.00			
4	0.51	0.75	0.76	1.00		
5	0.41	0.17	0.37	0.29	1.00	
6	0.37	0.42	0.70	0.76	0.20	1.00

Sur la diagonale, l'indice de liaison (associé à un seul groupe) est égal à la somme des carrés des valeurs propres du groupe (en tenant compte de la pondération de l'AFM), valeur qui s'interprète comme un indicateur de multidimensionnalité dont la racine est notée $\mathbf{N}_g(j) = \sqrt{\mathbf{L}_g(j,j)}$. Ici, par exemple, le groupe 2 est pratiquement unidimensionnel ($\mathbf{N}_g = 1.00$).

$RV(j,l) = \mathbf{L}_g(j,l) / [\mathbf{N}_g(j) \mathbf{N}_g(l)]$; ce coefficient, toujours compris entre 0 et 1, vaut 1 si les nuages d'individus induits par les deux groupes sont homothétiques.

Exemple d'utilisation conjointe des 2 indices :

- $RV(1,3) = .71 \approx RV(2,4) = .75$: les groupes 1 et 3 d'une part et 2 et 4 d'autre part ont des structures voisines, également proches de l'homothétie.
- $\mathbf{L}_g(1,3) = 1.05 > \mathbf{L}_g(2,4) = .80$: la structure commune aux groupes 1 et 3 est plus riche que la structure commune aux groupes 2 et 4.

6. Aides à l'interprétation des groupes

COORDONNEES ET AIDES A L'INTERPRETATION DES GROUPES ACTIFS
AXES 1 A 5

		COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
GRP.	P.REL DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
GR 1	0.25 1.61	0.78	0.62	0.37	0.17	0.09	22.6	45.3	60.7	46.4	31.6	0.38	0.24	0.09	0.02	0.00
GR 2	0.25 1.00	0.85	0.04	0.01	0.05	0.03	24.7	2.9	2.3	12.2	11.0	0.73	0.00	0.00	0.00	0.00
GR 3	0.25 1.37	0.92	0.47	0.18	0.10	0.12	26.7	34.3	29.3	27.2	42.9	0.62	0.16	0.02	0.01	0.01
GR 4	0.25 1.12	0.90	0.24	0.05	0.05	0.04	26.0	17.4	7.7	14.2	14.5	0.72	0.05	0.00	0.00	0.00
ENSEMBLE							100.0	100.0	100.0	100.0	100.0	0.59	0.13	0.03	0.01	0.00

COORDONNEES ET AIDES A L'INTERPRETATION DES GROUPES ILLUSTRATIFS
AXES 1 A 5

		COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
GRP.	P.REL DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
GR 5	0.25 2.65	0.30	0.64	0.20	0.14	0.29	8.5	47.0	31.9	37.7	108.1	0.03	0.16	0.01	0.01	0.03
GR 6	0.25 1.01	0.62	0.25	0.01	0.04	0.01	17.9	18.6	1.6	11.0	5.3	0.38	0.06	0.00	0.00	0.00
ENSEMBLE							26.4	65.6	33.5	48.7	113.4	0.13	0.13	0.01	0.01	0.02

Ces valeurs correspondent à l'analyse dans l'espace des groupes de variables (R^2).

DISTO : carré de la distance, dans R^2 , entre le point représentant un groupe et l'origine ; on retrouve la mesure $[\mathbf{N}_g]^2$ (cf. § 5).

P. REL : inverse du nombre de groupes actifs ; en AFM tous les groupes ont le même poids.

Les coordonnées sont toujours comprise entre 0 et 1. La coordonnée du groupe j pour l'axe s s'interprète aussi, dans l'espace des variables (de dimension l), comme l'inertie projetée de l'ensemble des variables du groupe j pour l'axe s (la somme, pour tous les groupes actifs, des coordonnées sur l'axe s est égale à l'inertie de l'axe s de l'analyse globale). Ainsi cette coordonnée vaut 1 si l'axe s de l'analyse globale coïncide avec la première composante principale du groupe j . Ici, par exemple, le premier facteur correspond à une direction d'inertie importante de chaque groupe.

La contribution est égale à la coordonnée divisée par la somme des coordonnées des groupes (en %).

- Pour les groupes actifs, ce rapport s'interprète véritablement comme la contribution du groupe au facteur. Ici, par exemple, ce sont surtout les groupes 1 et 3 qui contribuent à l'axe 2. La somme des contributions des groupes actifs vaut 100.

- Pour les groupes illustratifs, ce rapport ne s'interprète pas comme une contribution au plein sens du terme. Ici, par exemple, l'inertie du groupe 6 n'est pas négligeable (relativement à l'ensemble des groupes actifs) dans le plan (1, 2) ; mettre ce groupe en actif aurait renforcé ce plan. Ce rapport peut dépasser 100 : c'est le cas du groupe 5, « très » multidimensionnel, pour l'axe 5 (de très faible inertie pour les groupes actifs).

Cosinus carrés (=inertie projetée/inertie totale). Ils mesurent la qualité de représentation d'un groupe (en tant que point unique dans l'espace des groupes) par un facteur. Pour l'ensemble, on donne la qualité de représentation du nuage N_j de l'ensemble des groupes par un facteur. Ici, par exemple, le premier plan représente 72% (=59%+13%) de l'inertie totale de N_j . Les autres axes n'ajoutent pratiquement rien à la représentation de N_j (attention : même en gardant tous les axes, la représentation de N_j est en général imparfaite du fait de la contrainte d'être de rang 1 imposée aux axes de R^J).

7. Corrélations entre les variables canoniques et les variables générales

CORRELATIONS ENTRE LES VARIABLES CANONiques ET LES FACTEURS DE L'ANALYSE GLOBALE
AXES 1 A 5

CORRELATIONS					
FAC.	1	2	3	4	5
GR 1	0.89	0.96	0.89	0.48	0.42
GR 2	0.93	0.22	0.16	0.22	0.17
GR 3	0.97	0.89	0.90	0.57	0.66
GR 4	0.95	0.87	0.30	0.25	0.52

La s^e variable canonique du groupe j est la projection du nuage N_j^j (nuage des individus décrits uniquement par le groupe j) sur le s^e axe de l'analyse globale. La s^e variable générale est le s^e facteur sur les individus de l'analyse globale. Le coefficient de corrélation (dit canonique en AFM) entre ces deux facteurs montre dans quelle mesure le facteur de l'analyse globale peut être considéré comme une structure du groupe j . Ces coefficients sont toujours positifs ou nuls.

Ici, le premier facteur est commun aux quatre groupes ; le deuxième facteur aux trois groupes 1, 3 et 4 seulement, etc. Ces coefficients s'interprètent conjointement aux coordonnées des groupes : ici, par exemple, le facteur 2 de l'AFM est « présent » dans les groupes 1, 3 et 4 mais ne correspond à une direction d'inertie importante que pour les groupes 1 et 3.

8. Rapports (inertie inter)/(inertie totale)

RAPPORT : INERTIE INTER/INERTIE TOTALE
AXES 1 A 5

	RAPPORTS				
FAC.	1	2	3	4	5
	0.87	0.58	0.38	0.14	0.17

Ce rapport mesure l'intérêt global de la représentation simultanée de l'ensemble des nuages N_j^i . L'ensemble des points de ces nuages peut être partitionné en regroupant dans une même classe les points qui représentent un même individu (chaque classe contient autant de points que de groupes actifs ; il y a autant de classes que d'individus actifs). A cette partition, on applique le théorème de Huygens : Inertie totale=inertie inter+inertie intra.

Le rapport (inertie inter)/(inertie totale) se réfère à cette décomposition. Il montre dans quelle mesure chaque axe met en évidence une structure commune aux groupes. En cas de structure commune, les différentes représentations d'un même individu sont proches, ce qui se traduit par un rapport (inertie inter)/(inertie totale) élevé.

Ici, chacun des 3 premiers facteurs correspond à une structure commune ; on retrouve sous une autre forme le fait que cette structure est commune à tous les groupes uniquement pour l'axe 1.

Ce rapport ne décroît pas forcément avec le rang de l'axe (cf. axes 4 et 5) puisque ce n'est pas le critère utilisé mais, en pratique, décroît pour les tout premiers axes. Il ne s'additionne pas d'un axe à l'autre.

9. Aide à l'interprétation des individus actifs

COORDONNEES ET AIDES A L'INTERPRETATION DES INDIVIDUS ACTIFS
AXES 1 A 5

INDIVIDUS			COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
IDENT./GROUPE		POIDS DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1VAU		4.76 21.97	4.4	-1.0	0.0	0.3	0.5	26.4	3.4	0.0	1.2	4.0	0.87	0.05	0.00	0.00	0.01
	POUR GROUPE 1	72.15	3.9	-1.1	-2.4	-4.7	-1.1	0.6	0.0	7.2	13.0	2.3	0.21	0.02	0.08	0.30	0.02
	2	72.84	4.2	1.2	0.9	2.8	2.8	0.1	6.0	0.8	3.2	4.7	0.25	0.02	0.01	0.10	0.10
	3	91.68	4.7	-2.3	-0.8	-0.6	-0.7	0.3	2.0	0.8	0.5	1.2	0.24	0.06	0.01	0.00	0.01
	4	114.84	4.7	-1.8	2.5	3.8	1.0	0.2	0.8	7.1	6.3	0.2	0.19	0.03	0.05	0.12	0.01
1ING		4.76 2.58	-1.2	0.5	-0.2	-0.1	0.2	2.0	0.8	0.2	0.1	1.1	0.56	0.09	0.01	0.00	0.02
	POUR GROUPE 1	10.97	-1.2	1.0	-0.1	1.8	0.7	0.0	0.3	0.0	1.9	0.2	0.14	0.08	0.00	0.30	0.05
	2	6.86	-1.1	-0.2	-0.1	-0.6	-0.7	0.0	0.6	0.0	0.1	0.7	0.18	0.01	0.00	0.05	0.06
	3	13.97	-1.2	1.3	-0.1	-0.8	1.2	0.0	0.8	0.0	0.3	0.7	0.11	0.12	0.00	0.05	0.10
	4	9.48	-1.3	-0.1	-0.4	-0.7	-0.3	0.0	0.4	0.1	0.2	0.2	0.17	0.00	0.02	0.05	0.01
2ING		4.76 26.35	4.9	0.7	0.9	0.1	-0.1	32.8	1.6	6.0	0.2	0.1	0.91	0.02	0.03	0.00	0.00
	POUR GROUPE 1	52.09	3.1	1.6	-0.3	-3.8	-2.6	7.2	1.0	1.6	8.1	6.0	0.19	0.05	0.00	0.28	0.13
	2	107.85	5.2	1.5	1.1	3.4	3.4	0.2	0.9	0.1	5.7	10.7	0.25	0.02	0.01	0.11	0.11
	3	112.42	5.2	-0.9	0.9	-2.6	0.3	0.3	3.0	0.0	3.9	0.1	0.24	0.01	0.01	0.06	0.00
	4	149.26	6.0	0.5	1.8	3.5	-1.3	3.0	0.1	0.9	5.9	1.4	0.24	0.00	0.02	0.08	0.01
2BEA		4.76 6.16	-1.4	1.1	0.6	1.2	-0.2	2.7	4.1	2.4	17.1	0.7	0.32	0.19	0.05	0.22	0.01
	POUR GROUPE 1	47.73	-2.0	3.3	1.1	3.4	-0.1	0.8	6.1	0.4	2.7	0.0	0.08	0.23	0.03	0.25	0.00
	2	3.26	-0.1	0.1	0.2	0.1	0.0	4.0	1.1	0.2	0.6	0.0	0.00	0.00	0.01	0.00	0.00
	3	38.01	-2.9	0.8	0.6	1.0	-0.9	5.5	0.1	0.0	0.0	0.5	0.23	0.02	0.01	0.03	0.02
	4	9.53	-0.6	0.0	0.3	0.0	0.2	1.5	1.3	0.1	0.7	0.2	0.04	0.00	0.01	0.00	0.01

Chaque individu apparaît autant de fois qu'il y a de groupes actifs (individus partiels) plus 1 (individu moyen). L'individu moyen est situé au centre de gravité des individus partiels qui lui correspondent.

POIDS : commandé dans le menu pondérations ; la somme des poids des individus actifs vaut 1 ; par défaut poids =1// (ici=1/21).

DISTO : carré de la distance entre le point et l'origine. Il s'interprète comme dans toute analyse factorielle : ici, par exemple, le vin 2ING est très particulier. La somme des DISTO des individus partiels d'un même individu vaut, au coefficient $(JA)^2$ près (JA = nombre de groupes actifs) le DISTO de l'individu moyen.

Contributions

- *Pour les individus moyens*, il s'agit de la contribution au sens usuel i.e. l'inertie projetée du point rapportée à l'inertie totale de l'axe (en %). Ici, par exemple, les vins T1 et T2 contribuent pour 69% à la construction de l'axe 2. La somme des contributions pour les individus moyens actifs vaut 100.
- *Pour les individus partiels*, il s'agit de la contribution à l'inertie intra définie § 8 : elle indique dans quelle mesure l'individu partiel i^j est différent de l'individu moyen i . Ici, par exemple, le vin T2 a, du point de vue de l'axe 2, des images partielles très particulières pour les groupes 1 et 2 (contributions de 10.7 et 10.5), ce qui se retrouve dans les coordonnées, extrême pour l'une (-6.3), moyenne pour l'autre (-0.4).
- *Pour les individus illustratifs* (il n'y en a pas ici), cet indicateur ne s'interprète que comme un rapport d'inertie et non comme une véritable contribution.

Cosinus carrés = (inertie projetée du point)/(inertie totale du point). Pour un individu moyen, cet indicateur s'interprète comme usuellement : ainsi, par exemple, l'écart entre IVAU et l'origine est très bien représenté (.87) par le premier axe. Attention : pour les individus partiels, cet indicateur est par construction très faible (même si la forme du nuage partiel est bien respectée) et ne s'additionne pas d'un axe à l'autre.

10. Individus ayant les plus fortes contributions

INDIVIDUS AYANT LES PLUS FORTES CONTRIBUTIONS

AXE 1

INDIVIDUS	CTR	CUMUL
2ING	32.84	32.84
1VAU	26.40	59.24
1DAM	10.00	69.24
1CHA	5.75	74.99
2DAM	4.37	79.36

AXE 2

INDIVIDUS	CTR	CUMUL
T2	39.39	39.39
T1	29.71	69.10
1CHA	6.67	75.76
2BEA	4.07	79.83
1VAU	3.45	83.28

Sous-ensemble du tableau précédent (§ 9) des aides à l'interprétation des individus ; on a trié les individus moyens ayant les contributions les plus grandes. Par exemple, on retrouve que les vins T1 et T2 contribuent à 69% de l'inertie de l'axe 2.

Ceci est utile pour détecter les individus prépondérants lorsque les individus sont nombreux. Le cumul permet d'obtenir le plus petit nombre d'individus qui totalisent une contribution donnée à l'inertie d'un axe.

11. Individus ayant les plus fortes et les plus faibles inerties intra

INDIVIDUS AYANT LES PLUS FORTES INERTIES INTRA

AXE 1

INDIVIDUS	INER	CUMUL
2BEA	11.79	11.79
1DAM	11.40	23.20
2ING	10.67	33.87
2BOU	10.15	44.01

AXE 2

INDIVIDUS	INER	CUMUL
T2	23.07	23.07
T1	17.40	40.47
1VAU	8.86	49.33
2BEA	8.62	57.96

INDIVIDUS AYANT LES PLUS FAIBLES INERTIES INTRA

AXE 1

INDIVIDUS	INER	CUMUL
1ING	0.04	0.04
2DAM	0.24	0.28
1ROC	0.55	0.83
1TUR	1.15	1.98

AXE 2

INDIVIDUS	INER	CUMUL
PER1	0.39	0.39
4EL	0.55	0.94
DOM1	0.88	1.82
1FON	0.92	2.74

On décompose par individu l'inertie intra définie § 8. Ceci est utile pour détecter systématiquement :

- les individus de faible inertie intra, dont les images partielles sont "cohérentes" entre elles (et qui ainsi illustrent la structure commune) ; pour l'axe 1, c'est le cas de 1ING dont toutes les images partielles ont pratiquement la même coordonnée (comprise entre -1.1 et -1.3) ;
- ceux ayant une forte inertie intra, qui au contraire s'inscrivent le moins bien dans la structure commune ; pour l'axe 1, c'est le cas de 2BEA dont les coordonnées des images partielles varient de .1 pour le groupe 2 (ce vin est moyen du point de vue visuel) à 2.9 pour le groupe 3 (ce vin est particulier du point de vue de l'olfaction après agitation)

12. Individus partiels ayant les plus fortes inerties intra

INDIVIDUS PARTIELS AYANT LES PLUS FORTES INERTIES INTRA
AXE 1

INDIVIDUS		INNER	CUMUL
2ING	G 1	7.19	7.19
1DAM	G 1	6.50	13.69
2BOU	G 1	6.25	19.94
2BEA	G 3	5.50	25.45
DOMI	G 1	4.75	30.19

AXE 2

INDIVIDUS		INNER	CUMUL
T2	G 1	10.70	10.70
T2	G 2	10.45	21.15
T1	G 2	7.81	28.96
2BEA	G 1	6.14	35.09
1VAU	G 2	6.00	41.10

Ces inerties figurent déjà, sous le nom de contribution, dans les aides à l'interprétation des individus (§ 9).

Le tri est utile pour détecter les individus partiels les plus discordants par rapport à l'individu moyen qui leur est associé, et qui donc illustrent le mieux les écarts à la structure commune. Ainsi, pour l'axe 1, 2ING est particulier du point de vue du groupe 1 (sa coordonnée montre que ce point partiel est différent des autres dont les coordonnées sont comprises entre 5.2 et 6).

13. Aides à l'interprétation des modalités (centres de gravité)

Les centres de gravité (cdg) des individus présentant une même modalité pour une variable nominale se comportent comme des individus illustratifs (ainsi les classes comportant un seul individu sont identiques à cet individu). Le poids (P. REL) indiqué ici est la somme des poids des individus possédant la modalité. Ici, par exemple, les vins de la séquence de référence constituent 1/3 (=7/21) des vins.

La contribution rapporte l'inertie de la modalité (en tant que cdg) à l'inertie totale. Il ne s'agit pas d'une véritable contribution, même dans le cas où les variables nominales considérées sont actives, auquel cas cette quantité n'est que proportionnelle à la contribution au sens usuel. Prenons l'exemple des 7 vins de la séquence de référence. La contribution cumulée de ces 7 vins, calculée à partir du tableau complet dont une partie est reproduite au § 9, vaut 25%. Selon le théorème de Huygens, cette inertie peut être décomposée en inertie intra et inertie du cdg. C'est cette dernière qui vaut 19.9% : on peut dire que les caractéristiques des vins de la séquence de référence contribuent pour 19.9% au premier axe.

COORDONNEES, CONTRIBUTIONS, COSINUS CARRES DES MODALITES
AXES 1 A 5

MODALITES			COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
MODAL./GROUPE	P.REL	DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
<hr/>																	
1 . Appellation								9.8	10.6	10.1	19.0	29.2					
Saumur	52.38	0.59	-0.5	-0.4	-0.2	0.2	-0.2	4.3	4.7	4.7	3.7	11.1	0.48	0.21	0.09	0.05	0.10
POUR GROUPE 1		3.08	-0.5	-0.5	-0.2	0.8	0.2	0.0	0.3	0.0	2.3	2.3	0.10	0.08	0.01	0.20	0.02
2		2.16	-0.7	-0.2	-0.2	-0.5	-0.5	0.8	0.2	0.0	2.6	0.6	0.23	0.03	0.02	0.12	0.11
3		2.86	-0.4	-0.6	-0.3	0.7	-0.7	0.6	0.8	0.1	1.9	1.7	0.05	0.12	0.03	0.19	0.15
4		1.30	-0.5	-0.1	-0.3	-0.4	-0.1	0.0	1.1	0.0	1.6	0.3	0.19	0.00	0.05	0.11	0.00
<hr/>																	
Bourgueil	28.57	0.87	0.4	0.5	0.2	-0.4	0.4	1.3	5.3	2.2	13.6	18.0	0.18	0.29	0.05	0.20	0.20
POUR GROUPE 1		3.09	0.3	0.5	0.3	-0.8	-0.1	0.0	0.0	0.1	0.4	1.5	0.04	0.08	0.04	0.19	0.00
2		1.40	0.6	0.2	0.1	0.4	0.4	0.5	0.7	0.0	2.2	0.0	0.24	0.03	0.02	0.12	0.11
3		8.30	0.4	1.0	0.4	-1.4	1.4	0.0	1.7	0.2	2.8	5.3	0.02	0.12	0.02	0.22	0.24
4		1.17	0.2	0.3	0.0	0.0	0.0	0.5	0.2	0.3	0.7	1.1	0.04	0.10	0.00	0.00	0.00
<hr/>																	
Chinon	19.05	1.43	0.9	0.2	0.3	0.2	0.0	4.2	0.6	3.2	1.7	0.1	0.54	0.03	0.07	0.02	0.00
POUR GROUPE 1		8.62	1.0	0.6	0.0	-1.0	-0.5	0.2	0.9	0.5	3.0	1.1	0.12	0.05	0.00	0.12	0.03
2		5.19	1.1	0.4	0.3	0.8	0.7	0.4	0.1	0.0	0.8	1.8	0.22	0.03	0.02	0.12	0.11
3		2.01	0.4	0.1	0.3	0.0	-0.3	2.2	0.0	0.0	0.1	0.4	0.08	0.01	0.04	0.00	0.05
4		7.06	1.0	-0.3	0.7	0.9	0.2	0.2	1.4	0.7	1.2	0.1	0.15	0.02	0.07	0.13	0.01
<hr/>																	
2 . Terroir								33.1	82.6	18.4	1.3	13.2					
Séq. de référence	33.33	2.51	-1.4	0.6	0.2	0.0	0.2	19.9	7.8	1.5	0.0	5.0	0.82	0.13	0.01	0.00	0.02
POUR GROUPE 1		12.20	-1.5	1.0	1.0	1.8	0.5	0.0	1.3	6.1	11.4	0.5	0.17	0.07	0.09	0.25	0.02
2		4.98	-1.1	-0.3	-0.2	-0.7	-0.7	1.8	6.7	1.2	1.9	5.4	0.25	0.02	0.01	0.11	0.10
3		12.67	-1.7	1.2	0.3	-0.1	0.6	0.8	3.1	0.2	0.0	1.0	0.22	0.11	0.01	0.00	0.03
4		10.27	-1.5	0.5	-0.5	-1.0	0.4	0.1	0.1	3.4	3.3	0.4	0.23	0.02	0.02	0.09	0.02
<hr/>																	
Milieu 1	33.33	1.47	0.9	0.5	-0.5	-0.1	0.0	8.7	5.3	11.2	0.4	0.1	0.61	0.15	0.14	0.00	0.00
POUR GROUPE 1		8.37	1.1	0.8	-1.6	-0.8	-0.6	0.3	0.7	11.0	1.7	2.1	0.14	0.07	0.31	0.07	0.04
2		3.49	0.9	0.3	0.2	0.6	0.6	0.0	0.3	3.5	1.7	2.2	0.25	0.02	0.01	0.11	0.11
3		5.11	0.9	0.3	-0.5	-0.4	0.5	0.1	0.2	0.0	0.5	1.7	0.16	0.02	0.04	0.04	0.06
4		6.49	0.9	0.5	0.0	0.3	-0.5	0.1	0.0	2.1	0.5	1.8	0.12	0.04	0.00	0.01	0.04
<hr/>																	
Milieu 2	23.81	1.14	0.8	-0.2	0.4	0.0	-0.3	4.3	0.6	5.7	0.1	8.1	0.55	0.03	0.13	0.00	0.08
POUR GROUPE 1		6.70	0.9	-0.1	0.6	-1.2	-1.0	0.0	0.0	0.3	4.4	2.3	0.11	0.00	0.05	0.23	0.15
2		2.11	0.7	0.2	0.2	0.5	0.5	0.1	1.0	0.3	0.5	2.7	0.25	0.02	0.01	0.11	0.11
3		4.76	0.8	-0.4	0.2	0.2	-0.8	0.0	0.2	0.1	0.1	1.1	0.14	0.03	0.01	0.01	0.13
4		4.62	0.8	-0.5	0.5	0.7	0.1	0.0	0.6	0.1	1.3	0.8	0.14	0.05	0.06	0.12	0.00

...

COORDONNEES ET VALEURS-TEST DES MODALITES

CENTRES DE GRAVITE		VALEURS-TEST					COORDONNEES					DISTO
MODAL./GROUPE	P.REL	1	2	3	4	5	1	2	3	4	5	
1 . Appellation												
Saumur	52.38	0.3	0.4	0.3	1.1	1.7						
POUR GROUPE 1		-1.3	-1.4	-1.4	1.3	-2.2	-0.5	-0.4	-0.2	0.2	-0.2	0.59
2		-1.4	-1.1	-0.4	1.6	0.7	-0.5	-0.5	-0.2	0.8	0.2	3.08
3		-1.7	-1.8	-1.9	-1.8	-1.7	-0.7	-0.2	-0.2	-0.5	-0.5	2.16
4		-0.9	-1.5	-1.4	3.0	-2.3	-0.4	-0.6	-0.3	0.7	-0.7	2.86
		-1.1	-0.3	-1.5	-1.3	-0.5	-0.5	-0.1	-0.3	-0.4	-0.1	1.30
Bourgueil												
POUR GROUPE 1	28.57	0.6	1.2	0.8	-1.9	2.2	0.4	0.5	0.2	-0.4	0.4	0.87
2		0.5	0.6	0.4	-0.9	-0.2	0.3	0.5	0.3	-0.8	-0.1	3.09
3		0.8	0.9	0.9	0.9	0.8	0.6	0.2	0.1	0.4	0.4	1.40
4		0.6	1.5	1.0	-3.3	2.9	0.4	1.0	0.4	-1.4	1.4	8.30
		0.3	1.1	0.1	0.1	-0.2	0.2	0.3	0.0	0.0	0.0	1.17
...												
2 . Terroir												
Ség. de référence	33.33	1.5	4.7	0.5	-1.9	0.0						
POUR GROUPE 1		-2.4	1.5	0.7	0.0	1.2	-1.4	0.6	0.2	0.0	0.2	2.51
2		-2.4	1.4	1.5	2.4	1.0	-1.5	1.0	1.0	1.8	0.5	12.20
3		-1.8	-1.6	-1.5	-1.7	-1.7	-1.1	-0.3	-0.2	-0.7	-0.7	4.98
4		-2.6	2.1	1.0	-0.3	1.4	-1.7	1.2	0.3	-0.1	0.6	12.67
		-2.4	1.6	-1.9	-2.2	2.4	-1.5	0.5	-0.5	-1.0	0.4	10.27
Milieu 1												
POUR GROUPE 1	33.33	1.6	1.3	-1.8	-0.4	0.1	0.9	0.5	-0.5	-0.1	0.0	1.47
2		1.8	1.1	-2.4	-1.0	-1.1	1.1	0.8	-1.6	-0.8	-0.6	8.37
3		1.5	1.4	1.4	1.4	1.5	0.9	0.3	0.2	0.6	0.6	3.49
4		1.4	0.6	-1.4	-1.2	1.3	0.9	0.3	-0.5	-0.4	0.5	5.11
		1.4	1.7	0.2	0.7	-2.8	0.9	0.5	0.0	0.3	-0.5	6.49
Milieu 2												
POUR GROUPE 1	23.81	1.1	-0.4	1.2	0.2	-1.5	0.8	-0.2	0.4	0.0	-0.3	1.14
2		1.1	-0.1	0.7	-1.3	-1.6	0.9	-0.1	0.6	-1.2	-1.0	6.70
3		0.9	0.9	0.9	0.9	0.9	0.7	0.2	0.2	0.5	0.5	2.11
4		1.0	-0.5	0.6	0.4	-1.5	0.8	-0.4	0.2	0.2	-0.8	4.76
		1.0	-1.3	1.6	1.4	0.5	0.8	-0.5	0.5	0.7	0.1	4.62
Milieu 4												
POUR GROUPE 1	9.52	-0.2	-3.9	0.1	0.4	-0.1	-0.3	-3.1	0.1	0.2	0.0	10.16
2		-0.6	-3.8	0.4	-0.2	2.6	-0.8	-5.7	0.5	-0.4	2.8	74.38
3		-0.9	-0.9	-1.0	-0.9	-0.9	-1.2	-0.4	-0.3	-0.8	-0.8	5.93
4		0.5	-3.4	-0.3	1.8	-2.2	0.7	-4.3	-0.2	1.4	-2.0	54.17
		0.2	-3.3	0.4	0.5	-0.2	0.2	-2.2	0.2	0.5	-0.1	28.11

Les cdg sont regroupés par variables. Si l'on considère les cdg associés à une même variable nominale (et sauf données manquantes pour cette variable) :

- la somme pondérée des coordonnées des cdg est nulle, aussi bien pour les cdg moyens que pour les cdg partiels associés à un même groupe;

- la somme des poids vaut 1;
- la somme des contributions des cdg moyens est le rapport de corrélation entre cette variable et le facteur. Ici, par exemple, l'appellation est peu liée aux trois premiers facteurs ; la variable *terroir* est liée aux deux premiers facteurs, particulièrement au 2^e.

La valeur-test d'une modalité a son sens usuel de l'ACP ou de l'ACM de SPAD. C'est un indicateur qui prend en compte la coordonnée, l'effectif et l'inertie de l'axe. Il permet de comparer les coordonnées de cdg d'effectifs différents sur des axes différents. Il peut être lu en référence à une loi normale centrée réduite et de ce fait pourrait être converti en probabilité (probabilité d'observer, par tirage au hasard, un cdg de même effectif au moins aussi éloigné de l'origine). Les valeurs-tests ne sont calculées que si tous les individus ont le même poids. Ici, par exemple, les vins du milieu 4 sont caractérisés par l'axe 2, et ce pour les groupes 1, 3 et 4.

La valeur-test d'une variable pour un facteur (celle qui figure juste sous le libellé de la variable) est la conversion en variable normale centrée réduite du F de l'analyse de variance « expliquant » le facteur à partir de la variable. Ici, par exemple, on retrouve que l'appellation n'est pas liée aux premiers axes et que le terroir est lié aux deux premiers axes, en particulier au 2^e.

14. Centres de gravité ayant les plus fortes contributions

CENTRES DE GRAVITE AYANT LES PLUS FORTES CONTRIBUTIONS

AXE 1		
CENTRE DE GRAVITE	CTR	CUMUL
Séq. de référence	19.88	19.88
Milieu 1	8.67	28.55
Milieu 2	4.33	32.89

AXE 2		
CENTRE DE GRAVITE	CTR	CUMUL
Milieu 4	68.76	68.76
Séq. de référence	7.85	76.60
Milieu 1	5.32	81.92

Ce tableau reprend des données du tableau du § 13 (de même que le tableau du § 10 reprend des données du tableau du § 9).

15. Centres de gravité ayant les plus fortes et les plus faibles inerties intra

CENTRES DE GRAVITE AYANT LES PLUS FORTES INERTIES INTRA

AXE 1

CENTRE DE GRAVITE	INER	CUMUL
Milieu 4	10.44	10.44
Chinon	2.99	13.43
Séq. de référenc	2.70	16.13

AXE 2

CENTRE DE GRAVITE	INER	CUMUL
Milieu 4	39.86	39.86
Séq. de référenc	11.19	51.05
Bourgueil	2.54	53.59

CENTRE DE GRAVITE AYANT LES PLUS FAIBLES INERTIES INTRA

AXE 1

CENTRE DE GRAVITE	INER	CUMUL
Milieu 2	0.10	0.10
Milieu 1	0.41	0.52
Bourgueil	1.02	1.53

AXE 2

CENTRE DE GRAVITE	INER	CUMUL
Milieu 1	1.25	1.25
Milieu 2	1.86	3.11
Saumur	2.25	5.36

Ce tableau est analogue à celui du § 11.

Ici, par exemple, du point de vue de l'axe 2, les vins du milieu 4 sont perçus de façons très hétérogènes d'une phase à l'autre de la dégustation (on retrouve le fait que l'axe 2 met en évidence ces vins du point de vue des groupes 1, 3 et 4 uniquement).

16. Centres de gravité partiels ayant les plus fortes inerties intra

CENTRES DE GRAVITE PARTIELS AYANT LES PLUS FORTES INERTIES INTRA

AXE 1

CENTRE DE GRAVITE	INER	CUMUL
Milieu 4	3	4.06
Milieu 4	2	3.80
Chinon	3	2.23

AXE 2

CENTRE DE GRAVITE	INER	CUMUL
Milieu 4	2	18.16
Milieu 4	1	16.29
Séq. de référenc	2	6.72

Ce tableau reprend des données du tableau du § 13 (de même que le tableau du § 12 reprend des données du tableau du § 9).

17. Aides à l'interprétation des colonnes

COORDONNEES ET AIDES A L'INTERPRETATION DES VARIABLES CONTINUES ACTIVES
AXES 1 A 5

VARIABLES		COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
IDENT./GROUPE	DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
GROUPE 1																
Int. olfactive rep.	1.00	-0.6	-0.7	0.0	0.3	0.1	4.5	14.5	0.0	12.9	0.6	0.35	0.45	0.00	0.11	0.00
Qlt aromatique rep.	1.00	-0.8	0.1	0.4	0.1	0.0	9.0	0.2	9.1	2.4	0.4	0.70	0.01	0.13	0.02	0.00
Note fruitée repos	1.00	-0.7	0.2	0.5	0.3	0.2	6.6	0.7	20.9	7.6	5.9	0.51	0.02	0.29	0.06	0.04
Note florale repos	1.00	-0.4	0.4	-0.6	0.4	0.1	2.5	5.5	29.4	19.5	2.5	0.19	0.17	0.41	0.16	0.01
Note épicée repos	1.00	0.0	-0.9	-0.1	-0.2	0.4	0.0	24.4	1.2	4.0	22.3	0.00	0.75	0.02	0.03	0.13
ENSEMBLE							22.6	45.3	60.7	46.4	31.6	0.35	0.28	0.17	0.08	0.04
GROUPE 2																
Intensité visuelle	1.00	-0.9	-0.2	-0.1	-0.2	-0.2	7.9	1.5	1.1	4.3	4.1	0.78	0.06	0.02	0.05	0.03
Nuance orangé/violet	1.00	-0.9	-0.2	-0.1	-0.3	-0.2	7.6	1.4	1.2	6.5	4.4	0.74	0.05	0.02	0.07	0.03
Impression surface	1.00	-1.0	0.0	0.0	-0.1	-0.1	9.2	0.1	0.0	1.4	2.5	0.90	0.00	0.00	0.01	0.02
ENSEMBLE							24.7	2.9	2.3	12.2	11.0	0.81	0.04	0.01	0.04	0.03
GROUPE 3																
Intensité olfactive	1.00	-0.6	-0.6	-0.2	0.3	-0.1	2.4	5.2	1.6	5.4	0.2	0.39	0.33	0.05	0.09	0.00
Qualité olfactive	1.00	-0.8	0.4	0.2	0.0	-0.2	3.8	2.6	1.7	0.0	1.9	0.63	0.17	0.05	0.00	0.02
Note fruitée	1.00	-0.7	0.3	0.5	0.1	0.0	3.3	1.7	8.4	0.2	0.0	0.54	0.11	0.24	0.00	0.00
Note florale	1.00	-0.2	0.6	-0.6	0.1	0.3	0.2	5.1	14.0	1.1	7.7	0.03	0.33	0.41	0.02	0.10
Note épicée	1.00	-0.3	-0.7	0.0	-0.2	0.1	0.5	7.4	0.0	1.9	0.9	0.08	0.48	0.00	0.03	0.01
Note végétale	1.00	0.5	-0.6	0.0	0.4	-0.3	1.6	6.2	0.0	8.7	6.2	0.26	0.40	0.00	0.15	0.08
Note phénolique	1.00	-0.4	-0.3	-0.2	0.3	-0.5	0.9	1.1	2.0	6.1	20.8	0.15	0.07	0.06	0.11	0.26
Int. arom. en bouche	1.00	-0.9	0.0	0.1	0.2	-0.1	5.2	0.0	0.2	2.5	1.5	0.84	0.00	0.01	0.04	0.02
Pers. arom. bouche	1.00	-0.9	-0.1	-0.2	0.1	0.1	5.3	0.1	1.3	1.0	1.2	0.86	0.01	0.04	0.02	0.02
Qlt arom. en bouche	1.00	-0.7	0.6	0.0	-0.1	0.2	3.4	4.9	0.1	0.4	2.5	0.56	0.32	0.00	0.01	0.03
ENSEMBLE							26.7	34.3	29.3	27.2	42.9	0.43	0.22	0.08	0.05	0.05
GROUPE 4																
Intensité d'attaque	1.00	-0.8	0.0	-0.2	-0.3	0.2	3.6	0.0	1.1	3.3	2.7	0.71	0.00	0.04	0.07	0.04
Acidité	1.00	0.2	-0.4	0.3	0.2	0.4	0.1	1.8	3.1	2.8	8.7	0.03	0.14	0.11	0.06	0.13
Astringence	1.00	-0.8	-0.5	-0.2	-0.2	0.0	3.2	2.8	0.9	1.1	0.0	0.63	0.21	0.03	0.02	0.00
Alcool	1.00	-0.8	-0.2	0.1	0.0	-0.1	3.1	0.4	0.3	0.0	0.9	0.61	0.03	0.01	0.00	0.01
Equilibre	1.00	-0.8	0.5	-0.1	-0.2	0.0	3.1	3.6	0.2	1.5	0.1	0.60	0.28	0.01	0.03	0.00
Velouté	1.00	-0.8	0.4	-0.1	-0.2	0.0	3.5	2.5	0.5	1.7	0.0	0.69	0.19	0.02	0.04	0.00
Amertume	1.00	-0.4	-0.7	-0.1	-0.1	-0.1	0.7	5.5	0.2	0.2	0.2	0.14	0.42	0.01	0.00	0.00
Int. fin de bouche	1.00	-0.9	0.0	-0.1	-0.2	0.2	4.4	0.0	0.4	1.3	1.9	0.85	0.00	0.01	0.03	0.03
Harmonie	1.00	-0.9	0.3	-0.2	-0.2	0.0	4.2	0.9	1.0	2.3	0.0	0.81	0.07	0.03	0.05	0.00
ENSEMBLE							26.0	17.4	7.7	14.2	14.5	0.56	0.15	0.03	0.03	0.02

COORDONNEES ET AIDES A L'INTERPRETATION DES VARIABLES CONTINUES ILLUSTRATIVES
AXES 1 A 5

VARIABLES		COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
IDENT./GROUPE	DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
GROUPE 5																
	ENSEMBLE						8.5	47.0	31.9	37.7	108.1	0.09	0.19	0.06	0.04	0.08
GROUPE 6																
Qualité d'ensemble	1.00	-0.7	0.5	-0.1	-0.2	0.2	8.7	10.0	1.5	5.9	5.3	0.56	0.25	0.02	0.04	0.03
Typicité	1.00	-0.8	0.5	0.0	-0.2	0.0	9.2	8.6	0.1	5.1	0.0	0.59	0.22	0.00	0.03	0.00
	ENSEMBLE						17.9	18.6	1.6	11.0	5.3	0.57	0.24	0.01	0.04	0.01

Contrairement à l'ACM, même si elles sont actives (ce n'est pas le cas ici), les indicatrices n'apparaissent pas individuellement dans ce tableau (les modalités apparaissent avant, en tant que cdg) ; seuls figurent les ensembles d'indicatrices d'un même groupe (cf. groupe 5).

DISTO : pour les variables continues centrées et réduites, la distance à l'origine est constante et égale à 1. Lorsque les variables ne sont pas réduites, on retrouve la variance.

Comme en ACP, la coordonnée d'une variable continue, dans le cas centré réduit, est un coefficient de corrélation. Pour les colonnes non réduites, les coefficients de corrélation entre les variables continues et les facteurs apparaissent dans un tableau séparé.

Contribution = (inertie projetée du point)/(inertie projetée de l'ensemble des points).

- La somme des contributions des colonnes (variables continues ou indicatrices) d'un groupe est égale à la contribution du groupe (déjà vue dans le tableau du § 6) ; la somme des contributions vaut 100 lorsque l'on prend en compte l'ensemble des variables actives.
- Pour les colonnes illustratifs, ce rapport est calculé bien qu'il ne s'interprète pas comme une contribution au plein sens du terme.

Cosinus carrés = (inertie projetée du point)/(inertie totale du point). Dans le cas centré réduit, ce critère est le carré de la coordonnée. Ce rapport est calculé aussi pour le sous-nuage de l'ensemble des colonnes d'un groupe. Ainsi, par exemple, le plan (1,2) représente 63% (35% + 28%) de l'inertie du nuage des variables du groupe 1, valeur à comparer à l'inertie du premier plan de l'ACP séparée du groupe 1 : 75% (cf. tableau du § 2). Ceci permet de détecter si, en voulant représenter de façon optimale les variables de plusieurs groupes, on ne perd pas trop par rapport aux représentations optimales de chacun d'eux.

18. Aides à l'interprétation des axes partiels

COORDONNEES ET AIDES A L'INTERPRETATION DES AXES PARTIELS ACTIFS
AXES 1 A 5

VARIABLES				COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
IDENT./GROUPE		POIDS DISTO		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
GROUPE 1																		
AXE	1	1.00	1.00	-0.9	-0.1	0.3	0.3	0.1	22.2	0.8	10.7	27.0	7.7	0.77	0.01	0.07	0.10	0.02
AXE	2	0.68	1.00	-0.1	0.9	-0.1	0.2	-0.2	0.2	44.1	0.3	6.1	7.6	0.01	0.89	0.00	0.03	0.03
AXE	3	0.36	1.00	-0.1	-0.1	-0.9	0.3	0.1	0.1	0.3	49.3	8.4	2.7	0.01	0.01	0.83	0.09	0.02
AXE	4	0.13	1.00	0.0	0.1	0.1	-0.1	0.5	0.0	0.1	0.2	0.2	11.6	0.00	0.01	0.01	0.00	0.25
AXE	5	0.06	1.00	0.2	-0.1	0.2	0.5	-0.3	0.1	0.1	0.3	4.8	2.0	0.03	0.01	0.03	0.28	0.09
		ENSEMBLE							22.6	45.3	60.7	46.4	31.6	0.35	0.28	0.17	0.08	0.04
GROUPE 2																		
AXE	1	1.00	1.00	0.9	0.2	0.1	0.2	0.2	24.6	2.4	1.3	11.4	10.9	0.85	0.03	0.01	0.04	0.03
AXE	2	0.05	1.00	0.2	-0.4	-0.3	-0.2	-0.1	0.1	0.6	1.0	0.8	0.1	0.06	0.14	0.12	0.05	0.01
AXE	3	0.01	1.00	-0.1	-0.1	0.0	0.3	0.0	0.0	0.0	0.0	0.1	0.0	0.01	0.00	0.00	0.08	0.00
		ENSEMBLE							24.7	2.9	2.3	12.2	11.0	0.81	0.04	0.01	0.04	0.03
GROUPE 3																		
AXE	1	1.00	1.00	0.9	-0.3	0.0	-0.1	0.0	26.0	4.9	0.2	1.1	0.1	0.90	0.07	0.00	0.00	0.00
AXE	2	0.53	1.00	0.2	0.9	0.1	-0.2	0.3	0.7	28.6	0.5	7.9	13.5	0.04	0.74	0.01	0.06	0.07
AXE	3	0.22	1.00	0.1	0.1	-0.8	0.4	0.0	0.0	0.2	24.9	8.7	0.1	0.00	0.01	0.69	0.15	0.00
AXE	4	0.16	1.00	0.1	0.1	0.4	0.4	-0.6	0.0	0.2	3.3	7.6	22.1	0.01	0.02	0.13	0.17	0.37
AXE	5	0.11	1.00	-0.1	-0.2	0.1	0.2	0.4	0.0	0.3	0.2	1.0	4.9	0.01	0.04	0.01	0.04	0.12
		ENSEMBLE							26.7	34.3	29.2	26.4	40.6	0.43	0.22	0.08	0.05	0.05
GROUPE 4																		
AXE	1	1.00	1.00	0.9	-0.1	0.1	0.2	0.0	25.9	0.4	3.5	11.1	0.4	0.90	0.01	0.02	0.04	0.00
AXE	2	0.32	1.00	0.1	0.8	-0.1	-0.1	-0.1	0.0	15.7	0.8	1.1	1.3	0.01	0.68	0.02	0.01	0.01
AXE	3	0.12	1.00	0.0	0.3	0.3	0.2	0.5	0.0	0.8	1.7	0.7	10.2	0.00	0.09	0.09	0.02	0.23
AXE	4	0.06	1.00	0.1	0.1	-0.1	-0.1	0.1	0.0	0.0	0.1	0.2	0.3	0.01	0.01	0.01	0.01	0.01
AXE	5	0.06	1.00	0.0	0.3	0.4	0.2	-0.3	0.0	0.5	1.4	0.8	1.6	0.00	0.12	0.16	0.06	0.08
		ENSEMBLE							26.0	17.4	7.7	14.0	13.8	0.56	0.15	0.03	0.03	0.02
									CONTRIBUTION CUMULEE=100.0 99.9 99.9 99.0 97.0									

COORDONNEES ET AIDES A L'INTERPRETATION DES AXES PARTIELS ILLUSTRATIFS
AXES 1 A 5

VARIABLES			COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
IDENT./GROUPE		POIDS DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
GROUPE 5																	
AXE 1		1.00 1.00	0.1	-0.5	0.1	0.3	-0.5	0.1	16.9	1.0	25.0	79.0	0.00	0.23	0.01	0.09	0.21
AXE 2		0.88 1.00	0.3	0.6	0.2	-0.1	0.1	1.9	21.3	7.7	1.1	2.8	0.07	0.33	0.05	0.00	0.01
AXE 3		0.69 1.00	-0.5	0.0	0.3	0.0	0.1	5.5	0.1	11.6	0.3	5.1	0.27	0.00	0.10	0.00	0.02
AXE 4		0.50 1.00	-0.1	0.4	-0.2	0.0	-0.3	0.3	6.2	2.3	0.0	18.0	0.02	0.17	0.03	0.00	0.10
AXE 5		0.38 1.00	0.3	-0.3	0.4	-0.3	0.2	0.9	2.6	9.3	11.2	3.2	0.08	0.09	0.15	0.11	0.02
ENSEMBLE								8.5	47.0	31.9	37.7	108.1	0.09	0.19	0.06	0.04	0.08
GROUPE 6																	
AXE 1		1.00 1.00	-0.8	0.5	-0.1	-0.2	0.1	17.9	18.6	1.2	11.0	2.8	0.62	0.25	0.01	0.04	0.01
AXE 2		0.08 1.00	0.0	0.1	-0.2	0.0	0.3	0.0	0.0	0.4	0.0	2.6	0.00	0.00	0.03	0.00	0.09
ENSEMBLE								17.9	18.6	1.6	11.0	5.3	0.57	0.24	0.01	0.04	0.01
CONTRIBUTION CUMULEE=								26.4	65.6	33.5	48.7	113.4					

Les axes partiels sont issus des analyses séparées des groupes. Ils sont considérés comme des variables continues illustratives (centrées-réduites –donc DISTO=1– et pondérées) et donnent lieu aux mêmes aides à l'interprétation.

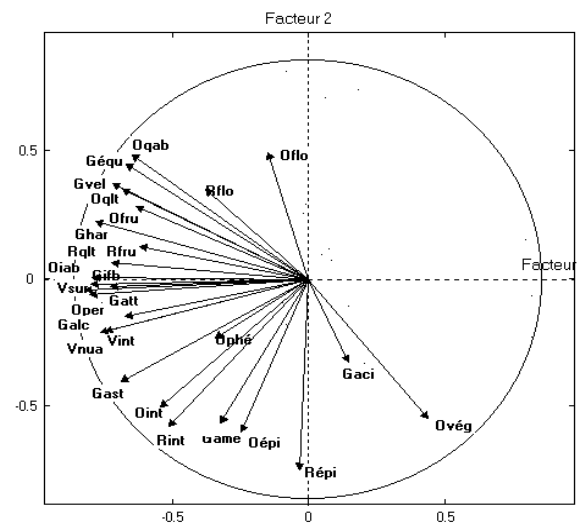
Le POIDS d'un axe est égal à la valeur propre qui lui est associée divisée par la première valeur propre de l'analyse du groupe dont l'axe est issu. Ainsi le poids du premier axe de chaque groupe vaut 1. Ici, par exemple, on retrouve que le groupe 2 est pratiquement unidimensionnel (seul son premier axe a un poids sensiblement différent de 0).

Il est équivalent de réaliser l'analyse globale à partir des colonnes actives ou à partir de l'ensemble des facteurs partiels actifs ainsi pondérés ; ainsi les rapports d'inertie figurant dans les colonnes contributions s'interprètent bien comme des contributions. Ici, par exemple :

- Les contributions correspondant au premier axe de chaque groupe actif contribuent également (entre 22.2% et 26%) à la contribution du premier axe de l'AFM;
- en se limitant aux 5 premiers axes de chacun des groupes actifs, on reconstitue presque parfaitement les 5 axes de l'AFM.

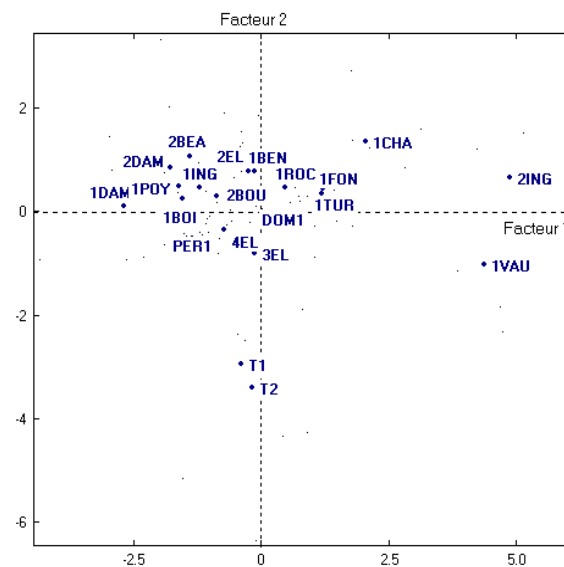
19. Représentation graphique des variables continues

Comme en ACP usuelle, les coordonnées des variables sont leurs coefficients de corrélation avec les facteurs. Ici, par exemple, les variables d'intensité et d'harmonie sont très liées au premier facteur que l'on synthétise donc sous le terme « puissance et harmonie » (la correspondance entre libellés courts et longs apparaît au §2).



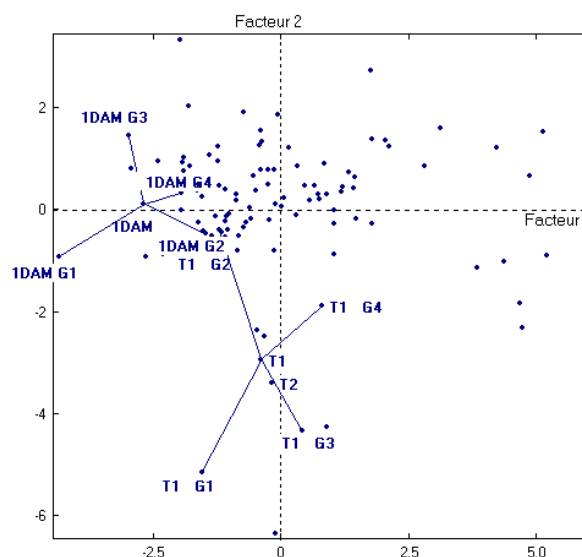
20. Représentation graphique des individus moyens

Cette représentation est strictement analogue à celle d'une analyse factorielle usuelle. Ici, par exemple, les vins 1VAU et 2ING ont été jugés particulièrement peu « puissants et harmonieux ». Le deuxième facteur met essentiellement en évidence le caractère particulier des vins T1 et T2.

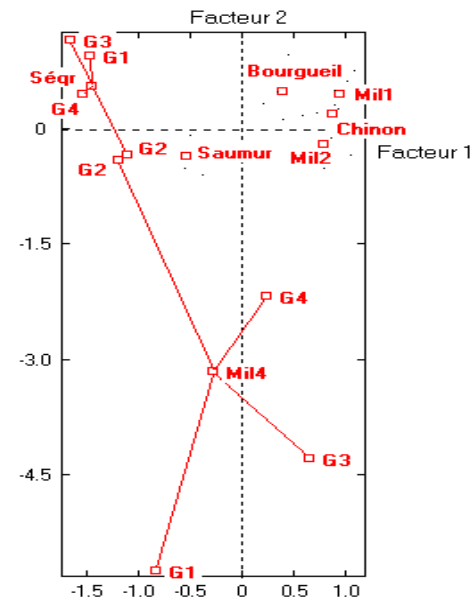


21. Représentation superposée des nuages (d'individus) partiels

Le libellé d'un individu partiel reprend celui de l'individu, complété par G suivi du numéro de groupe. L'ensemble des points partiels est très volumineux. On sélectionne donc plutôt quelques points partiels que l'on relie, par un segment, au point moyen homologue (CTRL+SHIFT+double-clic sur un point fait apparaître ces segments pour un individus). Ici par exemple, le vin T1 est surtout remarquable par ses caractéristiques olfactives (G1 et G3), à un moindre degré par ses caractéristiques gustatives (G4) et pas du tout par son aspect visuel (G2).

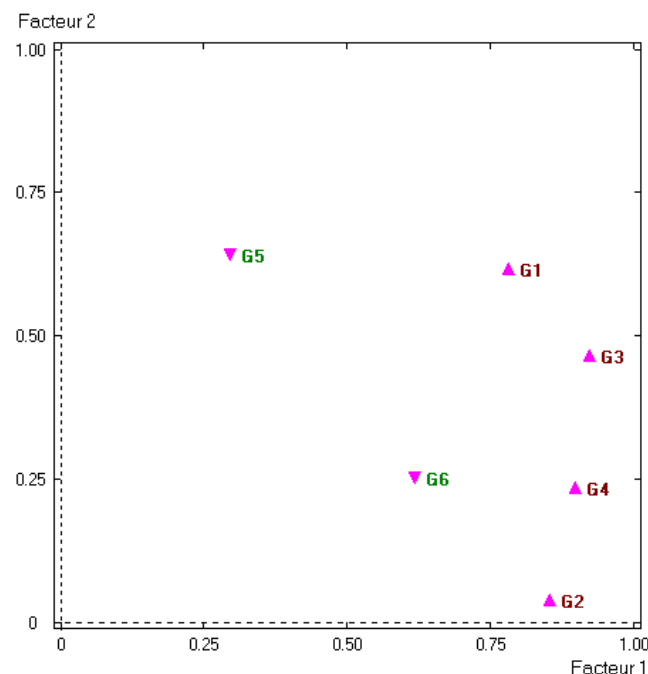


22. Représentation des variables nominales



Comme en ACP, les variables nominales sont représentées au travers des centres de gravité des individus possédant une même modalité. En outre, en AFM, on peut associer à chaque centre de gravité son image partielle, dont le libellé court se limite à G suivi du numéro de groupe. Ici, par exemple, les vins de la séquence de référence (Séqr) sont globalement puissants et harmonieux, quelle que soit la phase (de la dégustation) considérée.

23. Représentation des groupes



La coordonnée d'un groupe s'interprète comme un indice de liaison entre le groupe et le facteur de l'AFM. Elle est toujours comprise entre 0 et 1, d'où la présence du carré de côté 1 construit à partir de l'origine. Ici, par exemple, les 4 groupes actifs (G1-G4) sont également, et fortement, liés au premier axe. Corrélativement on peut interpréter la proximité entre groupes. Ici, par exemple, les deux olfactions (G1 et G3) apparaissent proches car à la fois elles « possèdent » le facteur commun « puissance et harmonie » (axe 1) et mettent bien en évidence la spécificité de T1 et T2 (axe 2). Les deux groupes « voisins » et « gustation » apparaissent également proches mais ceci en raison de leur faible liaison avec l'axe 2.

Ce plan représente assez bien les groupes (L'inertie projetée/inertie totale = 72% -cf §6-) et donc illustre bien les coefficients de liaison (cf §5). L'identité des axes de ce graphique avec ceux des autres représentations est essentielle pour la cohérence de l'interprétation de l'ensemble des résultats de l'AFM.

1. Présentation

1.1 Objet

Cette procédure permet d'extraire l'information de données multidimensionnelles recueillies dans des situations différentes. Prenons l'exemple d'un essai thérapeutique : les patients sont suivis pendant plusieurs mois et lors de chaque visite, le praticien décrit l'état pathologique d'un patient en situant ses performances sur un certain nombre d'échelles. On obtient des données à trois indices : un indice "patient", un indice "performance" et un indice "visite".

L'exploitation de telles données, à trois "voies", est un axe de recherche qui est loin d'être épuisé. En effet, les méthodes proposées à l'heure actuelle (y compris les méthodes exposées ici ...) n'abordent pas le problème dans sa complexité. En particulier, l'ordre dans lequel sont effectuées les visites, est rarement pris en compte. La méthode STATIS, tout comme l'analyse factorielle multiple (procédure AFMUL), privilégie une des "voies" car elle découpe les données en tranches. Ces tranches sont ensuite comparées entre elles à l'aide des outils de géométrie euclidienne habituels en analyse de données.

Méthode STATIS (option 2 de la procédure STTIS).

La méthode STATIS a été développée dans le laboratoire de probabilités et statistique de l'université Montpellier II par Yves Escoufier et son équipe. Les données sont découpées en q tableaux X_k (visites) de n lignes (patients) et p colonnes (performances). A chaque visite, on peut traduire géométriquement la variabilité inter-individuelle par un nuage multidimensionnel dans lequel chacun des n points représente un patient. Si les patients évoluaient tous de la même manière, ce nuage de points se déplacerait au cours des visites sans que les positions mutuelles des patients, les uns par rapport aux autres, changent.

La méthode STATIS donne une mesure de la déformation de ce nuage de patients au cours du temps. On compare les q nuages en calculant des distances euclidiennes, déduites du coefficient RV d'Escoufier (produit scalaire entre les matrices $W_k = X_k^t X_k$ normées). Ces distances définissent une configuration de q points dans laquelle le point n° k représente le nuage des patients lors de la k ème visite. On sait obtenir des représentations graphiques de cette configuration par la technique classique des axes principaux. Cette configuration est appelée "interstructure".

Pour déceler les réactions individuelles, on calcule la contribution de chaque patient à la distance euclidienne entre deux nuages. Enfin la méthode fournit un résumé de la position mutuelle des patients sous la forme d'un nuage compromis (c'est le nuage le plus proche, au sens de la distance utilisée précédemment, des différents nuages de patients à chaque visite). Dans la mesure où les nuages sont assez semblables d'une visite à l'autre, le nuage compromis décrit la structure commune à ces différents nuages, ou "intrastructure".

Cette méthode offre une certaine souplesse d'utilisation. En effet, puisqu'on compare les tranches X_k par l'intermédiaire des nuages de lignes, peu importe comment ces nuages ont été obtenus. En d'autres termes, les tableaux X_k peuvent ne pas avoir le même nombre de colonnes et ces colonnes peuvent correspondre à des groupes de variables différents.

Les seules contraintes sur les tableaux X_k sont les suivantes : les tableaux décrivent les mêmes objets ; pour un tableau, la pondération des objets est uniforme et la distance entre deux objets correspond à la distance euclidienne classique entre les lignes du tableau. Par exemple, il est possible de comparer les résultats d'enquêtes portant sur les mêmes individus, à des dates différentes ou sur des sujets d'enquêtes différents, en prenant comme tableau X_k , le tableau des coordonnées factorielles issues de l'analyse des correspondances multiples des réponses à l'enquête n°k.

Méthode STATIS DUALE (option 3 de la procédure STTIS).

La "stratégie" STATIS peut être utilisée pour comparer les nuages particuliers que sont les nuages de variables lorsque les variables des tableaux X_k sont centrées et réduites. En effet, le nuage des variables centrées et réduites est un nuage multidimensionnel "calibré" puisque tous les points M sont à la distance 1 de l'origine O. De plus, le cosinus de l'angle entre deux vecteurs OM n'est autre que le coefficient de corrélation linéaire entre les deux variables.

La distance euclidienne entre deux nuages de variables est déduite du produit scalaire entre les matrices de corrélations $V_k = {}^tX_k X_k / n$ (qu'il est inutile de normer puisque les nuages sont "calibrés"). Comme précédemment, ces distances définissent une configuration d'interstructure de q points, chaque point représentant un nuage de variables. On calcule ensuite la contribution de chaque variable à la distance euclidienne entre deux nuages, puis un résumé sous la forme d'un nuage compromis (ou d'une matrice de corrélations compromise, ce qui est équivalent).

Il faut que les tableaux soient du type : individus x variables continues et qu'ils concernent les mêmes variables, peu importe si les groupes d'individus sont différents. Les variables n'ont pas besoin d'être centrées et réduites car le programme calcule les matrices de corrélations avant de dérouler la méthode STATIS DUALE.

Analyse triadique de Jaffrenou (option 1 de la procédure STTIS).

La distance euclidienne entre deux nuages est en fait une distance euclidienne entre matrices qu'on peut employer sur les tableaux X_k eux-mêmes, normés ou non. Comme précédemment, ces distances définissent encore une configuration d'interstructure de q points dans laquelle chaque point représente un tableau. On calcule ensuite la contribution de chaque ligne et de chaque colonne à la distance euclidienne entre deux tableaux. En revanche, on ne sait pas définir un tableau compromis qui ait les bonnes propriétés d'un nuage compromis.

Les lignes des tableaux doivent correspondre aux mêmes objets, les colonnes également. Les données peuvent être de n'importe quelle nature.

La procédure utilise en entrée les fichiers NDIC et NDON (créés dans la procédure SELEC). Dans le fichier NDON, les q tableaux sont juxtaposés et il ne doit pas y avoir de données manquantes. En sortie, la procédure fournit les fichiers NGUS1, NDIC1, NDON1, et dans le cas des options 2 et 3, les fichiers supplémentaires NGUS, NDICB, NDONB.

1.2 Editions

On obtient la situation des tableaux (ou des nuages suivant l'option) les uns par rapport aux autres ; l'interprétation des distances entre tableaux (ou nuages) en termes de contributions des lignes et (ou) des colonnes ; l'étude de la configuration de l'interstructure ; et l'étude du nuage compromis lorsqu'il existe.

1.3 Paramètres

Les paramètres de la procédure se répartissent en trois catégories :

- **Les paramètres de définition de l'analyse** : choix de l'option (LMETH) ; normalisation des tableaux si on choisit la première option (LNORM).
- **Les paramètres de fonctionnement** : majorant du nombre de tableaux (NXTAB) ; nombre d'axes retenus pour le graphique de l'interstructure (NAXTA) ; couples de points de l'interstructure pour lesquels on veut décomposer la distance en contribution des lignes et (ou) des colonnes (LNUMT) ; suivant l'option choisie, nombre d'axes retenus pour le graphique du nuage des lignes compromis (NAXLI), nombre d'axes retenus pour le graphique du nuage des variables compromis (NAXCO).

Les paramètres d'édition contrôlent le listage des coordonnées et des contributions des points de l'interstructure (NEDIT), et suivant l'option choisie, des lignes compromis (NEDLI), ou des variables compromis (NEDCO).

1.4 Après les paramètres

Après les paramètres, on indique le découpage en tableaux du fichier d'entrée en donnant pour chacun des tableaux la liste des colonnes qui lui sont affectées. On termine par un enregistrement contenant le mot-clé FIN.

2. Instructions de commande

Les valeurs par défaut sont indiquées entre parenthèses, à la suite du nom du paramètre.

- | | |
|----------------|---|
| (1) PROC STTIS | Analyse conjointe de tableaux |
| (2) | <i>titre de l'analyse</i> |
| (3) NXTAB (15) | majorant du nombre de tableaux. |
| LMETH (2) | choix de l'option |
| | 1 : comparaison des tableaux |
| | 2 : des nuages de lignes |
| | 3 : des matrices de corrélations |
| LNORM (1) | normalisation des tableaux dans l'option 1 |
| | 0 : pas de normalisation |
| | 1 : normalisation |
| NAXTA (TOT) | nombre d'axes retenus pour le graphique de l'interstructure |
| | TOT : tous les axes |
| LNUMT (0) | couples de points de l'interstructure pour lesquels on veut décomposer la distance |
| | 0 : aucun |
| | TOT : tous les couples de points |
| | k : le point n°k et tous les autres |
| NAXLI (TOT) | nombre d'axes retenus pour le graphique du nuage des lignes compromis |
| | TOT : tous les axes |
| NAXCO (TOT) | nombre d'axes retenus pour la représentation graphique de la matrice des corrélations compromis |
| | TOT : tous les axes |
| NEDIT (5) | nombre d'axes pour lesquels seront éditées les coordonnées des points de l'interstructure |
| | TOT : tous les axes |
| NEDLI (5) | nombre d'axes pour lesquels seront éditées les coordonnées des lignes compromises |
| | TOT : tous les axes |
| NEDCO (5) | nombre d'axes pour lesquels seront éditées les coordonnées des variables compromis |
| | TOT : tous les axes |
| (4) | Découpage du fichier d'entrée : liste des colonnes qui constituent un tableau. |
| (5) | FIN |

3. Présentation détaillée des paramètres

NXTAB

majorant du nombre de tableaux

- *valeurs possibles* : toutes les valeurs plus grandes que 2
- *valeur par défaut* : 15

LMETH

choix de l'option

- *valeurs possibles* : 1 (comparaison des tableaux)
2 (comparaison des nuages de lignes)
3 (comparaison des matrices de corrélations)
- *valeur par défaut* : 2

Si LMETH = 1, on compare q tableaux de n lignes et p colonnes. Les lignes des tableaux doivent correspondre aux mêmes objets, les colonnes également. Il peut y avoir des colonnes ou des lignes entièrement nulles, mais pas de données manquantes. Dans le fichier NDON, les q tableaux sont juxtaposés : chacun des n enregistrements correspond à une ligne et contient les p valeurs du premier tableau, puis les p valeurs du deuxième tableau et ainsi de suite jusqu'au $q^{\text{ième}}$ tableau.

Si LMETH = 2, on compare les nuages de lignes. Les lignes des q tableaux doivent correspondre aux mêmes individus. Les tableaux peuvent ne pas avoir le même nombre de colonnes et ces colonnes peuvent appartenir à des groupes de variables différents. Dans le fichier NDON, les q tableaux sont juxtaposés : chacun des n enregistrements correspond à une ligne et contient les p_1 valeurs du premier tableau, puis les p_2 valeurs du deuxième tableau et ainsi de suite jusqu'au $q^{\text{ième}}$ tableau. Il ne doit pas y avoir de données manquantes.

Si LMETH = 3, on compare les matrices de corrélations. Les données des q tableaux portent sur les mêmes variables. Les tableaux peuvent ne pas avoir le même nombre de lignes et ces lignes peuvent correspondre à des groupes d'individus différents. il ne doit pas y avoir de données manquantes. Dans le fichier NDON, les q tableaux juxtaposés doivent avoir, impérativement, le même nombre de lignes. Cet obstacle peut être contourné de la manière suivante : Si les tableaux de données sont hétérogènes et ne portent pas sur les mêmes individus, on crée un fichier NDON dans lequel les lignes manquantes d'un tableau sont remplacées par des lignes de zéros.

LNORM**normalisation des tableaux si LMETH = 1**

- *valeurs possibles* : 0 (pas de normalisation)
1 (normalisation)
- *valeur par défaut* : 1

Si LNORM = 0, on compare les tableaux bruts.

Si LNORM = 1, on compare les tableaux normés. Si deux tableaux sont différents, cela peut provenir de deux causes : soit c'est l'ordre de grandeur des données qui n'est pas le même d'un tableau à l'autre, soit c'est la répartition des données à l'intérieur du tableau qui n'est pas la même d'un tableau à l'autre. Il faut choisir de normer les tableaux lorsqu'on ne s'intéresse qu'à la structure interne du tableau et non à l'ordre de grandeur des données.

NAXTA**nombre d'axes retenus pour la représentation graphique de l'interstructure**

- *valeurs possibles* : de 1 au nombre de tableaux comparés
- *valeur par défaut* : TOT (tous les axes)

Ce paramètre détermine le nombre d'axes qui seront sauvegardés, et qui pourront donc être édités. Si la valeur codée s'avérait trop grande, elle serait ramenée à la valeur maximale autorisée. Si on veut reconstruire complètement le nuage, il faut garder tous les axes (TOT) pour pouvoir faire une classification des tableaux à partir des coordonnées factorielles.

LNUMT**couples de points de l'interstructure pour lesquels on veut décomposer la distance**

- *valeurs possibles* : 0 (aucun)
TOT ou -1 (tous les couples de points)
k (le point n°k et tous les autres)
- *valeur par défaut* : 0

Si LNUMT = 0, on n'obtient qu'un résultat global, à savoir la décomposition de la somme des distances au carré entre tous les couples de points, en contributions des lignes et des colonnes (si LMETH = 1), des lignes (si LMETH = 2), ou des colonnes (si LMETH = 3).

Si $LNUMT = TOT$, en plus du résultat global précédent, on obtient, pour chaque couple de points, la décomposition de la distance au carré entre ces deux points, en contributions des lignes et des colonnes (si $LMETH = 1$), des lignes (si $LMETH = 2$), ou des colonnes (si $LMETH = 3$).

Si $LNUMT = k$, en plus du résultat global, on obtient la décomposition de la distance au carré entre le point $n^{\circ}k$ et les $q-1$ autres points, en contributions des lignes et des colonnes (si $LMETH = 1$), des lignes (si $LMETH = 2$), ou des colonnes (si $LMETH = 3$).

NAXLI

nombre d'axes retenus pour la représentation graphique du nuage des lignes compromis si $LMETH = 2$

- *valeurs possibles* : de 1 au nombre de lignes d'un tableau
- *valeur par défaut* : TOT (tous les axes)

Ce paramètre détermine le nombre d'axes qui seront sauvegardés, et qui pourront donc être édités. Si la valeur codée s'avérait trop grande, elle serait ramenée à la valeur maximale autorisée. Si on veut reconstruire complètement le nuage, il faut garder tous les axes (TOT) pour pouvoir faire une classification des individus compromis à partir des coordonnées factorielles.

NAXCO

nombre d'axes retenus pour la représentation graphique de la matrice des corrélations compromis si $LMETH = 3$

- *valeurs possibles* : de 1 au nombre de colonnes d'un tableau
- *valeur par défaut* : TOT (tous les axes)

Ce paramètre détermine le nombre d'axes qui seront sauvegardés, et qui pourront donc être édités. Si la valeur codée s'avérait trop grande, elle serait ramenée à la valeur maximale autorisée.

NEDIT

nombre d'axes pour lesquels on édite les coordonnées

- *valeurs possibles* : de 1 à valeur de $NAXTA$
 TOT (tous les axes)
- *valeur par défaut* : 5

Les coordonnées factorielles, les contributions à l'inertie d'un axe et les cosinus carrés, sont listés par groupe de 5. Si $NEDIT$ n'est pas un multiple de 5, les tableaux des coordonnées, des contributions et des cosinus carrés sont complétés par des zéros.

NEDLI

nombre d'axes pour lesquels on édite les coordonnées des lignes compromis si LMETH = 2

- *valeurs possibles* : de 1 à valeur de NAXLI
TOT (tous les axes)
- *valeur par défaut* : 5

Les coordonnées factorielles, les contributions à l'inertie d'un axe et les cosinus carrés, sont listés par groupe de 5. Si NEDLI n'est pas un multiple de 5, les tableaux des coordonnées, des contributions et des cosinus carrés sont complétés par des zéros.

NEDCO

nombre d'axes pour lesquels on édite les coordonnées des variables compromis si LMETH = 3

- *valeurs possibles* : de 1 à valeur de NAXCO
TOT (tous les axes)
- *valeur par défaut* : 5

Les coordonnées factorielles, les contributions à l'inertie d'un axe et les cosinus carrés, sont listés par groupe de 5. Si NEDCO n'est pas un multiple de 5, les tableaux des coordonnées, des contributions et des cosinus carrés sont complétés par des zéros.

4. Exemples de commande

Les données portent sur les profils socioprofessionnels des 30 cantons ruraux de l'Hérault, au cours des recensements de 1954, 1962, 1968, 1975 et 1982. Pour les quatre premiers recensements, les professions sont réparties en 9 classes. En 82, la nomenclature de l'INSEE a été modifiée, et la classe "clergé, armée, police" a été ventilée dans les 8 autres classes. Le fichier de données comporte donc 30 lignes et (9x4)+8 colonnes.

Pour utiliser l'option n°1, il faut que les tableaux (correspondant aux recensements), portent sur les mêmes individus et les mêmes variables. On ne peut donc comparer que les quatre premiers tableaux.

Dans l'option n°2, on compare les nuages d'individus, les variables pouvant être différentes d'un tableau à l'autre. On peut donc traiter l'ensemble des cinq recensements.

La procédure STTIS doit être précédée par la procédure SELEC qui permet de créer les fichiers NDIC et NDON.

4.1 Exemple de comparaison de tableaux (option n°1)

```

-----+-----1-----+-----2-----+-----3-----+-----4-----+-----5-----+-----6-----+
NDICA = 'base.lad', NDONA = 'base.dar' : fichiers de base
NDIC = 'candic.din', NDON = 'candon.bin' : fichiers sorties de SELEC
PROC SELEC
===== exemple 1 : sélection des (9x4) premières variables =====
LSELI = LIST LZERO = NOREC
CONT ACT 1--36
FIN
ACT 1--30
FIN

NDIC1 = 'dic1.bin', NDON1 = 'don1.bin', NGUS1 = 'gus1.bin' : sorties STTIS
PROC STTIS
===== exemple 1 : tableaux des cantons de l'Hérault pour 54, 62, 68, 75 =====
NXTAB = 4 LMETH = 1 LNORM = 0 NAXTA = 3 LNUMT = 4 NEDIT = 3
GROU V1--V9
GROU V10--V18
GROU V19--V27
GROU V28--V36
FIN

STOP
-----+-----1-----+-----2-----+-----3-----+-----4-----+-----5-----+-----6-----+

```

L'utilisateur utilise l'option n°1 (LMETH = 1) de la procédure STTIS pour effectuer une comparaison des quatre tableaux (NXTAB = 4), définis par le découpage du fichier NDON selon les quatre groupes de variables :

```

GROU V1--V9
GROU V10--V18
GROU V19--V27
GROU V28--V36.

```

L'utilisateur demande la décomposition détaillée des distances entre le tableau n°4 et les trois autres (LNUMT = 4). Il demande qu'on retienne 3 axes pour le graphique de l'interstructure (NAXTA = 3) et qu'on édite les coordonnées factorielles des tableaux, les contributions et les cosinus carrés sur ces 3 axes (NEDIT = 3).

4.2 Exemple de comparaison de nuages (option n°2)

```

-----+-----1-----+-----2-----+-----3-----+-----4-----+-----5-----+-----6-----+
PROC SELEC
===== exemple 2 : sélection de toutes les variables =====
LSELI = LIST LZERO = NOREC
CONT ACT 1--44
FIN
ACT 1--30
FIN

NDICB = 'dic.bin', NDONB = 'don.bin', NGUS = 'gus.bin'
NDIC1 = 'dic1.bin', NDON1 = 'don1.bin', NGUS1 = 'gus1.bin' : sorties STTIS
PROC STTIS
===== exemple 2 : nuages des cantons de l'Hérault pour 54, 62, 68, 75, 82 =====
NXTAB = 5 LMETH = 2 NAXTA = TOT LNUMT = TOT NAXLI = 10 >
NEDIT = 3 NEDLI = 2
GROU V1--V9
GROU V10--V18
GROU V19--V27
GROU V28--V36
GROU V37--V44
FIN

STOP
-----+-----1-----+-----2-----+-----3-----+-----4-----+-----5-----+-----6-----+

```

L'utilisateur utilise l'option n°2 (LMETH = 2) de la procédure STTIS pour effectuer une comparaison des cinq nuages de lignes (NXTAB = 5), définis par le découpage du fichier NDON selon les cinq groupes de variables :

```

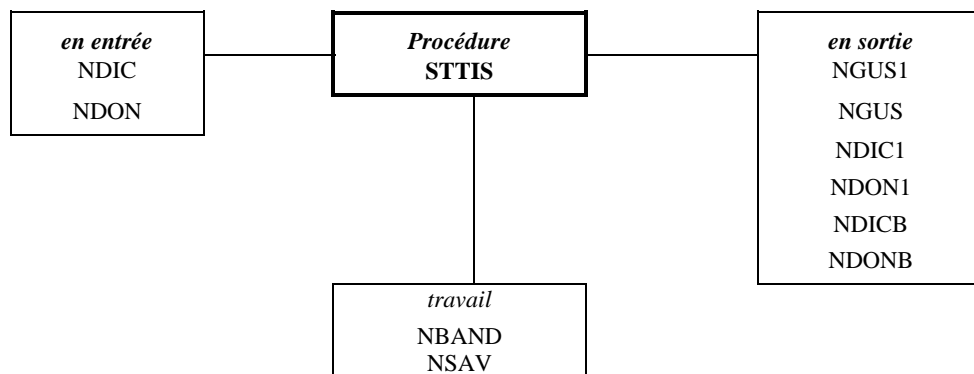
GROU V1--V9
GROU V10--V18
GROU V19--V27
GROU V28--V36
GROU V37--V44.

```

L'utilisateur demande la décomposition détaillée des distances entre tous les couples de nuages (LNUMT = TOT). Pour le graphique de l'interstructure, l'utilisateur demande qu'on retienne tous les axes (NAXTA = TOT) et qu'on édite les coordonnées factorielles des nuages, les contributions et les cosinus carrés sur 3 axes (NEDIT = 3). Pour la représentation graphique des lignes compromis, l'utilisateur demande qu'on retienne 10 axes (NAXLI = 10) et qu'on édite les coordonnées factorielles des individus compromis, les contributions et les cosinus carrés sur deux axes (NEDLI = 2).

5. Fichiers nécessaires à l'exécution

- en lecture NDIC (dictionnaire utile)
NDON (données utiles)
- en écriture NGUS1 (coordonnées factorielles des points de l'interstructure)
NDIC1 (dictionnaire utile associé au fichier NGUS1)
NDON1 (données utiles associées au fichier NGUS1)
NGUS (coordonnées factorielles des lignes compromis si
LMETH = 2, ou des variables compromis si LMETH = 3)
NDICB (dictionnaire utile associé au fichier NGUS)
NDONB (données utiles associées au fichier NGUS)
- de travail NSAV, NBAND



PROCEDURE STTIS

La méthode

Dans l'option n°1 de la procédure STTIS, on calcule des distances euclidiennes entre tableaux de données. Cela revient à assimiler un tableau à un vecteur (on empile les colonnes du tableau les unes sur les autres) et utiliser le produit scalaire habituel entre vecteurs. Dans les deux autres options, on compare des nuages de points par l'intermédiaire de matrices. La distance entre nuages se déduit donc d'un produit scalaire entre matrices, lui-même calculé comme un produit scalaire entre vecteurs.

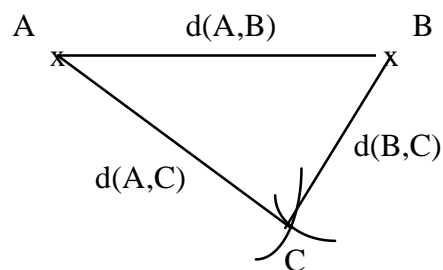
Ces distances sont ensuite décomposées pour pouvoir être interprétées, et utilisées pour obtenir un nuage compromis lorsque cela est possible. Un certain nombre de résultats sont fournis sous la forme de nuages de points qu'on exploite graphiquement par des cartes factorielles qui (contrairement à ce qui se fait en analyse en composantes principales ou en analyse des correspondances) ne passent pas forcément par le centre de gravité du nuage.

La méthode est basée sur des concepts relativement simples de géométrie euclidienne que nous rappelons brièvement dans les trois premiers paragraphes.

1. Géométrie d'un nuage de points

1.1 Configuration de points

Connaissant les distances entre trois points A, B et C, on peut dessiner le nuage des points A, B et C, à l'aide d'un compas et d'une règle graduée. Pour trois points, un plan suffit. Pour n points, un espace de dimension $n-1$ suffit.



1.2 Relations entre produits scalaires, distances et nuage de points

Choisissons un quatrième point O quelconque. Connaissant les longueurs des trois vecteurs OA, OB, OC et les produits scalaires (OA|OB), (OA|OC) et (OB|OC), on peut calculer les distances mutuelles entre les trois points A, B et C. Par exemple, la distance d(A,B) est égale à :

$$d(A, B) = \sqrt{\|OA\|^2 + \|OB\|^2 - 2(OA|OB)},$$

où $\|OA\| = \sqrt{(OA|OA)}$ désigne la longueur (ou norme) du vecteur OA.

On peut donc calculer les distances à partir des produits scalaires.

$$\mathbf{W}_O = \begin{bmatrix} \|OA\|^2 & (OA|OB) & (OA|OC) \\ (OA|OB) & \|OB\|^2 & (OB|OC) \\ (OA|OC) & (OB|OC) & \|OC\|^2 \end{bmatrix} \Rightarrow \mathbf{D} = \begin{bmatrix} 0 & d(A,B) & d(A,C) \\ d(A,B) & 0 & d(B,C) \\ d(A,C) & d(B,C) & 0 \end{bmatrix}$$

Réciproquement, connaissant les distances d(A,B), d(A,C) et d(B,C), on sait calculer les produits scalaires, mais pas à partir de n'importe quel point. Soit G le centre de gravité du triangle ABC. On peut déduire des distances d(A,B), d(A,C) et d(B,C), les normes des vecteurs GA, GB, GC et les produits scalaires (GA|GB), (GA|GC) et (GB|GC) en utilisant la relation $GA+GB+GC = 0$. Par exemple, le produit scalaire (GA|GB) est égal à :

$$\begin{aligned} (GA|GB) &= \frac{1}{2} (\|GA\|^2 + \|GB\|^2 - d^2(A, B)) \\ \text{avec } \|GA\|^2 &= \frac{1}{3} (d^2(A, B) + d^2(A, C)) - \frac{1}{3^2} (d^2(A, B) + d^2(A, C) + d^2(B, C)) \\ \text{et } \|GB\|^2 &= \frac{1}{3} (d^2(A, B) + d^2(B, C)) - \frac{1}{3^2} (d^2(A, B) + d^2(A, C) + d^2(B, C)) \end{aligned}$$

On peut donc déduire les distances des produits scalaires "centrés" (c'est-à-dire calculés depuis le centre de gravité), ou le contraire.

$$\mathbf{D} = \begin{bmatrix} 0 & d(A,B) & d(A,C) \\ d(A,B) & 0 & d(B,C) \\ d(A,C) & d(B,C) & 0 \end{bmatrix} \Leftrightarrow \mathbf{W}_G = \begin{bmatrix} \|GA\|^2 & (GA|GB) & (GA|GC) \\ (GA|GB) & \|GB\|^2 & (GB|GC) \\ (GA|GC) & (GB|GC) & \|GC\|^2 \end{bmatrix}$$

Enfin, connaissant la matrice des produits scalaires \mathbf{W}_O , on sait en déduire la matrice des produits scalaires centrés \mathbf{W}_G . Il suffit d'effectuer le bicentrage (centrage des lignes suivi d'un centrage en colonnes, ou l'inverse) de \mathbf{W}_O pour obtenir \mathbf{W}_G .

En conclusion, une configuration de points est caractérisée par la matrice W_O des produits scalaires calculés à partir d'un point origine O . Le choix de l'origine dépend des données traitées. Le centre de gravité G du nuage a de bonnes propriétés mathématiques et une interprétation commode en tant que point "moyen" (c'est la raison pour laquelle on le prend souvent comme origine), mais ce n'est pas toujours le point le plus adéquat (cf § 3.2., § 5.6 et § 6.5).

1.3 Représentation graphique d'un nuage multidimensionnel

Considérons un nuage moins simplet que le nuage formé des trois points A, B, C . Par exemple, un tableau de données décrivant 16 individus à l'aide de 40 variables. A chaque individu, on associe un point ayant 40 coordonnées. Le nuage des individus est, a priori, dans un espace de dimension 40. En utilisant le raisonnement du § 1.1, on en déduit que ce nuage de $n = 16$ individus est dans un espace dont la dimension est au plus égale à 15. Par conséquent, bien que les individus aient chacun 40 coordonnées, ils sont regroupés en réalité dans un sous-espace de l'espace de dimension 40.

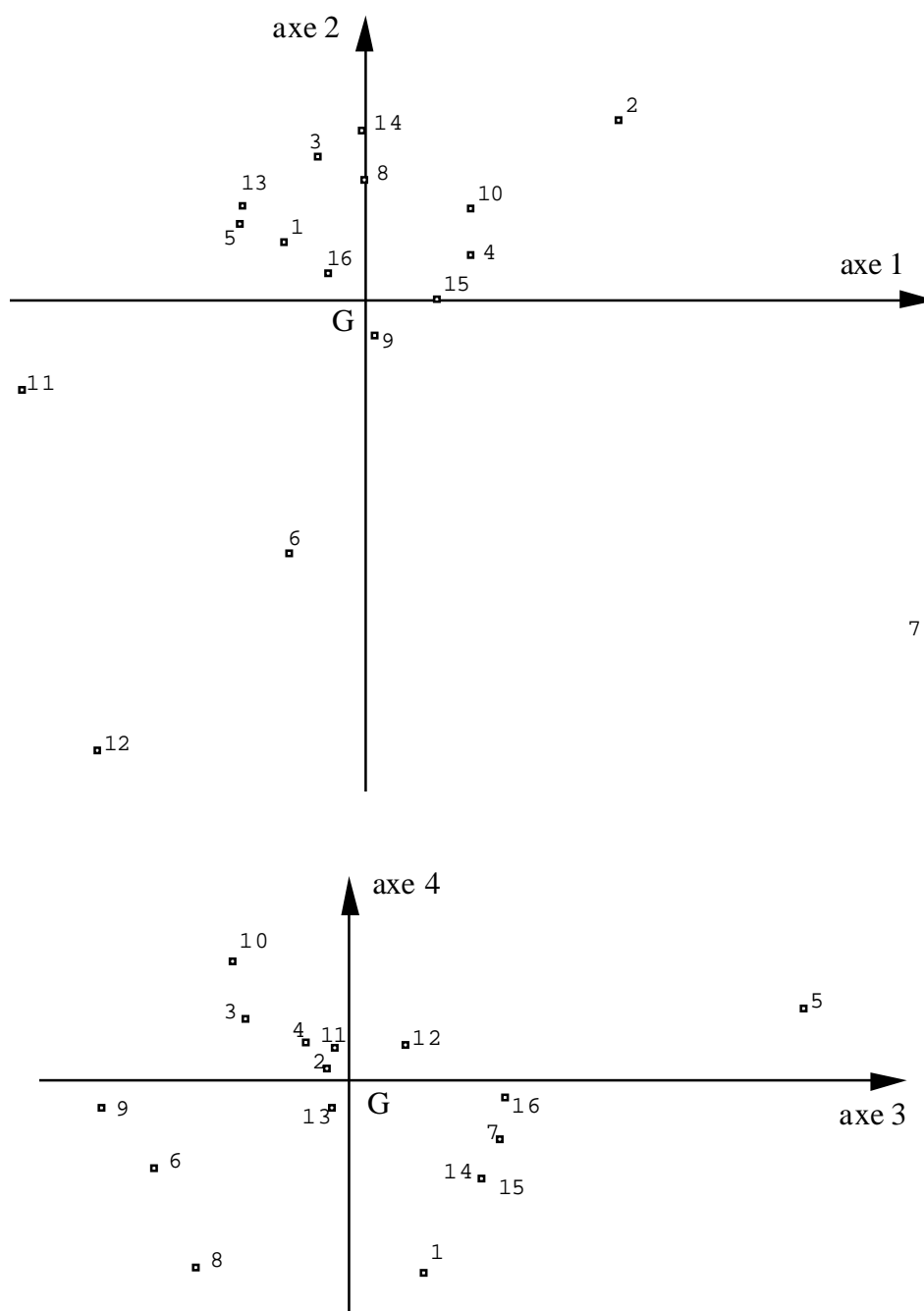
Même en tenant compte de cette réduction de dimension, on ne peut plus dessiner le nuage avec une règle et un compas. Il faut utiliser des techniques mathématiques plus sophistiquées. Pour avoir une idée de la forme du nuage, on choisit un point origine et on décompose l'inertie du nuage par rapport à ce point, en contributions individuelles. On effectue ensuite des coupes du nuage par la technique des axes principaux d'inertie.

Pour notre exemple, les contributions individuelles, calculées en pourcentages de l'inertie du nuage par rapport au centre de gravité, sont les suivantes.

ind n°5				ind. n°7				ind. n°12							
5%	6%	3%	3%	9%	6%	20%	5%	4%	6%	7%	12%	3%	4%	6%	2%

On en conclut que le nuage est formé d'un noyau constitué de treize individus proches de l'individu "moyen", et de trois individus qui fusent hors du noyau (les individus n°7, n°12 et n°5). Les deux cartes factorielles suivantes nous montrent où se situent ces points. On a pris la même échelle sur les deux cartes pour bien mettre en évidence l'effondrement de l'inertie au fur et à mesure qu'on avance dans l'indice de l'axe principal.

Sur cette projection, on retrouve les individus dont on avait remarqué la forte contribution à l'inertie, à savoir les n°7 et n°12. On ne repère pas le n°5 parce qu'on n'a récupéré que 49% de l'inertie totale du nuage. Il manque 51% qui correspondent à "l'épaisseur" du nuage qu'on ne voit pas. La projection sur les deux axes suivants restitue 23% d'inertie supplémentaire et c'est sur cette coupe du nuage que se distingue l'individu n°5.



1.4 Calcul des coordonnées factorielles

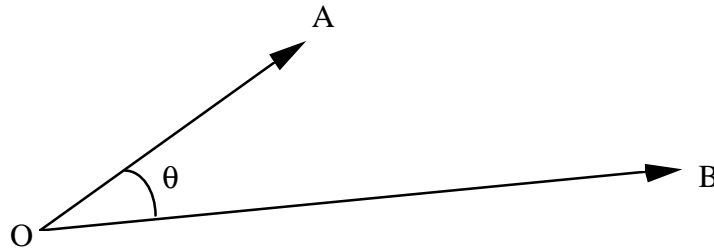
Les coordonnées factorielles s'obtiennent à partir des valeurs propres et des vecteurs propres de la matrice de produits scalaires.

La dimension du sous-espace occupé par le nuage est égale au nombre de valeurs propres non nulles de \mathbf{W}_G . Combien a-t-on de coordonnées factorielles ? Tout dépend du point origine choisi. Dans l'exemple des trois points A, B, C, deux axes suffisent pour reconstruire complètement le nuage si on choisit pour origine un point du plan formé par ces trois points (par exemple le centre de gravité G du triangle). En revanche, si on prend un point origine quelconque, il faudra trois axes.

2. Faut-il comparer des vecteurs normés ou non normés ?

2.1 Représentation graphique lorsque les vecteurs ne sont pas normés

Un produit scalaire est calculé sur des vecteurs. Prenons deux vecteurs quelconques OA et OB. On note $\|OA\|$ la norme du vecteur OA. Le produit scalaire entre OA et OB est égal à : $(OA|OB) = \|OA\| \|OB\| \cos\theta$.



Sur le graphique, on lit la distance entre A et B, la longueur du vecteur OA et celle du vecteur OB, mais on ne peut pas lire le produit scalaire. Il n'y a qu'un seul cas pour lequel on "voit" la valeur du produit scalaire sur le graphique : c'est lorsque les vecteurs sont perpendiculaires. Alors, $\cos \theta = 0$ et le produit scalaire est nul.

2.2 Intérêt de la normalisation

Prenons par exemple un tableau de données de 16 individus et deux variables. A la première colonne du tableau brut, on associe le vecteur OA ayant pour coordonnées : $x_{11}=12,4$ $x_{21}=10,1$... $x_{16,1}=12,2$, et à la deuxième colonne, le vecteur OB. Divisons maintenant chaque vecteur par sa norme :

$$\|OA\| = \sqrt{\sum_{i=1}^{16} x_{i1}^2} = 37,15 \quad \text{et} \quad \|OB\| = \sqrt{\sum_{i=1}^{16} x_{i2}^2} = 12,50.$$

On constate que les deux vecteurs normés sont égaux, ce qui n'était pas évident à la lecture des données brutes.

	données brutes		données normées	
ind. n°1	12,4	4,1	0,33	0,33
ind. n°2	10,1	3,4	0,27	0,27
...	7,7	2,6	0,21	0,21
	9,7	3,2	0,26	0,26
	15	5	0,40	0,40
	1,5	0,5	0,04	0,04
	14,8	4,9	0,40	0,40
	10,5	3,5	0,28	0,28
	1,7	0,6	0,05	0,05

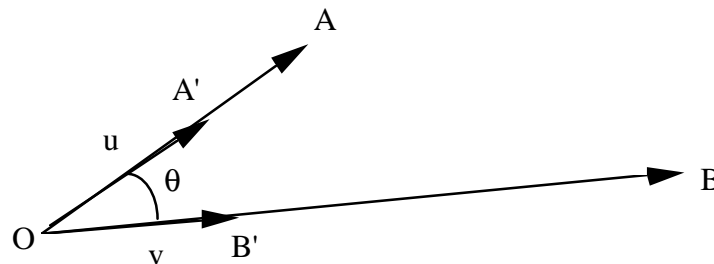
	données brutes	données normées
	7,7 2,6	0,21 0,21
	0 0	0 0
	0 0	0 0
	7,6 2,5	0,20 0,20
	11,7 3,9	0,31 0,31
	2,9 1	0,08 0,08
ind. n°16	12,2 4,1	0,33 0,33

La répartition des individus selon la première variable est, à un changement d'échelle près, la même que celle donnée par la deuxième variable. Deux individus dont les valeurs sont voisines pour la première variable, sont encore voisins pour la deuxième variable. Et inversement, deux individus dont les valeurs diffèrent pour la première variable, restent éloignés pour la deuxième variable. La ressemblance entre les deux vecteurs normés signifie que les deux répartitions ont la même "structure".

2.3 Représentation graphique lorsque les vecteurs sont normés

Notons $OA' = \frac{OA}{||OA||}$, le vecteur OA normé, et $OB' = \frac{OB}{||OB||}$. Le produit scalaire entre

OA' et OB' se réduit à : $(OA'|OB') = \cos \theta$ et la distance entre les points A' et B' se simplifie en : $d(A', B') = \sqrt{1+1-2\cos\theta}$. Les points A' et B' sont à la distance 1 de l'origine O.



Sur le graphique, on lit la distance entre A' et B' , mais aussi le produit scalaire $(OA'|OB')$ avec la règle de lecture suivante : plus l'angle est petit, plus le produit scalaire $(OA'|OB') = \cos$ est proche de 1, plus l'angle se rapproche de 180° , plus le produit scalaire est proche de -1 et un angle droit correspond à un produit scalaire nul.

3. Exemples de calcul de produits scalaires

3.1 Produits scalaires centrés entre individus

Dans l'option n°2 du programme d'analyse conjointe de tableaux, nous comparons les nuages de lignes, chacun des nuages correspondant à un tableau, par l'intermédiaire des matrices de produits scalaires centrés \mathbf{W}_G .

Nous allons donner le calcul de la matrice \mathbf{W}_G pour le tableau de données suivant, croisant $n = 16$ individus et $p = 40$ variables.

individu n°1	0	12,4	7,7	3,5	0	0	0	0	0	0	...
individu n°2	0	10,1	2,6	4,7	0	3	0	5,8	6,8	0	
...	0	7,7	0	3,8	0	0	0	0	2,3	0	
	1,4	9,7	0	2,9	2,5	1,5	3,5	0	5,7	0	
	0	15	0	0	0	0	0	0	0	0	
	0	1,5	0	0	0	0	0	0	3,2	0	
	0	14,8	4,4	0	0	10,2	0	9,4	10,8	0	
	0	10,5	0	0	0	0	0	0	0	0	
	0	1,7	0	0	0	0	0	0	4,3	0	
	0	7,7	7,8	7,7	0	0	0	0	7,7	0	
	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0	0	0	0	
	0	7,6	0	0	0	0	0	0	0	0	
	1,3	11,7	0	0	0	0	0	0	2,5	0	
	0,9	2,9	0	15	1,5	4	1	3,7	2,9	0	
individu n°16	0	12,2	0	3,8	1,5	1,5	1,4	0	3,7	0	

On centre ce tableau en colonnes (on enlève à chaque colonne sa moyenne), ce qui a pour effet de déplacer l'origine sur l'individu "moyen" G .

individu n°1	-0,23	4,56	6,29	0,91	-0,34	-1,26	-0,37	-1,18	-3,12	...
individu n°2	-0,23	2,26	1,19	2,11	-0,34	1,74	-0,37	4,62	3,68	
...	-0,23	-0,14	-1,41	1,21	-0,34	-1,26	-0,37	-1,18	-0,82	
	1,18	1,86	-1,41	0,31	2,16	0,24	3,13	-1,18	2,58	
	-0,23	7,16	-1,41	-2,59	-0,34	-1,26	-0,37	-1,18	-3,12	
	-0,23	-6,34	-1,41	-2,59	-0,34	-1,26	-0,37	-1,18	0,08	
	-0,23	6,96	2,99	-2,59	-0,34	8,94	-0,37	8,22	7,68	
	-0,23	2,66	-1,41	-2,59	-0,34	-1,26	-0,37	-1,18	-3,12	
	-0,23	-6,14	-1,41	-2,59	-0,34	-1,26	-0,37	-1,18	1,18	
	-0,23	-0,14	6,39	5,11	-0,34	-1,26	-0,37	-1,18	4,58	
	-0,23	-7,84	-1,41	-2,59	-0,34	-1,26	-0,37	-1,18	-3,12	
	-0,23	-7,84	-1,41	-2,59	-0,34	-1,26	-0,37	-1,18	-3,12	
	-0,23	-0,24	-1,41	-2,59	-0,34	-1,26	-0,37	-1,18	-3,12	
	1,08	3,86	-1,41	-2,59	-0,34	-1,26	-0,37	-1,18	-0,62	
	0,68	-4,94	-1,41	12,41	1,16	2,74	0,63	2,52	-0,22	
individu n°16	-0,23	4,36	-1,41	1,21	1,16	0,24	1,03	-1,18	0,58	

Soient M_j et $M_{j'}$, les points du nuage, associés aux individus $n^\circ i$ et $n^\circ i'$. Le produit scalaire $(GM_i | GM_{i'}) = \sum_{j=1}^{40} x_{ij} x_{i'j}$ est l'élément (i, i') de la matrice \mathbf{W}_G .

Par exemple :

$$\begin{aligned} (\mathbf{W}_G)_{11} &= (GM_1 | GM_1) = \|GM_1\|^2 = (-0,23)^2 + (4,56)^2 + (6,29)^2 + \dots \\ (\mathbf{W}_G)_{12} &= (\mathbf{W}_G)_{21} = (GM_1 | GM_2) = (-0,23)^2 + (4,56 \times 2,26) + (6,29 \times 1,19) + \dots \end{aligned}$$

La distance entre les individus $n^\circ 1$ et $n^\circ 2$ est égale à

$$d(1,2) = \sqrt{0 + (4,56 - 2,26)^2 + (6,29 - 1,19)^2 + \dots}$$

mais aussi (cf § 1.2) à $\sqrt{\|GM_1\|^2 + \|GM_2\|^2 - 2(GM_1 | GM_2)}$.

3.2 Produits scalaires entre variables

Dans l'option $n^\circ 3$ du programme d'analyse conjointe de tableaux, nous comparons les matrices de corrélations entre les variables, chaque matrice correspondant à un tableau. La corrélation entre la variable j et la variable j' se calcule à partir du tableau de données centré et réduit en colonnes (on divise chaque colonne, centrée au préalable, par son écart-type) :

individu $n^\circ 1$	-0,47	0,92	2,35	0,23	-0,46	-0,49	-0,41	-0,44	-0,98	...
individu $n^\circ 2$	-0,47	0,45	0,45	0,54	-0,46	0,67	-0,41	1,74	1,16	
...	-0,47	-0,03	-0,53	0,31	-0,46	-0,49	-0,41	-0,44	-0,26	
	2,46	0,37	-0,53	0,08	2,9	0,09	3,47	-0,44	0,81	
	-0,47	1,44	-0,53	-0,66	-0,46	-0,49	-0,41	-0,44	-0,98	
	-0,47	-1,28	-0,53	-0,66	-0,46	-0,49	-0,41	-0,44	0,03	
	-0,47	1,4	1,12	-0,66	-0,46	3,44	-0,41	3,09	2,42	
	-0,47	0,53	-0,53	-0,66	-0,46	-0,49	-0,41	-0,44	-0,98	
	-0,47	-1,24	-0,53	-0,66	-0,46	-0,49	-0,41	-0,44	0,37	
	-0,47	-0,03	2,39	1,3	-0,46	-0,49	-0,41	-0,44	1,44	
	-0,47	-1,58	-0,53	-0,66	-0,46	-0,49	-0,41	-0,44	-0,98	
	-0,47	-1,58	-0,53	-0,66	-0,46	-0,49	-0,41	-0,44	-0,98	
	-0,47	-0,05	-0,53	-0,66	-0,46	-0,49	-0,41	-0,44	-0,98	
	2,25	0,78	-0,53	-0,66	-0,46	-0,49	-0,41	-0,44	-0,19	
	1,41	-0,99	-0,53	3,15	1,55	1,05	0,7	0,95	-0,07	
individu $n^\circ 16$	-0,47	0,88	-0,53	0,31	1,55	0,09	1,14	-0,44	0,18	

A partir des termes x_{ij} du tableau centré et réduit, la corrélation $corr(j, j')$ entre la variable j et la variable j' , est égale à $\frac{1}{16} \sum_{i=1}^{16} x_{ij} x_{i'j'}$. La corrélation est un produit scalaire, particulier, entre les vecteurs colonnes du tableau centré et réduit.

Appliqué aux vecteurs colonnes j et j' du tableau centré du § 3.1, ce produit scalaire est égal à la covariance $cov(j, j')$ entre les variables j et j' , et la norme du vecteur colonne j est égale à l'écart-type $\sigma(j)$ de la variable j . Puisque $corr(j, j') = \frac{cov(j, j')}{\sigma(j)\sigma(j')}$, on remarque que les corrélations sont des produits scalaires entre vecteurs normés. La matrice des corrélations \mathbf{V} est par conséquent une matrice de produits scalaires normés, mais non centrés (puisque le tableau ci-dessus n'est pas centré en lignes), du type \mathbf{W}_O .

3.3 Nuage des variables

Pour le produit scalaire défini au § 3.2, les vecteurs sont normés. Ce qui donne un nuage de points très particulier : les points sont tous à la distance 1 de l'origine O . Ils sont donc répartis sur l'écorce d'une sphère multidimensionnelle de centre O et de rayon 1.

Pour obtenir une représentation graphique de ce nuage de variables, on diagonalise la matrice des produits scalaires \mathbf{V} , et on effectue une projection plane sur les axes principaux. Le plan coupe la sphère selon le cercle de centre O et de rayon 1, et les points apparaissent à l'intérieur de ce cercle. Une telle représentation graphique est appelée communément "cercle des corrélations".

Remarque

Les corrélations sont interprétables d'un point de vue statistique et d'un point de vue géométrique puisque ce sont des produits scalaires normés. La matrice de corrélations est, par conséquent, lisible. En revanche, le nuage des variables centrées et réduites n'a pas beaucoup d'intérêt car on ne s'intéresse pas, en général, à la distance entre deux variables. Ce nuage est alors utilisé comme un simple résumé visuel de l'information contenue dans la matrice de corrélations. Et, en plus, ce résumé est peu fidèle puisqu'on n'observe le nuage qu'à travers quelques coupes.

4. Procédure STTIS, option n°1

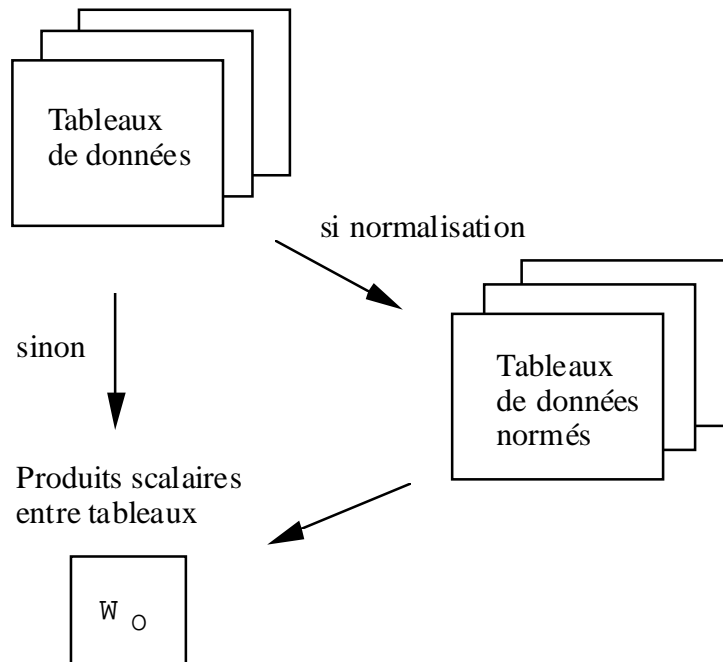
Dans l'option n°1, on compare directement les tableaux de données. Les lignes des tableaux doivent correspondre aux mêmes objets, les colonnes également. Les données peuvent être de n'importe quelle nature.

4.1 Produits scalaires et distances entre tableaux de données

On peut imaginer le tableau de données \mathbf{X}_k comme un grand vecteur de $n_6 p$ éléments, obtenu en empilant les colonnes les unes sur les autres. Au premier tableau de données \mathbf{X}_1 , on associe un premier vecteur OA . Au deuxième tableau de données \mathbf{X}_2 , on associe un deuxième vecteur OB .

On calcule ensuite le produit scalaire $(OA|OB)$ et on en déduit la distance entre les deux tableaux $d(\mathbf{X}_1, \mathbf{X}_2) = \sqrt{\|OA\|^2 + \|OB\|^2 - 2(OA|OB)}$.

Dans le cas où seule la structure interne du tableau nous intéresse, on norme au préalable les vecteurs OA et OB. La norme $\|OA\|$ du vecteur OA est égale à la racine carrée de la somme des éléments du tableau au carré. On calcule ensuite le produit scalaire entre les deux vecteurs normés et on en déduit comme précédemment leur distance. Lorsqu'on choisit de normer les tableaux, le produit scalaire est évidemment compris entre -1 et $+1$.



4.2 Configuration de l'interstructure

A partir de la matrice de produits scalaires entre tableaux, on peut construire un nuage dont chaque point correspond à un tableau.

Pour obtenir une représentation graphique de ce "nuage de tableaux", on prend comme origine le centre de gravité G de ce nuage. On calcule la matrice \mathbf{W}_G , des produits scalaires entre tableaux exprimés depuis le point origine G, qu'on diagonalise ensuite pour obtenir des coupes planes du nuage par la technique des axes principaux.

4.3 Tests de permutation

Lorsque les tableaux de données sont centrés (en lignes ou en colonnes, ou bicentrés), et qu'on utilise l'option n°1 avec normalisation, le produit scalaire est alors un banal coefficient de corrélation entre les deux grands vecteurs de $n \times p$ éléments.

Il est alors possible de construire un test de signification du produit scalaire. Pour cela, on compare la valeur du produit scalaire avec les $n!$ valeurs obtenues en permutant les n lignes d'un des deux tableaux. Un produit scalaire, suffisamment élevé (en valeur absolue) pour ne pas être dû qu'au seul hasard, sera signalé par une étoile. Cela indique que les tableaux sont "corrélés" de façon significative (cette corrélation pouvant être positive ou négative).

Néanmoins, cette méthode étant généralement utilisée sur des tableaux de données recueillies dans des situations similaires, on ne peut pas tomber sur une corrélation négative. Par conséquent, un produit scalaire signalé par une étoile détectera des tableaux significativement voisins.

4.4 Décomposition de la distance entre tableaux

Le vecteur obtenu en empilant les colonnes du tableau de données est en réalité structuré. Certains éléments de ce vecteur appartiennent à la même ligne du tableau (ce sont les valeurs d'un même individu), ou à la même colonne (ce sont les valeurs d'une même variable). La distance au carré qu'on vient de calculer, s'écrit :

$$d^2(\mathbf{X}_1, \mathbf{X}_2) = \sum_{i=1}^n \sum_{j=1}^p \left((\mathbf{X}_1)_{ij} - (\mathbf{X}_2)_{ij} \right)^2$$

Il est possible de la décomposer en contributions dues aux lignes, puis en contributions dues aux colonnes.

4.4.1 Pour chaque couple de tableaux

La quantité $\sum_{j=1}^p \left((\mathbf{X}_1)_{ij} - (\mathbf{X}_2)_{ij} \right)^2$ mesure l'écart entre les valeurs de l'individu n° i dans

le premier tableau et les valeurs de ce même individu dans le deuxième tableau. Exprimée en pourcentage du carré de la distance entre les deux tableaux, cette quantité mesure la contribution de l'individu n° i dans la dissemblance constatée entre les deux tableaux.

D'une manière analogue, la quantité $\sum_{i=1}^n \left((\mathbf{X}_1)_{ij} - (\mathbf{X}_2)_{ij} \right)^2$ mesure les variations entre

les valeurs de la variable j dans le premier tableau et celles de cette même variable dans le deuxième tableau. Exprimée en pourcentage du carré de la distance, cette quantité mesure la contribution de la variable j dans l'écart constaté entre les deux tableaux.

4.4.2 Globalement pour l'ensemble des q tableaux

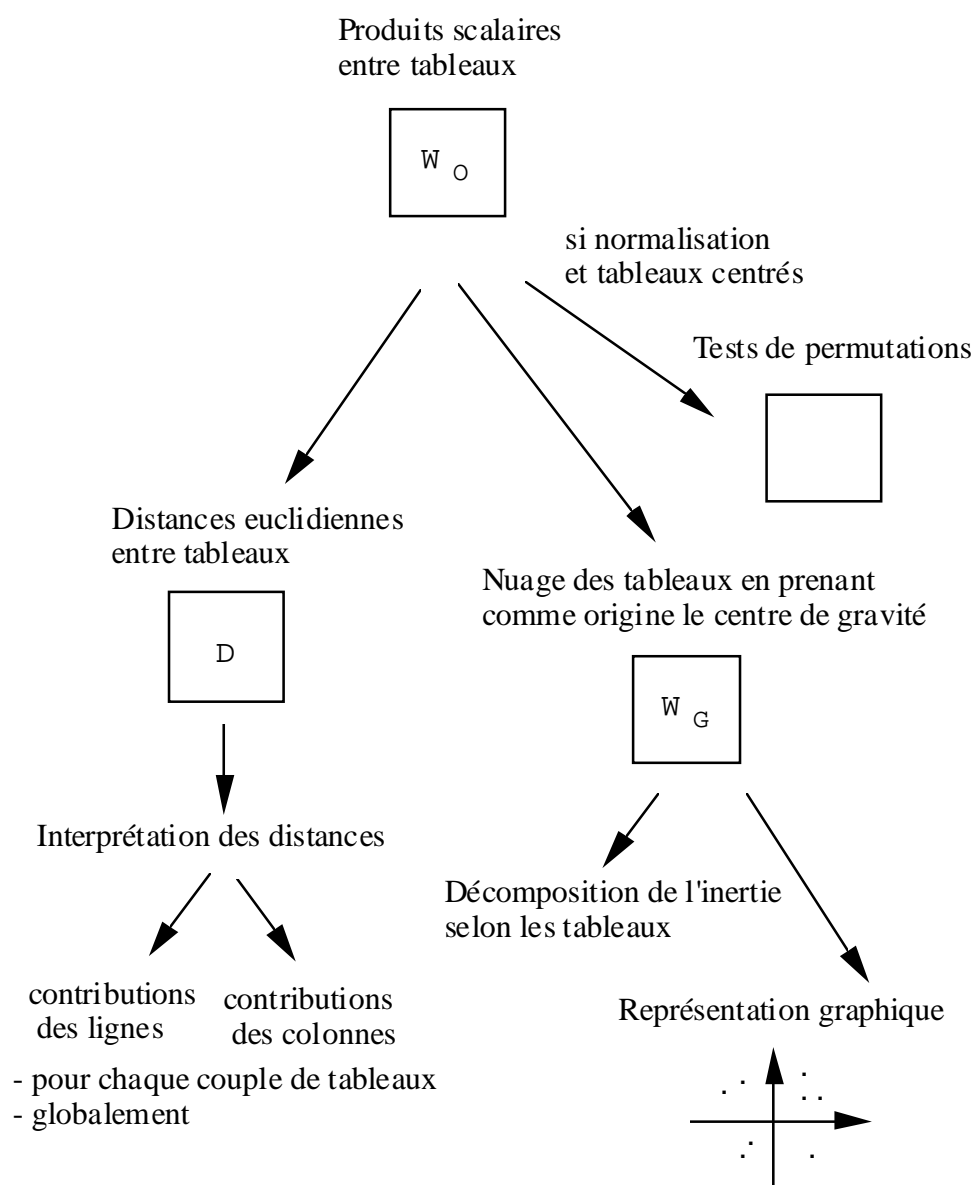
Si on considère maintenant l'ensemble des tableaux, la somme $\sum_{k=1}^q \sum_{k'=1}^q [d^2(\mathbf{X}_k, \mathbf{X}_{k'})]$ des carrés des distances entre chaque couple de tableaux est une mesure de la dispersion des points dans le nuage des tableaux. De la même manière que précédemment, cette quantité se décompose selon les individus et selon les variables :

$$\sum_{k=1}^q \sum_{k'=1}^q \left[\sum_{j=1}^p ((\mathbf{X}_k)_{ij} - (\mathbf{X}_{k'})_{ij})^2 \right] \text{ mesure la part due à l'individu } n^{\circ} i,$$

$$\sum_{k=1}^q \sum_{k'=1}^q \left[\sum_{i=1}^n ((\mathbf{X}_k)_{ij} - (\mathbf{X}_{k'})_{ij})^2 \right] \text{ mesure la part due à la variable } n^{\circ} j.$$

Exprimées en pourcentages de la somme totale, ces contributions donnent une explication globale de la dispersion des points dans le nuage de l'interstructure, par les individus d'une part, et par les variables d'autre part.

4.5 Déroulement du programme



5. Procédure STTIS, option n°2

Dans l'option n°2, on compare les nuages des lignes par l'intermédiaire des matrices $(W_G)_k$ de produits scalaires centrés (cf § 3.1). Les lignes des tableaux X_k doivent correspondre aux mêmes objets d'un tableau à l'autre.

Les "objets" ne sont pas forcément des "individus" : l'option n°2 de la procédure STTIS suppose seulement que la pondération de ces objets est uniforme et que la distance entre deux objets correspond à la distance euclidienne classique entre les deux lignes du tableau.

5.1 Produits scalaires et distances entre nuages de lignes

On utilise le produit scalaire entre matrices $(\mathbf{W}_G)_k$ *normées*, car on ne s'intéresse qu'à la répartition des individus dans le nuage, sans tenir compte de la dilatation de ce nuage dans l'espace. Le produit scalaire entre matrices de type \mathbf{W} est positif. Ici, on l'utilise entre matrices normées, il est donc compris entre 0 et 1. Ce produit scalaire est connu sous le nom de coefficient RV.

Les distances au carré entre nuages d'individus sont égales à :

$$d^2(\mathbf{w}_k, \mathbf{w}_{k'}) = \sum_{i=1}^n \sum_{j=1}^n \left((\mathbf{w}_k)_{ji} - (\mathbf{w}_{k'})_{ji} \right)^2$$

où \mathbf{W}_k désigne la matrice $(\mathbf{W}_G)_k$ normée.

L'interprétation de la distance entre deux nuages multidimensionnels est plus délicate que celle de la distance entre deux tableaux, et cela nécessite quelques précisions.

Si le coefficient RV est égal à 1, la distance, égale (cf § 2.3) à $\sqrt{2 - 2RV}$, est nulle et les deux nuages *normés* sont identiques. Cela signifie que les deux nuages sont les mêmes à une dilatation près.

Hélas, ce n'est jamais le cas. Si on trouve un RV de l'ordre de 0.8 ou 0.9, la distance n'est plus nulle et les nuages normés ne sont plus superposables. Cela peut provenir de plusieurs causes :

- Les nuages peuvent appartenir à des espaces de dimension différente (la dimension d'un nuage a été évoquée dans le § 1.4; elle correspond au nombre de valeurs propres non nulles de la matrice de produits scalaires \mathbf{W}_G).

Ce cas a une interprétation intéressante en évaluation sensorielle. Des produits (lignes) sont évalués au moyen de descripteurs sensoriels (colonnes) par un ensemble de juges (tableaux). Même si les juges ont à leur disposition une dizaine de descripteurs sensoriels, on peut avoir le cas d'un juge qui n'utilise que deux dimensions pour placer les produits (une dimension qui correspond à l'appréciation hédonique du juge, et une autre dimension qui correspond à un descripteur très facile à noter, par exemple le goût salé). Un juge plus expert distinguera plus finement les produits en les plaçant dans un espace de dimension quatre, par exemple.

- Si les nuages sont de même dimension, ils peuvent ne pas avoir la même forme (la forme d'un nuage est donnée par la répartition des valeurs propres de la matrice de produits scalaires \mathbf{W}_G). On peut avoir un nuage "allongé" si la première valeur propre est prépondérante, et un nuage "rond" si les premières valeurs propres sont du même ordre de grandeur.
- Si les nuages ont à peu près la même forme, c'est peut-être la répartition des points à l'intérieur d'un nuage qui diffère d'un nuage à l'autre.

Fort heureusement, dans la plupart des cas, on compare des nuages qui ont grosso modo la même forme, et le programme permet de s'en assurer (on donne, pour chaque nuage, la répartition des valeurs propres de la matrice de produits scalaires \mathbf{W}_G). La distance entre nuages provient alors essentiellement de déformations internes dues à quelques individus dont la position, à l'intérieur du nuage, varie d'un nuage à l'autre.

5.2 Configuration de l'interstructure

Les produits scalaires RV entre matrices \mathbf{W}_k définissent une configuration de q points dans laquelle chaque point M représente un nuage d'individus. Les matrices \mathbf{W}_k étant normées, les vecteurs OM sont de longueur 1. Le produit scalaire RV correspond au cosinus entre ces vecteurs (cf 2.3). Les RV étant compris entre 0 et 1, ces angles sont aigus. La configuration de l'interstructure correspond aux baleines d'un parapluie presque fermé.

5.3 Tests de permutation

Comme dans le cas de l'option n°1, il est possible de construire un test de signification du produit scalaire. Un produit scalaire, suffisamment élevé pour ne pas être dû qu'au seul hasard, sera signalé par une étoile et détectera deux nuages relativement semblables.

5.4 Décomposition de la distance entre nuages

On décompose les distances entre nuages en contributions dues aux individus.

5.4.1 Pour chaque couple de nuages de lignes

La quantité $\sum_{i=1}^n ((\mathbf{W}_1)_{ii} - (\mathbf{W}_2)_{ii})^2$ représente le carré de la distance entre la ligne i de \mathbf{W}_1 et la même ligne i de \mathbf{W}_2 .

La ligne i de \mathbf{W}_1 contient les produits scalaires $(GM_i | GM_i)$ entre l'individu n° i et les $n-1$ autres. De la connaissance des éléments de cette ligne et des éléments diagonaux de \mathbf{W}_1 , on peut en déduire la distance entre l'individu n° i et chacun des $n-1$ autres individus. Autrement dit, cette ligne est caractéristique de la position de l'individu n° i dans le nuage.

La quantité $\sum_{i=1}^n ((\mathbf{W}_1)_{ii} - (\mathbf{W}_2)_{ii})^2$ mesure, par conséquent, le changement de position de l'individu n° i d'un nuage à l'autre. On l'exprime en termes de contribution, c'est-à-dire en pourcentage du carré de la distance entre les deux nuages.

5.4.2 Globalement pour l'ensemble des q nuages de lignes

Si on considère maintenant l'ensemble des q nuages, la somme des carrés des distances entre chaque couple de nuages, c'est-à-dire chaque couple $(\mathbf{W}_k, \mathbf{W}_{k'})$, se décompose selon les lignes en :

$$\sum_{k=1}^q \sum_{k'=1}^q \left[\sum_{i=1}^n ((\mathbf{W}_k)_{ii'} - (\mathbf{W}_{k'})_{ii'})^2 \right]$$
 qui mesure la part due à l'individu n° i dans la dispersion des points dans la configuration de l'interstructure. Ces quantités sont exprimées en pourcentage.

5.5 Nuage des lignes compromis

Calcul de la matrice compromis

On sait calculer la matrice compromis \mathbf{W} la plus "proche" (au sens du produit scalaire RV) des matrices $\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_q$. Cette matrice compromis est une moyenne pondérée des q matrices précédentes qui satisfait à la condition suivante : parmi toutes les combinaisons linéaires \mathbf{A} des \mathbf{W}_k , c'est celle qui maximise

$$\sum_{k=1}^q (\text{RV}(\mathbf{A}, \mathbf{W}_k))^2$$
. La matrice compromis est encore une matrice normée.

En d'autres termes, la somme des cosinus au carré entre le vecteur OC qui représente le compromis et les vecteurs OM_k qui représentent les nuages, doit être maximum. Il faut donc placer le vecteur OC dans le parapluie de telle sorte que les angles entre OC et les OM_k soient les plus petits possibles.

Nuage compromis

La matrice compromis est encore une matrice de produits scalaires centrés entre les lignes, de type \mathbf{W}_G , à partir de laquelle on construit le nuage compromis qui résume "au mieux" les positions mutuelles des individus.

Contrairement à une matrice de corrélations, la matrice compromis ne mérite pas qu'on l'édite car, excepté pour les termes diagonaux qui donnent le carré de la distance des individus compromis au centre de gravité, elle est totalement illisible.

Interprétation du compromis

Le compromis n'est pas une moyenne, au sens habituel. Pour le comprendre, reprenons l'exemple de l'évaluation sensorielle donné au § 5.1. Si plusieurs juges ont des opinions concordantes, ils fournissent des nuages de produits relativement semblables. Les RV, c'est-à-dire les cosinus, sont proches de 1, et les angles sont pratiquement nuls (le parapluie du § 5.2 est presque fermé). Inversement, un juge qui ne serait pas en accord avec les autres, fournirait un nuage de produits particulier qui aurait un faible RV avec les nuages précédents, donc un angle aigu assez ouvert (ce juge correspond à une baleine du parapluie qui serait arrachée).

Le vecteur OC qui correspond au compromis, ne peut pas être placé ailleurs qu'au milieu des baleines fermées du parapluie, donc loin de la baleine arrachée. En d'autres termes, le compromis est un résumé de l'opinion majoritaire qui pénalise le juge qui n'est pas en accord avec les autres.

5.6 Graphique de l'interstructure

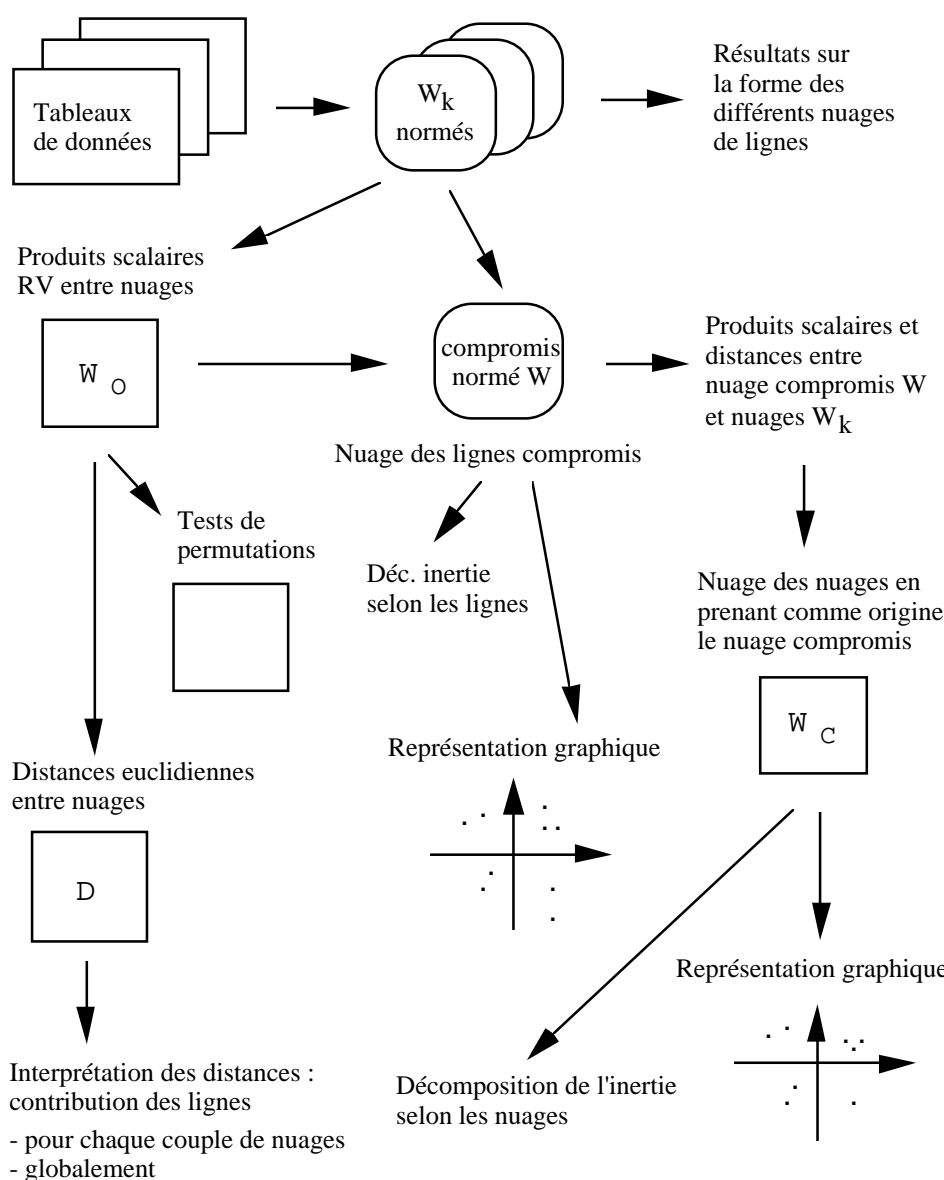
Pour obtenir une représentation graphique du nuage de l'interstructure ("nuage des nuages" puisque chaque point représente un nuage d'individus), on prend comme origine le point C correspondant au nuage compromis (on ne s'intéresse qu'aux extrémités des baleines du parapluie du § 5.2).

Il faut donc connaître la matrice \mathbf{W}_C des produits scalaires entre matrices \mathbf{W}_k , calculés depuis le point origine C. Soit M_k (respectivement $M_{k'}$), le point correspondant au nuage k (respectivement k'). Le produit scalaire $(CM_k | CM_{k'})$ est égal (cf § 1.2) à :

$$\frac{1}{2} [d^2(\mathbf{w}, \mathbf{w}_k) + d^2(\mathbf{w}, \mathbf{w}_{k'}) - d^2(\mathbf{w}_k, \mathbf{w}_{k'})].$$

On diagonalise ensuite la matrice \mathbf{W}_C pour obtenir des coupes planes du nuage par la technique des axes principaux.

5.7 Déroulement du programme



5.8 Comparaison avec la procédure AFMUL

L'option 2 de la procédure STTIS et la procédure AFMUL (analyse factorielle multiple) sont très voisines dans leur conception mathématique puisqu'elles traitent le même type de données (des groupes de variables sur les mêmes individus) avec un objectif commun : la construction d'un nuage d'individus "résumé". Les deux méthodes diffèrent essentiellement par la construction de ce nuage "résumé" et l'exploitation des résultats annexes.

Dans les deux méthodes, les nuages sont au préalable "normés" pour pouvoir être comparés. La normalisation est de type géométrique dans AFMUL (on ramène l'inertie axiale maximum du nuage à 1) et de type algébrique dans STTIS (on norme la matrice \mathbf{W}_G), ce qui produit des effets à peu près similaires.

Plus précisément, si on note $\lambda_1, \dots, \lambda_s, \dots$ les valeurs propres rangées par ordre décroissant de la matrice des produits scalaires \mathbf{W}_G (c'est-à-dire l'inertie expliquée par les axes principaux de l'analyse en composantes principales du tableau), le coefficient de contraction du nuage est égal à $\frac{1}{\lambda_1}$ dans AFMUL et $\frac{1}{\sqrt{\lambda_1^2 + \dots + \lambda_s^2 + \dots}}$ dans STTIS.

En revanche, la différence porte sur le calcul du nuage "résumé". Ce nuage est obtenu en faisant la moyenne des nuages contractés dans AFMUL, alors que dans STTIS, il s'agit d'une moyenne pondérée des nuages contractés qui a pour effet de donner un poids plus important aux nuages qui se ressemblent et de pénaliser un nuage qui serait à l'écart des autres.

Le compromis de STTIS est intéressant dans la mesure où on cherche à extraire des données une opinion majoritaire. Le "résumé" de AFMUL a l'avantage de ne pas être sensible à la dimension des différents nuages.

6. Procédure STTIS, option n°3

Dans l'option n°3, on compare les nuages des variables centrées et réduites par l'intermédiaire des matrices de corrélations (cf § 3.2). Il faut que les tableaux soient du type : individus x variables continues et qu'ils concernent les mêmes variables.

6.1 Produits scalaires entre matrices de corrélations

Nous avons rappelé (§ 3.3) que lorsque les variables sont centrées et réduites, les nuages des variables sont "calibrés" car sphériques de rayon 1. Il est par conséquent inutile de normer les matrices \mathbf{V}_k avant de calculer le produit scalaire. On a encore un produit scalaire entre matrices de type \mathbf{W} , c'est un produit scalaire positif. Puisque les matrices ne sont pas normées, le produit scalaire n'est plus compris entre 0 et 1.

Les distances au carré entre matrices de corrélations correspondent aux distances

$$d^2(\mathbf{v}_k, \mathbf{v}_{k'}) = \sum_{j=1}^p \sum_{j'=1}^p \left((\mathbf{v}_k)_{jj'} - (\mathbf{v}_{k'})_{jj'} \right)^2$$

qu'on décompose ensuite en contributions dues aux variables.

6.2 Configuration de l'interstructure

Les produits scalaires entre matrices \mathbf{V}_k définissent une configuration de q points dans laquelle chaque point représente une des q matrices de corrélations (ou nuages de variables).

6.3 Décomposition de la distance

6.3.1 Pour chaque couple de matrices de corrélations

La quantité $\sum_{j=1}^p \left((V_1)_{jj'} - (V_2)_{jj'} \right)^2$ représente le carré de la distance entre la colonne j de V_1 et la colonne j de V_2 .

La colonne j de V_1 contient les corrélations entre la variable $n^\circ j$ et les $p-1$ autres. Autrement dit, la liaison de la variable $n^\circ j$ avec les autres, ou encore la position de la variable $n^\circ j$ dans le nuage des variables.

La quantité $\sum_{j=1}^p \left((V_1)_{jj'} - (V_2)_{jj'} \right)^2$ mesure la modification des relations linéaires de la variable $n^\circ j$ avec les autres, entre les deux tableaux. On l'exprime en termes de contribution, c'est-à-dire en pourcentage du carré de la distance entre les deux matrices de corrélations.

6.3.2 Globalement pour l'ensemble des q matrices de corrélations

Si on considère maintenant l'ensemble des q matrices de corrélations, la somme des carrés des distances entre chaque couple $(V_k, V_{k'})$ se décompose selon les variables :

$\sum_{k=1}^q \sum_{k'=1}^q \left[\sum_{j=1}^p \left((V_k)_{jj'} - (V_{k'})_{jj'} \right)^2 \right]$ mesure la part due à la variable $n^\circ j$ dans la dispersion des points du nuage de l'interstructure. Ces quantités sont exprimées en pourcentage.

6.4 Matrice de corrélations compromis

Comme dans l'option n°2, on sait calculer la matrice compromis V la plus proche (au sens du produit scalaire précédent) des q matrices de corrélations V_1, V_2, \dots, V_q . Cette matrice V est une moyenne pondérée des q matrices précédentes. C'est encore une matrice de corrélations.

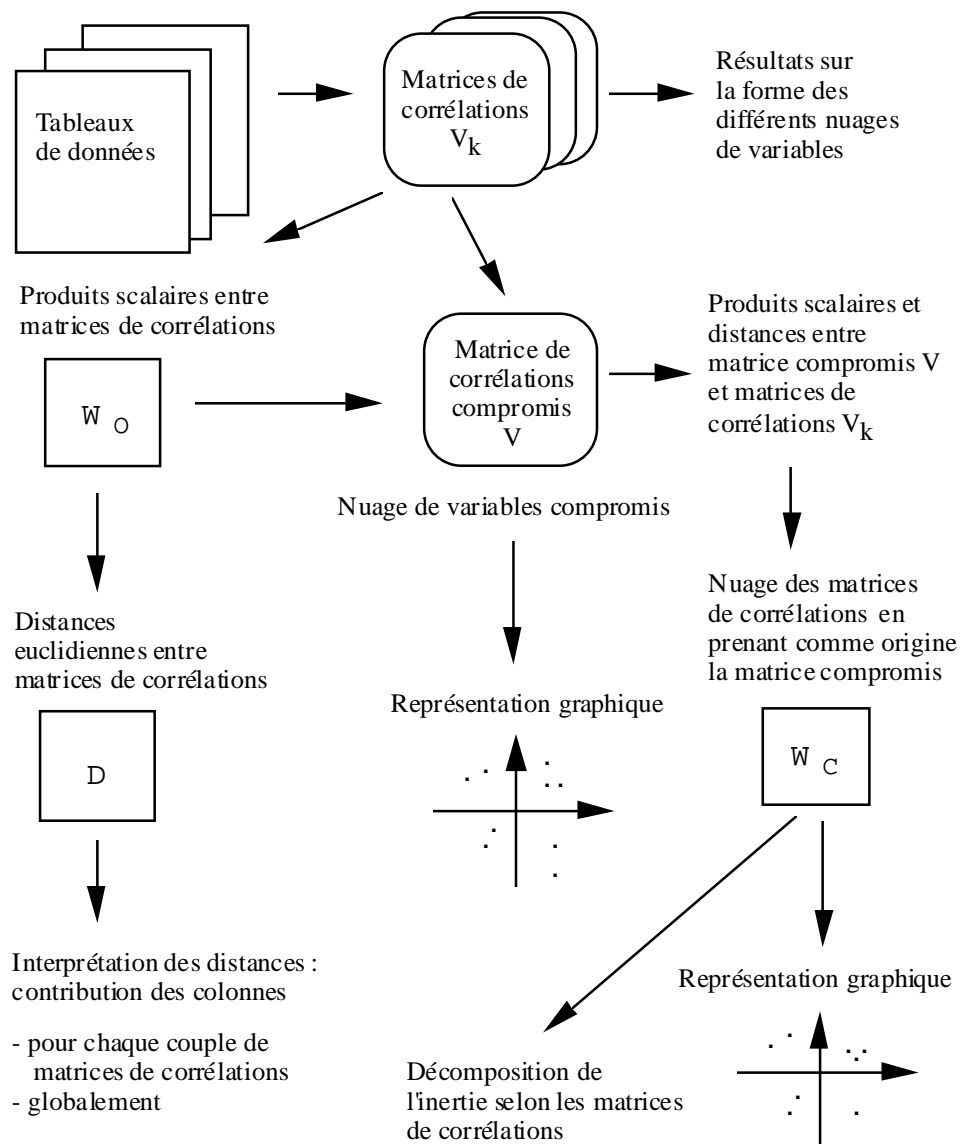
A partir de la matrice V , on construit le nuage compromis des variables, dont les projections planes sur les axes principaux donnent les cercles de corrélations compromis.

6.5 Graphique de l'interstructure

Pour obtenir une représentation graphique du nuage de l'interstructure ("nuage des matrices de corrélations" puisque chaque point représente une matrice de corrélations), on prend comme origine le point C correspondant à la matrice de corrélations compromis. Il faut donc connaître la matrice W_C des produits scalaires entre matrices V_k , calculés depuis le point origine C.

On diagonalise ensuite la matrice W_C pour obtenir des coupes planes du nuage par la technique des axes principaux.

6.6 Déroulement du programme



7. Références

DAZE F. - LE BARZIC J. F. (1996) - L'Analyse des données évolutives, méthodes et applications - Technip.

ESCOUFIER Y. (1973) – Le traitement des variables vectorielles – Biometrics, 29, 751-760.

JAFFRENOU P. A. (1978) – Sur l'analyse des familles finies de variables vectorielles – Thèse de 3^e cycle, Lyon I.

LAVIT Ch. (1988) – Analyse conjointe de tableaux quantitatifs – Masson.

KAZI-AOUAL F. (1993) – Approximations to permutation tests for data analysis – Rapport de recherche 93-06, biométrie, ENSA.M, Montpellier.

LAVIT Ch et al. (1994) – The ACT (STATIS method) – Computational Statistics & Data Analysis, 18, 97-119.

Exemple 2 : Analyse Conjointe de Tableaux - Méthode STATIS

Dans cette partie, nous présentons l'application des trois options sur un exemple.

- 1) Comparaison de tableaux
- 2) Comparaison de nuages de lignes
- 3) Comparaison de matrices de corrélations

Les données portent sur les profils socioprofessionnels des 30 cantons ruraux de l'Hérault, au cours des recensements de 1954, 1962, 1968, 1975 et 1982. Pour les quatre premiers recensements, les professions sont réparties en 9 classes (exploitants agricoles, ouvriers agricoles, industriels et commerçants, professions libérales, cadres moyens, employés, ouvriers, services, clergé armée police). En 82, la nomenclature de l'INSEE a été modifiée (les "cadres moyens" s'appellent dorénavant "professions intermédiaires"), et la classe "clergé, armée, police" a été supprimée et répartie, suivant le grade, dans les classes "professions intermédiaires" et "employés". Le fichier de données comporte donc 30 lignes et $(9 \times 4) + 8$ colonnes.

Pour utiliser l'option n°1, il faut que les tableaux (correspondant aux recensements), portent sur les mêmes individus et les mêmes variables. On ne peut comparer que les quatre premiers tableaux.

Dans l'option n°2, on compare les nuages d'individus, les variables pouvant être différentes d'un tableau à l'autre. On peut traiter l'ensemble des cinq recensements.

Pour utiliser l'option n°3, il faut que les variables soient les mêmes pour tous les tableaux. On ne peut comparer que les quatre premiers tableaux.

Les cantons ruraux de l'Hérault sont à dominante agricole pour la moitié d'entre eux, industrielle pour quelques uns et tertiaire pour le reste. Ce qui entraîne que les pourcentages des catégories socioprofessionnelles ne sont pas du même ordre de grandeur. Les pourcentages des catégories "services" et "clergé, armée, police" sont de l'ordre de quelques unités. Les pourcentages des catégories agricoles sont de l'ordre de quelques dizaines. Nous avons par conséquent utilisé les trois options de la procédure STTIS sur les données centrées et réduites par tableau.

1. Sorties de l'option n°1 : Comparaison des tableaux non normés

1.1 Normes des tableaux

```
1. NORMES DES TABLEAUX
NORME DU TABLEAU 1 =      16.43
NORME DU TABLEAU 2 =      16.43
NORME DU TABLEAU 3 =      16.43
NORME DU TABLEAU 4 =      16.43
```

Ce résultat permet de contrôler que les normes sont du même ordre de grandeur et que, par conséquent, il n'est pas nécessaire de normer les tableaux pour les comparer. Ici, les tableaux ($n \times p$) du fichier NDON sont centrés et réduits en colonnes. On a donc

$$\frac{1}{n} \sum_i \sum_j x_{ij}^2 = p \text{ et la norme de chacun des tableaux est égale à } \sqrt{np} = \sqrt{30 \times 9}.$$

1.2 Situation des tableaux les uns par rapport aux autres

```
PRODUITS SCALAIRES ENTRE TABLEAUX
-----|-----
      | 1      2      3      4
-----|-----
1 | 270.000
2 | 229.832 270.000
3 | 203.253 234.850 270.000
4 | 156.759 181.595 206.877 270.000
-----|-----
      | 1      2      3      4
```

Lorsque la normalisation des tableaux n'est pas demandée (LNORM = 0), il n'y a pas de test de permutations.

```
DISTANCES EUCLIDIENNES ENTRE TABLEAUX
-----|-----
      | 1      2      3      4
-----|-----
1 | 0.000
2 | 8.963 0.000
3 | 11.554 8.385 0.000
4 | 15.049 13.297 11.236 0.000
-----|-----
      | 1      2      3      4
```

On constate que l'écart entre les tableaux se creuse avec le temps, avec une rupture entre le tableau n°3 (1968) et le tableau n°4 (1975). Le tableau n°4 est assez éloigné des trois autres.

1.3 Interprétation des distances entre les tableaux

DECOMPOSITION DE LA DISTANCE AU CARRE ENTRE LE TABLEAU 4 ET LES AUTRES
CONTRIBUTIONS DES LIGNES EN POURCENTAGE

		TABLEAU		
LIGNE	MOYENNE	1	2	3
AGDE	1.3	0.9	1.7	1.1
BEDARIEUX	3.9	5.2	2.0	4.6
CAPESTANG	1.1	0.7	0.8	1.7
FLORENSAC	3.1	2.1	1.9	5.4
MONTAGNAC	1.3	0.9	1.3	1.7
MURVIEL	1.3	1.2	1.1	1.4
OLARGUES	2.4	2.6	2.4	2.2
OLONZAC	1.6	2.2	1.5	1.0
PEZENAS	1.1	1.3	0.9	1.2
ROUJAN	2.1	1.5	1.8	3.0
ST-CHINIAN	1.2	0.6	1.0	2.0
ST-GERVAIS	5.3	4.7	5.0	6.2
ST-PONS	7.1	7.7	8.8	4.7
SALVETAT	5.7	6.0	3.8	7.4
SERVIAN	1.1	1.3	1.0	1.1
LE-CAYLAR	5.4	3.1	4.2	8.7
CLERMONT	0.8	0.3	0.8	1.2
GIGNAC	0.9	1.1	0.6	0.9
LODEVE	2.5	1.7	3.2	2.5
LUNAS	3.6	2.6	4.4	3.8
ANIANE	3.5	3.6	2.3	4.5
CASTRIES	4.6	6.5	5.4	1.7
CLARET	5.3	4.2	5.9	5.9
FRONTIGNAN	2.0	1.6	2.1	2.3
GANGES	4.1	6.4	3.7	2.1
LUNEL	1.9	1.9	1.0	2.9
MATELLES	12.1	16.8	15.3	4.1
MAUGUIO	7.6	5.2	8.5	9.1
MEZE	0.8	0.9	0.7	0.7
ST-MARTIN	5.6	5.3	6.6	4.9

DECOMPOSITION DE LA DISTANCE AU CARRE ENTRE LE TABLEAU 4 ET LES AUTRES
CONTRIBUTIONS DES COLONNES EN POURCENTAGE

		TABLEAU		
COLONNE	MOYENNE	1	2	3
EAGR	3.4	4.1	3.0	3.0
OAGR	6.5	6.9	6.3	6.4
INDU	12.1	12.0	9.5	14.8
LIBE	22.3	28.1	27.6	11.3
CADR	16.1	12.7	21.1	14.6
EMPL	12.5	13.4	8.9	15.0
OUVR	5.0	4.2	4.4	6.4
SERV	11.5	11.0	10.5	13.1
CLER	10.5	7.4	8.7	15.4

L'utilisateur peut demander la décomposition de la distance au carré entre un tableau particulier et les autres tableaux en précisant le numéro LNUMT de ce tableau. Dans notre exemple, on a demandé la décomposition de la distance au carré entre le tableau n°4 et les trois autres (LNUMT = 4).

La première colonne du tableau des contributions correspond à la moyenne des trois autres. Un balayage rapide de cette colonne montre que l'écart constaté entre le tableau n°4 et les autres tableaux est essentiellement dû au canton des Matelles, et aux catégories socioprofessionnelles "professions libérales" et, dans une moindre mesure "cadres moyens". L'examen des trois autres colonnes permet d'interpréter plus finement la distance entre le tableau n°4 et chacun des trois autres tableaux.

1.4 Etude du nuage des tableaux en prenant comme origine le centre de gravité

Pour obtenir une représentation graphique de la configuration de l'interstructure dans laquelle chaque point représente un tableau, on prend comme origine le centre de gravité de ces quatre points. Et on fournit les informations nécessaires pour comprendre comment sont situés les tableaux les uns par rapport aux autres.

LA SOMME DES DISTANCES AU CARRE (PROPORTIONNELLE A L'INERTIE DU NUAGE) SE DECOMPOSE
SELON LES LIGNES

AGDE	=	1.45 %
BEDARIEUX	=	3.53 %
CAPESTANG	=	1.03 %
FLORENSAC	=	2.37 %
MONTAGNAC	=	0.96 %
MURVIEL	=	1.52 %
OLARGUES	=	2.19 %
OLONZAC	=	1.68 %
PEZENAS	=	1.82 %
ROUJAN	=	2.13 %
ST-CHINIAN	=	1.10 %
ST-GERVAIS	=	5.36 %
ST-PONS	=	6.41 %
SALVETAT	=	5.22 %
SERVIAN	=	1.03 %
LE-CAYLAR	=	4.16 %
CLERMONT	=	0.74 %
GIGNAC	=	1.08 %
LODEVE	=	2.64 %
LUNAS	=	2.74 %
ANIANE	=	3.32 %
CASTRIES	=	5.17 %
CLARET	=	4.61 %
FRONTIGNAN	=	2.68 %
GANGES	=	4.98 %
LUNEL	=	2.09 %
MATELLES	=	15.81 %
MAUGUIO	=	5.78 %
MEZE	=	0.99 %
ST-MARTIN	=	5.39 %

SELON LES COLONNES

EAGR	=	3.39 %
OAGR	=	5.62 %
INDU	=	9.47 %
LIBE	=	24.58 %
CADR	=	15.41 %
EMPL	=	12.34 %
OUVR	=	4.97 %
SERV	=	12.34 %
CLER	=	11.87 %

SELON LES TABLEAUX

TABLEAU 1	=	29.11 %
TABLEAU 2	=	15.24 %
TABLEAU 3	=	15.56 %
TABLEAU 4	=	40.08 %

L'inertie du nuage des tableaux par rapport au centre de gravité est proportionnelle à la somme des distances au carré entre tous les couples de tableaux. On peut donc décomposer cette inertie en contribution des lignes et des colonnes (comme on l'a fait dans le paragraphe précédent pour la distance au carré entre le tableau n°4 et les trois autres tableaux). On obtient une information très synthétique où on retrouve encore que la dispersion du nuage des tableaux est essentiellement due au canton des Matelles et aux professions libérales. Enfin, la décomposition de cette inertie selon les tableaux renseigne sur l'éloignement de chaque tableau par rapport au centre de gravité.

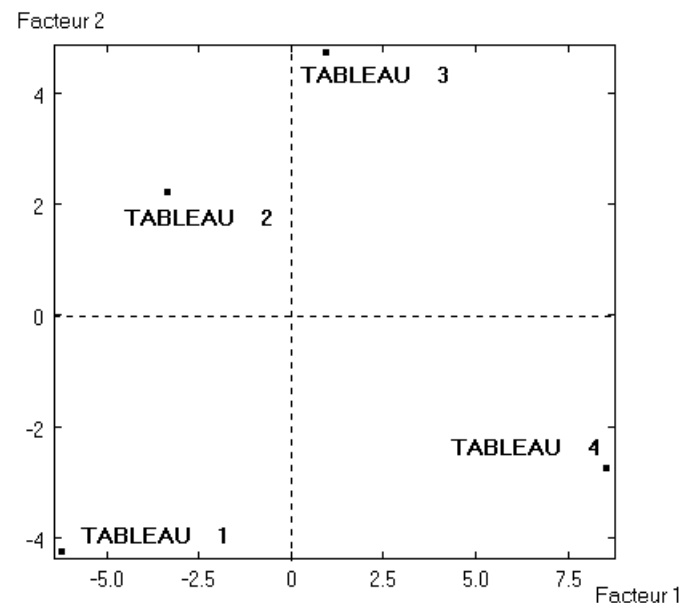
DECOMPOSITION DE L'INERTIE SELON LES AXES PRINCIPAUX
VALEURS PROPRES
TRACE DE LA MATRICE : 203.4173
HISTOGRAMME DES 3 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	124.3122	61.11	61.11	*****
2	52.7304	25.92	87.03	*****
3	26.3746	12.97	100.00	*****

EDITION SOMMAIRE DES VALEURS PROPRES SUIVANTES
4 = 0.0002

COORDONNEES DES TABLEAUX SUR LES AXES PRINCIPAUX

tab.	DISTO	COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
		1	2	3	0	0	1	2	3	0	0	1	2	3	0	0
1	59.22	-6.22	-4.24	1.61	0.00	0.00	0.31	0.34	0.10	0.00	0.00	0.65	0.30	0.04	0.00	0.00
2	31.01	-3.32	2.24	-3.86	0.00	0.00	0.09	0.10	0.57	0.00	0.00	0.36	0.16	0.48	0.00	0.00
3	31.66	0.96	4.72	2.90	0.00	0.00	0.01	0.42	0.32	0.00	0.00	0.03	0.70	0.27	0.00	0.00
4	81.53	8.58	-2.73	-0.65	0.00	0.00	0.59	0.14	0.02	0.00	0.00	0.90	0.09	0.01	0.00	0.00



Puisque le nuage ne contient que quatre points, et qu'il est observé depuis son centre de gravité, il est logique d'obtenir exactement trois axes principaux d'inertie. L'utilisateur a demandé que les coordonnées factorielles, contributions, et cosinus carrés soient édités sur trois axes (NEDIT = 3).

Les cartes factorielles de l'interstructure peuvent être obtenues par une procédure graphique à partir du fichier NGUS1 sur lequel on a stocké, à la demande de l'utilisateur, NAXTA coordonnées.

1.5 Autre exemple

Des juges décrivent un produit à l'aide de descripteurs sensoriels et le produit leur est présenté plusieurs fois. On organise les données en tableaux croisant les juges (lignes) et les répétitions (colonnes), chaque descripteur correspondant à un tableau. On centre en lignes puis en colonnes ces différents tableaux. Si les expertises étaient exactement les mêmes d'une répétition à l'autre, tous ces tableaux bicentrés seraient nuls et les distances entre tableaux, nulles.

Les distances entre tableaux permettent de déceler les descripteurs difficiles à analyser. La décomposition de ces distances en contributions des lignes, fait ressortir les juges qui sont inconstants dans leur appréciation. Sorties de l'option n°2

2. Sorties de l'option n°2 : Comparaison des nuages de lignes

2.1 Résultats sur la forme des différents nuages de lignes

Pour chacun des tableaux X_k , on donne, pour information, une mesure de la dilatation initiale du nuage des lignes (c'est la norme de la matrice $W_k = X_k^t X_k$), et la répartition de l'inertie du nuage normé par rapport au centre de gravité, selon les NAXLI premiers axes principaux (NAXLI = 10).

Ce sont les pourcentages d'inertie que fournirait une analyse en composantes principales non normée (diagonalisation de la matrice des covariances) du tableau, avec une pondération uniforme des lignes. Les valeurs propres sont celles de la matrice W_k normée (alors que les valeurs propres de l'analyse en composantes principales sont celles de W_k/n).

DILATATION DU NUAGE 1
 NORME DE W 1 = 145.22
 REPARTITION DE L'INERTIE SELON LES AXES PRINCIPAUX DU NUAGE 1
 VALEURS PROPRES
 TRACE DE LA MATRICE : 1.8592
 HISTOGRAMME DES 10 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.8919	47.97	47.97	*****
2	0.3381	18.18	66.15	*****
3	0.2110	11.35	77.50	*****
4	0.1644	8.84	86.34	*****
5	0.0907	4.88	91.22	*****
6	0.0838	4.51	95.73	*****
7	0.0531	2.85	98.58	*****
8	0.0264	1.42	100.00	***
9	0.0000	0.00	100.00	*
10	0.0000	0.00	100.00	*

DILATATION DU NUAGE 2
 NORME DE W 2 = 136.07
 REPARTITION DE L'INERTIE SELON LES AXES PRINCIPAUX DU NUAGE 2
 VALEURS PROPRES
 TRACE DE LA MATRICE : 1.9842
 HISTOGRAMME DES 10 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.8773	44.21	44.21	*****
2	0.2723	13.72	57.94	*****
3	0.2459	12.39	70.33	*****
4	0.2152	10.85	81.17	*****
5	0.1873	9.44	90.61	*****
6	0.0982	4.95	95.56	*****
7	0.0647	3.26	98.82	*****
8	0.0233	1.18	100.00	***
9	0.0000	0.00	100.00	*
10	0.0000	0.00	100.00	*

DILATATION DU NUAGE 3
 NORME DE W 3 = 135.94
 REPARTITION DE L'INERTIE SELON LES AXES PRINCIPAUX DU NUAGE 3
 VALEURS PROPRES
 TRACE DE LA MATRICE : 1.9861
 HISTOGRAMME DES 10 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.8431	42.45	42.45	*****
2	0.3884	19.55	62.00	*****
3	0.2498	12.58	74.58	*****
4	0.2223	11.19	85.78	*****
5	0.1334	6.72	92.49	*****
6	0.0776	3.91	96.40	*****
7	0.0439	2.21	98.61	*****
8	0.0276	1.39	100.00	***
9	0.0000	0.00	100.00	*
10	0.0000	0.00	100.00	*

DILATATION DU NUAGE 4
 NORME DE W 4 = 124.75
 REPARTITION DE L'INERTIE SELON LES AXES PRINCIPAUX DU NUAGE 4
 VALEURS PROPRES
 TRACE DE LA MATRICE : 2.1643
 HISTOGRAMME DES 10 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.7640	35.30	35.30	*****
2	0.4575	21.14	56.44	*****
3	0.3520	16.26	72.70	*****
4	0.2033	9.39	82.09	*****
5	0.1419	6.55	88.65	*****
6	0.1092	5.04	93.69	*****
7	0.0836	3.86	97.55	*****
8	0.0529	2.45	100.00	*****
9	0.0000	0.00	100.00	*
10	0.0000	0.00	100.00	*

DILATATION DU NUAGE 5
 NORME DE W 5 = 122.84
 REPARTITION DE L'INERTIE SELON LES AXES PRINCIPAUX DU NUAGE 5
 VALEURS PROPRES
 TRACE DE LA MATRICE : 1.9538
 HISTOGRAMME DES 10 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.7811	39.98	39.98	*****
2	0.5010	25.64	65.62	*****
3	0.3070	15.71	81.34	*****
4	0.1687	8.63	89.97	*****
5	0.1119	5.73	95.70	*****
6	0.0496	2.54	98.24	*****
7	0.0345	1.76	100.00	****
8	0.0000	0.00	100.00	*
9	0.0000	0.00	100.00	*
10	0.0000	0.00	100.00	*

Ces résultats permettent de contrôler que les cinq nuages de lignes ont des dimensions comparables. Ils sont dans un espace de dimension 8 (7 pour le dernier). Et ils ont, à peu près, la même forme (le quatrième nuage est plus ramassé que les autres avec une inertie de 35.30% sur le premier axe).

2.2 Situation des nuages les uns par rapport aux autres

PRODUITS SCALAIRES RV ENTRE NUAGES					
	1	2	3	4	5
1	1.000				
2	0.871	1.000			
3	0.722	0.868	1.000		
4	0.600	0.676	0.735	1.000	
5	0.475	0.530	0.602	0.705	1.000

TESTS DE PERMUTATIONS :

* INDIQUE QUE LES NUAGES SONT PEU DIFFERENTS

	1	2	3	4	5
1	*				
2	*	*			
3	*	*	*		
4	*	*	*	*	
5	*	*	*	*	*

DISTANCES EUCLIDIENNES ENTRE NUAGES

	1	2	3	4	5
1	0.000				
2	0.508	0.000			
3	0.745	0.515	0.000		
4	0.894	0.805	0.728	0.000	
5	1.025	0.969	0.892	0.768	0.000

L'information que nous avait fournie l'option n°1 est encore vraie, à savoir que l'écart entre les nuages se creuse avec le temps, avec une rupture entre le nuage n°3 et le nuage n°4, rupture confirmée par la position du nuage n°5 qui semble à l'écart des quatre autres. Les produits scalaires sont toutefois déclarés significativement élevés par le test de permutations. Cela signifie que dans l'ensemble, les cantons restent positionnés les uns par rapport aux autres de la même manière d'un recensement à l'autre.

L'utilisateur peut demander la décomposition de la distance au carré entre un nuage particulier et les autres nuages en précisant le numéro LNUMT de ce tableau. Dans notre exemple, on a demandé la décomposition de la distance au carré entre chacun des nuages et les quatre autres (LNUMT = TOT). L'information obtenue est redondante car on obtient deux fois chaque décomposition : entre le nuage n°1 et le nuage n°2, puis entre le nuage n°2 et le nuage n°1, etc.

DECOMPOSITION DE LA DISTANCE AU CARRE ENTRE LE NUAGE 1 ET LES AUTRES
CONTRIBUTIONS DES LIGNES EN POURCENTAGE

LIGNE	MOYENNE	NUAGE			
		2	3	4	5
AGDE	1.3	1.8	1.1	1.1	1.1
BEDARIEUX	2.7	4.0	2.4	2.1	2.1
CAPESTANG	0.6	0.8	0.4	0.4	0.8
FLORENSAC	1.7	1.0	1.5	2.3	2.0
MONTAGNAC	1.1	0.7	0.4	1.8	1.5
MURVIEL	1.6	1.7	1.1	1.3	2.3
OLARGUES	2.6	3.9	2.1	2.9	1.4
OLONZAC	2.6	3.4	2.3	1.9	3.0
PEZENAS	1.1	2.1	0.8	1.1	0.6
ROUJAN	1.6	1.0	2.3	1.3	1.8
ST-CHINIAN	1.1	1.1	0.6	0.5	2.3
ST-GERVAIS	9.6	11.9	12.5	7.7	6.3
ST-PONS	7.6	4.5	7.2	8.9	9.9
SALVETAT	4.0	6.0	3.0	2.5	4.5
SERVIAN	1.4	0.9	1.0	1.7	2.0
LE-CAYLAR	3.9	5.4	3.0	3.7	3.6
CLERMONT	0.8	0.9	0.4	0.9	0.9
GIGNAC	1.7	2.0	1.2	2.1	1.4
LODEVE	3.7	2.9	2.9	3.7	5.3
LUNAS	1.5	1.1	1.4	1.8	1.7
ANTIANE	3.0	4.8	2.4	1.9	2.9
CASTRIES	5.4	3.4	4.7	7.0	6.6
CLARET	3.4	4.1	2.0	3.4	4.2
FRONTIGNAN	3.9	4.4	3.4	3.7	4.0
GANGES	5.7	4.8	6.5	6.6	4.8
LUNEL	1.5	1.0	2.3	1.7	1.1
MATELLES	16.8	13.3	26.0	14.1	13.9
MAUGUIO	2.8	2.0	1.0	4.4	3.9
MEZE	2.3	2.2	1.5	3.6	1.7
ST-MARTIN	3.1	3.0	2.7	4.1	2.5

DECOMPOSITION DE LA DISTANCE AU CARRE ENTRE LE NUAGE 2 ET LES AUTRES
CONTRIBUTIONS DES LIGNES EN POURCENTIAGE

LIGNE	MOYENNE	NUAGE			
		1	3	4	5
AGDE	1.7	1.8	2.1	1.6	1.2
BEDARIEUX	2.2	4.0	1.4	1.6	1.8
CAPESTANG	0.7	0.8	0.9	0.5	0.8
FLORENSAC	1.7	1.0	2.3	2.0	1.5
MONTAGNAC	1.1	0.7	0.5	1.9	1.4
MURVIEL	1.7	1.7	1.0	1.5	2.5
OLARGUES	2.4	3.9	2.6	1.9	1.1
OLONZAC	1.8	3.4	0.7	0.9	2.2
PEZENAS	1.1	2.1	0.3	1.6	0.3
ROUJAN	1.9	1.0	3.6	1.4	1.5
ST-CHINIAN	1.2	1.1	0.8	0.6	2.3
ST-GERVAIS	11.6	11.9	13.6	11.3	9.8
ST-PONS	9.4	4.5	12.0	10.4	10.6
SALVETAT	3.0	6.0	1.5	1.7	2.9
SERVIAN	1.5	0.9	1.1	1.9	2.1
LE-CAYLAR	4.4	5.4	4.1	4.5	3.5
CLERMONT	1.0	0.9	0.9	0.8	1.3
GIGNAC	1.4	2.0	0.9	1.1	1.4
LODEVE	2.8	2.9	1.7	2.8	3.7
LUNAS	2.3	1.1	2.3	3.2	2.5
ANIANE	2.7	4.8	1.4	2.6	2.2
CASTRIS	4.4	3.4	3.5	4.9	5.8
CLARET	4.5	4.1	2.2	4.7	6.9
FRONTIGNAN	3.7	4.4	3.5	3.4	3.5
GANGES	5.2	4.8	5.9	5.7	4.3
LUNEL	1.6	1.0	2.4	2.1	0.8
MATELLES	13.3	13.3	20.8	8.1	11.0
MAUGUIO	4.4	2.0	2.0	7.7	6.0
MEZE	2.2	2.2	1.8	2.7	2.1
ST-MARTIN	3.3	3.0	2.3	4.8	2.9

DECOMPOSITION DE LA DISTANCE AU CARRE ENTRE LE NUAGE 3 ET LES AUTRES
CONTRIBUTIONS DES LIGNES EN POURCENTAGE

LIGNE	MOYENNE	NUAGE			
		1	2	4	5
AGDE	1.3	1.1	2.1	1.0	1.2
BEDARIEUX	2.3	2.4	1.4	3.0	2.5
CAPESTANG	0.7	0.4	0.9	0.5	0.9
FLORENSAC	2.8	1.5	2.3	3.9	3.6
MONTAGNAC	1.0	0.4	0.5	1.9	1.1
MURVIEL	1.4	1.1	1.0	1.5	2.0
OLARGUES	2.1	2.1	2.6	2.3	1.5
OLONZAC	1.5	2.3	0.7	0.9	2.1
PEZENAS	0.8	0.8	0.3	1.8	0.5
ROUJAN	2.2	2.3	3.6	1.3	1.4
ST-CHINIAN	1.1	0.6	0.8	0.8	2.4
ST-GERVAIS	11.7	12.5	13.6	10.6	10.3
ST-PONS	6.9	7.2	12.0	3.4	5.1
SALVETAT	2.3	3.0	1.5	1.8	2.8
SERVIAN	1.4	1.0	1.1	1.7	1.7
LE-CAYLAR	4.0	3.0	4.1	5.9	3.1
CLERMONT	0.9	0.4	0.9	0.7	1.5
GIGNAC	1.0	1.2	0.9	1.1	0.9
LODEVE	2.4	2.9	1.7	2.1	2.9
LUNAS	2.4	1.4	2.3	3.7	2.3
ANTIANE	2.5	2.4	1.4	3.4	2.5
CASTRIES	4.6	4.7	3.5	4.3	5.9
CLARET	4.2	2.0	2.2	5.5	7.2
FRONTIGNAN	3.5	3.4	3.5	3.7	3.6
GANGES	5.1	6.5	5.9	3.5	4.5
LUNEL	2.2	2.3	2.4	3.4	0.8
MATELLES	17.5	26.0	20.8	10.0	13.1
MAUGUIO	4.7	1.0	2.0	9.0	7.1
MEZE	2.8	1.5	1.8	4.6	3.3
ST-MARTIN	2.6	2.7	2.3	2.9	2.6

DECOMPOSITION DE LA DISTANCE AU CARRE ENTRE LE NUAGE 4 ET LES AUTRES
CONTRIBUTIONS DES LIGNES EN POURCENTAGE

LIGNE	MOYENNE	NUAGE			
		1	2	3	5
AGDE	1.3	1.1	1.6	1.0	1.7
BEDARIEUX	2.5	2.1	1.6	3.0	3.4
CAPESTANG	0.7	0.4	0.5	0.5	1.2
FLORENSAC	2.5	2.3	2.0	3.9	1.8
MONTAGNAC	1.8	1.8	1.9	1.9	1.4
MURVIEL	2.1	1.3	1.5	1.5	3.9
OLARGUES	2.4	2.9	1.9	2.3	2.6
OLONZAC	1.4	1.9	0.9	0.9	1.9
PEZENAS	1.6	1.1	1.6	1.8	1.8
ROUJAN	1.5	1.3	1.4	1.3	1.9
ST-CHINIAN	1.3	0.5	0.6	0.8	3.2
ST-GERVAIS	8.3	7.7	11.3	10.6	3.6
ST-PONS	6.1	8.9	10.4	3.4	1.6
SALVETAT	2.9	2.5	1.7	1.8	5.6
SERVIAN	2.0	1.7	1.9	1.7	2.5
LE-CAYLAR	5.6	3.7	4.5	5.9	8.2
CLERMONT	1.3	0.9	0.8	0.7	2.9
GIGNAC	1.5	2.1	1.1	1.1	1.6
LODEVE	3.1	3.7	2.8	2.1	3.9
LUNAS	3.0	1.8	3.2	3.7	3.2
ANIANE	2.4	1.9	2.6	3.4	1.8
CASTRIS	5.0	7.0	4.9	4.3	3.9
CLARET	5.2	3.4	4.7	5.5	7.2
FRONTIGNAN	3.3	3.7	3.4	3.7	2.5
GANGES	5.0	6.6	5.7	3.5	4.4
LUNEL	2.4	1.7	2.1	3.4	2.5
MATELLES	10.1	14.1	8.1	10.0	8.3
MAUGUIO	6.0	4.4	7.7	9.0	2.8
MEZE	4.2	3.6	2.7	4.6	6.0
ST-MARTIN	3.6	4.1	4.8	2.9	2.6

DECOMPOSITION DE LA DISTANCE AU CARRE ENTRE LE NUAGE 5 ET LES AUTRES
CONTRIBUTIONS DES LIGNES EN POURCENTAGE

LIGNE	MOYENNE	NUAGE			
		1	2	3	4
AGDE	1.3	1.1	1.2	1.2	1.7
BEDARIEUX	2.4	2.1	1.8	2.5	3.4
CAPESTANG	0.9	0.8	0.8	0.9	1.2
FLORENSAC	2.2	2.0	1.5	3.6	1.8
MONTAGNAC	1.4	1.5	1.4	1.1	1.4
MURVIEL	2.7	2.3	2.5	2.0	3.9
OLARGUES	1.6	1.4	1.1	1.5	2.6
OLONZAC	2.3	3.0	2.2	2.1	1.9
PEZENAS	0.8	0.6	0.3	0.5	1.8
ROUJAN	1.7	1.8	1.5	1.4	1.9
ST-CHINIAN	2.5	2.3	2.3	2.4	3.2
ST-GERVAIS	7.5	6.3	9.8	10.3	3.6
ST-PONS	6.8	9.9	10.6	5.1	1.6
SALVETAT	4.0	4.5	2.9	2.8	5.6
SERVIAN	2.1	2.0	2.1	1.7	2.5
LE-CAYLAR	4.6	3.6	3.5	3.1	8.2
CLERMONT	1.6	0.9	1.3	1.5	2.9
GIGNAC	1.3	1.4	1.4	0.9	1.6
LODEVE	3.9	5.3	3.7	2.9	3.9
LUNAS	2.4	1.7	2.5	2.3	3.2
ANIANE	2.3	2.9	2.2	2.5	1.8
CASTRIS	5.6	6.6	5.8	5.9	3.9
CLARET	6.4	4.2	6.9	7.2	7.2
FRONTIGNAN	3.4	4.0	3.5	3.6	2.5
GANGES	4.5	4.8	4.3	4.5	4.4
LUNEL	1.3	1.1	0.8	0.8	2.5
MATELLES	11.6	13.9	11.0	13.1	8.3
MAUGUIO	5.0	3.9	6.0	7.1	2.8
MEZE	3.3	1.7	2.1	3.3	6.0
ST-MARTIN	2.7	2.5	2.9	2.6	2.6

Les cinq nuages de lignes, bien que significativement proches, ne sont pas superposables (les distances entre nuages ne sont pas nulles) . Cela peut provenir de deux causes : soit les nuages n'ont pas la même forme, soit ils n'ont pas la même structure. Nous avons remarqué que les pourcentages d'inertie des axes des différents nuages n'étaient pas très différents. On en déduit que les cinq nuages ont, grosso modo, la même allure et que c'est par leur structure interne qu'ils diffèrent.

L'examen des contributions des lignes, au carré de la distance entre deux nuages, met en évidence que les cantons "les Matelles", "St Gervais" et, dans une moindre mesure, "St Pons" évoluent différemment des autres cantons puisque leur position à l'intérieur du nuage varie d'un recensement à l'autre.

LA SOMME DES DISTANCES AU CARRE SE DECOMPOSE SELON LES LIGNES		
AGDE	=	1.28 %
BEDARIEUX	=	2.32 %
CAPESTANG	=	0.71 %
FLORENSAC	=	2.24 %
MONTAGNAC	=	1.37 %
MURVIEL	=	1.98 %
OLARGUES	=	1.98 %
OLONZAC	=	1.98 %
PEZENAS	=	0.98 %
ROUJAN	=	1.66 %
ST-CHINIAN	=	1.61 %
ST-GERVAIS	=	9.16 %
ST-PONS	=	7.67 %
SALVETIAT	=	3.22 %
SERVIAN	=	1.79 %
LE-CAYLAR	=	4.28 %
CLERMONT	=	1.14 %
GIGNAC	=	1.38 %
LODEVE	=	3.46 %
LUNAS	=	2.32 %
ANLANE	=	2.50 %
CASTRIES	=	5.43 %
CLARET	=	5.02 %
FRONTIGNAN	=	3.56 %
GANGES	=	5.03 %
LUNEL	=	1.67 %
MATELLES	=	13.27 %
MAUGUIO	=	4.96 %
MEZE	=	2.94 %
ST-MARTIN	=	3.08 %

Ce tableau résume les décompositions détaillées par couple de nuages puisqu'il donne la décomposition de la somme des distances au carré entre nuages, selon les lignes.

2.3 Etude du "nuage des nuages" en prenant comme origine le nuage compromis

Pour obtenir une représentation graphique de la configuration de l'interstructure dans laquelle chaque point représente un nuage de lignes, on prend comme origine le nuage de lignes compromis. Et on fournit les informations nécessaires pour comprendre comment sont situés les nuages les uns par rapport aux autres.

CONSTRUCTION DU NUAGE COMPROMIS

= MOYENNE PONDEREE DES NUAGES PAR LES COEFFICIENTS

1 = 0.230
2 = 0.248
3 = 0.246
4 = 0.230
5 = 0.202

PRODUITS SCALAIRES ET DISTANCES ENTRE LE NUAGE COMPROMIS ET LES DIFFERENTS NUAGES

1 : PRODUIT SCALAIRE = 0.857 , DISTANCE = 0.535
2 : PRODUIT SCALAIRE = 0.923 , DISTANCE = 0.392
3 : PRODUIT SCALAIRE = 0.917 , DISTANCE = 0.407
4 : PRODUIT SCALAIRE = 0.858 , DISTANCE = 0.533
5 : PRODUIT SCALAIRE = 0.752 , DISTANCE = 0.704

La somme des distances au carré entre le nuage compromis et chaque nuage, est égale à l'inertie 1,3848 du "nuage des nuages" par rapport au nuage compromis. Ces distances au carré sont exprimées en pourcentages de l'inertie dans le tableau suivant. Par exemple, le nuage n°1 est à la distance $0.535 = \sqrt{0.29}$ du nuage compromis, et sa contribution à l'inertie est égale à $0.29/1.3848$, soit 20.68%.

DECOMPOSITION DE L'INERTIE SELON LES NUAGES

NUAGE 1 = 20.68 %
NUAGE 2 = 11.08 %
NUAGE 3 = 11.97 %
NUAGE 4 = 20.48 %
NUAGE 5 = 35.80 %

Puisque le nuage contient cinq points, et qu'il est observé depuis un sixième point, il est logique d'obtenir exactement cinq axes principaux d'inertie. L'utilisateur a demandé que les coordonnées factorielles, contributions, et cosinus carrés soient édités sur trois axes (NEDIT = 3).

DECOMPOSITION DE L'INERTIE SELON LES AXES PRINCIPAUX

VALEURS PROPRES

TRACE DE LA MATRICE : 1.3848

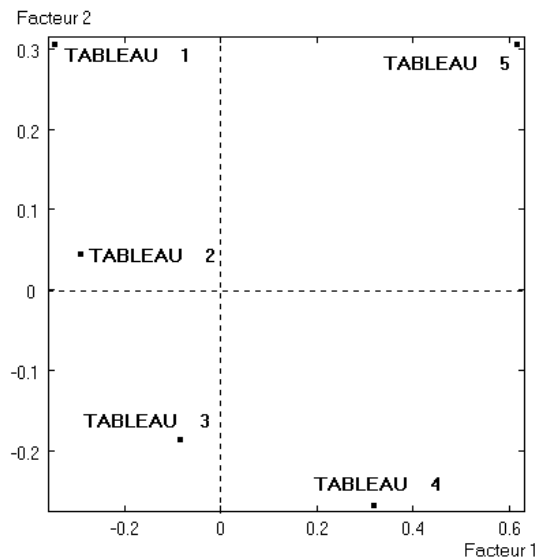
HISTOGRAMME DES 5 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.6964	50.29	50.29	*****
2	0.2974	21.48	71.77	*****
3	0.2296	16.58	88.34	*****
4	0.0876	6.32	94.67	*****
5	0.0738	5.33	100.00	*****

COORDONNEES DES NUAGES SUR LES AXES PRINCIPAUX

NUAGE	DISTO	COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	0.29	-0.35	0.31	-0.21	-0.16	-0.03	0.17	0.32	0.20	0.30	0.02	0.42	0.33	0.16	0.09	0.00
2	0.15	-0.29	0.05	0.06	0.01	0.25	0.12	0.01	0.02	0.00	0.85	0.55	0.01	0.03	0.00	0.41
3	0.17	-0.08	-0.19	0.28	-0.22	-0.02	0.01	0.12	0.34	0.53	0.00	0.04	0.21	0.47	0.28	0.00
4	0.28	0.32	-0.27	-0.30	-0.10	0.08	0.15	0.24	0.40	0.12	0.09	0.36	0.25	0.32	0.04	0.02
5	0.50	0.62	0.31	0.11	-0.06	0.05	0.55	0.32	0.05	0.04	0.04	0.77	0.19	0.02	0.01	0.01

Les cartes factorielles de l'interstructure peuvent être obtenues par une procédure graphique à partir du fichier NGUS1 sur lequel on a stocké, à la demande de l'utilisateur, NAXTA coordonnées.



2.4 Etude du nuage compromis en prenant comme origine le centre de gravité

La matrice compromis est encore une matrice de produits scalaires centrés entre les lignes, de type W_G , à partir de laquelle on construit le nuage compromis qui résume "au mieux" les positions mutuelles des cantons.

DECOMPOSITION DE L'INERTIE SELON LES LIGNES

AGDE	=	2.07 %
BEDARIEUX	=	4.30 %
CAPESTANG	=	1.15 %
FLORENSAC	=	1.31 %
MONTAGNAC	=	1.45 %
MURVIEL	=	1.58 %
OLARGUES	=	3.96 %
OLONZAC	=	2.34 %
PEZENAS	=	1.22 %
ROUJAN	=	1.54 %
ST-CHINIAN	=	1.44 %
ST-GERVAIS	=	10.34 %
ST-PONS	=	4.71 %
SALVETAT	=	3.23 %
SERVIAN	=	2.35 %
LE-CAYLAR	=	4.69 %
CLERMONT	=	0.85 %
GIGNAC	=	1.08 %
LODEVE	=	3.95 %
LUNAS	=	2.34 %
ANIANE	=	2.66 %
CASTRIES	=	3.45 %
CLARET	=	4.14 %
FRONTIGNAN	=	5.72 %
GANGES	=	7.42 %
LUNEL	=	1.76 %
MATELLES	=	9.52 %
MAUGUIO	=	2.89 %
MEZE	=	4.69 %
ST-MARTIN	=	1.86 %

REPRESENTATION GRAPHIQUE DU NUAGE COMPROMIS
 DECOMPOSITION DE L'INERTIE SELON LES AXES PRINCIPAUX
 VALEURS PROPRES
 TRACE DE LA MATRICE : 2.2985
 HISTOGRAMME DES 29 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.8354	36.35	36.35	*****
2	0.3654	15.90	52.24	*****
3	0.2897	12.61	64.85	*****
4	0.2048	8.91	73.76	*****
5	0.1417	6.16	79.92	*****
6	0.0903	3.93	83.85	*****
7	0.0669	2.91	86.76	*****
8	0.0577	2.51	89.27	*****
9	0.0452	1.96	91.23	*****
10	0.0447	1.95	93.18	*****

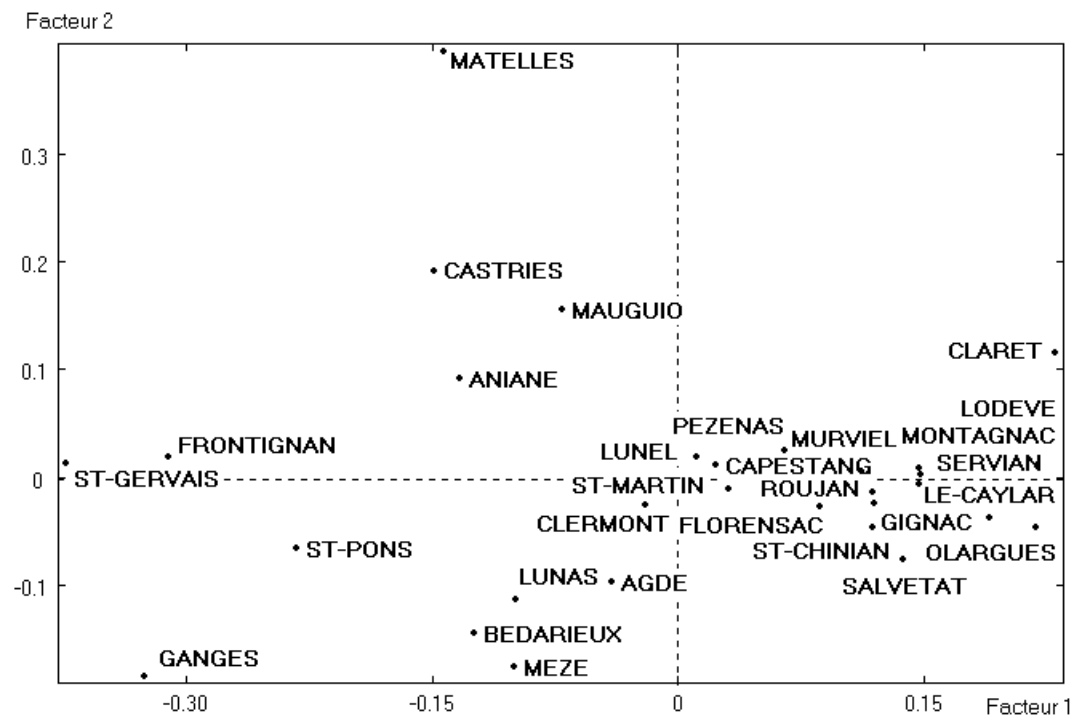
Comme dans toute représentation graphique, on fournit les informations nécessaires pour comprendre comment sont situées les lignes compromis les unes par rapport aux autres. Par exemple, ST-GERVAIS est à la distance $\sqrt{0.24}$ du centre de gravité du nuage compromis, et sa contribution à l'inertie est égale à $0.24/2.2985$, soit 10.34%.

L'utilisateur a demandé que soient retenus NAXLI axes principaux et que les coordonnées factorielles, contributions, et cosinus carrés soient édités sur deux axes (NAXLI = 10, NEDLI = 2).

COORDONNEES DES LIGNES SUR LES AXES PRINCIPAUX

		COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
LIGNE	DISTO	1	2	0	0	0	1	2	0	0	0	1	2	0	0	0
AGDE	0.05	-0.04	-0.10	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.03	0.19	0.00	0.00	0.00
BEDARIEUX	0.10	-0.12	-0.14	0.00	0.00	0.00	0.02	0.06	0.00	0.00	0.00	0.16	0.21	0.00	0.00	0.00
CAPESTANG	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.00
FLORENSAC	0.03	0.09	-0.03	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.25	0.02	0.00	0.00	0.00
MONTAGNAC	0.03	0.15	0.01	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.65	0.00	0.00	0.00	0.00
MURVIEL	0.04	0.11	0.01	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.35	0.00	0.00	0.00	0.00
OLARGUES	0.09	0.22	-0.05	0.00	0.00	0.00	0.06	0.01	0.00	0.00	0.00	0.52	0.02	0.00	0.00	0.00
OLONZAC	0.05	0.19	-0.04	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.67	0.02	0.00	0.00	0.00
PEZENAS	0.03	0.06	0.03	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.15	0.02	0.00	0.00	0.00
ROUJAN	0.04	0.12	-0.01	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.40	0.00	0.00	0.00	0.00
ST-CHINIAN	0.03	0.12	-0.05	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.42	0.06	0.00	0.00	0.00
ST-GERVAIS	0.24	-0.37	0.01	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.59	0.00	0.00	0.00	0.00
ST-PONS	0.11	-0.23	-0.06	0.00	0.00	0.00	0.06	0.01	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00
SALVETAT	0.07	0.14	-0.08	0.00	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.26	0.08	0.00	0.00	0.00
SERVIAN	0.05	0.15	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.40	0.00	0.00	0.00	0.00
LE-CAYLAR	0.11	0.15	-0.01	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00
CLERMONT	0.02	-0.02	-0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.03	0.00	0.00	0.00
GIGNAC	0.02	0.12	-0.02	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.58	0.02	0.00	0.00	0.00
LODEVE	0.09	0.22	0.01	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.53	0.00	0.00	0.00	0.00
LUNAS	0.05	-0.10	-0.11	0.00	0.00	0.00	0.01	0.03	0.00	0.00	0.00	0.18	0.23	0.00	0.00	0.00
ANIANE	0.06	-0.13	0.09	0.00	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.29	0.14	0.00	0.00	0.00
CASTRIES	0.08	-0.15	0.19	0.00	0.00	0.00	0.03	0.10	0.00	0.00	0.00	0.28	0.46	0.00	0.00	0.00
CLARET	0.10	0.23	0.12	0.00	0.00	0.00	0.06	0.04	0.00	0.00	0.00	0.56	0.14	0.00	0.00	0.00
FRONTIGNAN	0.13	-0.31	0.02	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.73	0.00	0.00	0.00	0.00
GANGES	0.17	-0.33	-0.18	0.00	0.00	0.00	0.13	0.09	0.00	0.00	0.00	0.62	0.20	0.00	0.00	0.00
LUNEL	0.04	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
MATELLES	0.22	-0.14	0.40	0.00	0.00	0.00	0.02	0.43	0.00	0.00	0.00	0.09	0.71	0.00	0.00	0.00
MAUGUIO	0.07	-0.07	0.16	0.00	0.00	0.00	0.01	0.07	0.00	0.00	0.00	0.08	0.37	0.00	0.00	0.00
MEZE	0.11	-0.10	-0.17	0.00	0.00	0.00	0.01	0.08	0.00	0.00	0.00	0.09	0.28	0.00	0.00	0.00
ST-MARTIN	0.04	0.03	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00

Les cartes factorielles du nuage compromis peuvent être obtenues par une procédure graphique à partir du fichier NGUS sur lequel on a stocké, à la demande de l'utilisateur, NAXLI coordonnées.



2.5 Trajectoires des individus

Dans les anciennes versions du programme, on proposait le dessin de trajectoires, obtenu en projetant les différentes positions d'un individu sur le plan principal 1-2 du nuage compromis. Cette technique a été abandonnée car le plan principal du compromis n'est pas un plan optimal pour ces trajectoires.

On s'est aperçu que la lecture de ce graphique conduisait à des interprétations fantaisistes, particulièrement dans le cas de données du type produits (lignes), descripteurs (colonnes), juges (tableaux) puisque, dans ce cas, le nuage compromis peut être assez éloigné du nuage des produits donné par un juge. En revanche, lorsque le plan principal du nuage compromis est proche des plans principaux des différents nuages (pour des données décrivant un phénomène qui évolue lentement dans le temps, par exemple), les trajectoires étaient relativement fiables.

Dans la procédure STTIS, l'information donnée par les trajectoires, est remplacée par la décomposition des distances entre nuages en contributions des individus. Cette décomposition (algébrique, et par conséquent exacte) permet de repérer les individus qui ont une trajectoire anormale.

Il reste à savoir dans quel sens évolue cette trajectoire. Une analyse en composantes principales du tableau formé par les différents tableaux juxtaposés peut très bien faire l'affaire puisque le plan principal de cette analyse est alors optimal pour représenter l'ensemble des points de la trajectoire.

2.6 Autre exemple

On dispose de trois tableaux de données portant sur un ensemble de sites. Les trois tableaux ont été fournis par des chercheurs appartenant à des domaines scientifiques différents. Le premier tableau est fourni par un pédologue et décrit les caractéristiques physico-chimiques du sol pour chaque site. Le second tableau, fourni par un botaniste, donne la présence ou l'absence de diverses espèces végétales sur chaque site. Le troisième tableau, fourni par un biologiste, donne l'abondance de certaines espèces animales sur chacun des sites. Dans quelle mesure la classification de ces sites dépend du spécialiste envoyé sur le terrain ?

3. Sorties de l'option n°3 : Comparaison des matrices de corrélations

3.1 Résultats sur la forme des différents nuages de variables

Pour chacun des tableaux on édite la matrice de corrélations et la répartition de l'inertie du nuage des variables, selon les axes principaux (NAXCO = TOT). Ce sont les valeurs propres et les pourcentages d'inertie que fournirait une analyse en composantes principales normée (diagonalisation de la matrice des corrélations) du tableau.

NUAGE 1

MATRICE DE CORRELATIONS 1

	1	2	3	4	5	6	7	8	9
1	1.000								
2	-0.094	1.000							
3	-0.441	-0.161	1.000						
4	-0.497	-0.328	0.570	1.000					
5	-0.400	-0.527	0.335	0.716	1.000				
6	-0.702	-0.233	0.412	0.489	0.628	1.000			
7	-0.645	-0.656	0.241	0.471	0.580	0.597	1.000		
8	-0.568	-0.125	0.417	0.607	0.557	0.674	0.378	1.000	
9	0.302	-0.270	-0.172	0.061	0.180	-0.193	-0.054	0.136	1.000

VALEURS PROPRES

TRACE DE LA MATRICE : 9.0000

HISTOGRAMME DES 9 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	4.3173	47.97	47.97	*****
2	1.6366	18.18	66.15	*****
3	1.0214	11.35	77.50	*****
4	0.7956	8.84	86.34	*****
5	0.4388	4.88	91.22	*****
6	0.4056	4.51	95.73	*****
7	0.2569	2.85	98.58	*****
8	0.1277	1.42	100.00	***
9	0.0000	0.00	100.00	*

NUAGE 2

MATRICE DE CORRELATIONS 2

	1	2	3	4	5	6	7	8	9
1	1.000								
2	0.131	1.000							
3	-0.355	-0.337	1.000						
4	-0.629	-0.432	0.467	1.000					
5	-0.148	-0.514	0.158	0.698	1.000				
6	-0.667	-0.183	0.101	0.524	0.292	1.000			
7	-0.742	-0.682	0.247	0.534	0.278	0.482	1.000		
8	-0.428	-0.260	0.280	0.605	0.476	0.549	0.254	1.000	
9	-0.080	-0.154	-0.034	0.167	0.059	0.131	0.071	0.201	1.000

VALEURS PROPRES

TRACE DE LA MATRICE : 9.0000

HISTOGRAMME DES 9 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	3.9792	44.21	44.21	*****
2	1.2351	13.72	57.94	*****
3	1.1151	12.39	70.33	*****
4	0.9762	10.85	81.17	*****
5	0.8496	9.44	90.61	*****
6	0.4454	4.95	95.56	*****
7	0.2935	3.26	98.82	*****
8	0.1059	1.18	100.00	***
9	0.0000	0.00	100.00	*

NUAGE 3

MATRICE DE CORRELATIONS 3

	1	2	3	4	5	6	7	8	9
1	1.000								
2	0.158	1.000							
3	-0.292	-0.166	1.000						
4	-0.377	-0.197	-0.073	1.000					
5	-0.390	-0.389	0.146	0.735	1.000				
6	-0.750	-0.322	-0.036	0.462	0.457	1.000			
7	-0.809	-0.588	0.109	0.184	0.248	0.699	1.000		
8	-0.422	-0.334	0.207	0.582	0.665	0.389	0.258	1.000	
9	0.050	-0.262	-0.195	0.559	0.326	-0.053	0.008	0.170	1.000

VALEURS PROPRES

TRACE DE LA MATRICE : 9.0000

HISTOGRAMME DES 9 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	3.8204	42.45	42.45	*****
2	1.7599	19.55	62.00	*****
3	1.1320	12.58	74.58	*****
4	1.0074	11.19	85.78	*****
5	0.6044	6.72	92.49	*****
6	0.3518	3.91	96.40	*****
7	0.1988	2.21	98.61	*****
8	0.1252	1.39	100.00	***
9	0.0000	0.00	100.00	*

NUAGE 4

MATRICE DE CORRELATIONS 4

	1	2	3	4	5	6	7	8	9
1	1.000								
2	0.264	1.000							
3	-0.254	-0.416	1.000						
4	-0.312	-0.108	-0.171	1.000					
5	-0.521	-0.511	0.056	0.630	1.000				
6	-0.600	-0.186	-0.182	0.449	0.556	1.000			
7	-0.703	-0.473	0.240	-0.213	0.120	0.198	1.000		
8	-0.453	-0.140	0.081	0.204	0.361	0.323	0.104	1.000	
9	-0.013	-0.413	0.416	0.088	0.243	-0.057	-0.052	-0.147	1.000

VALEURS PROPRES

TRACE DE LA MATRICE : 9.0000

HISTOGRAMME DES 9 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	3.1771	35.30	35.30	*****
2	1.9023	21.14	56.44	*****
3	1.4637	16.26	72.70	*****
4	0.8453	9.39	82.09	*****
5	0.5899	6.55	88.65	*****
6	0.4539	5.04	93.69	*****
7	0.3478	3.86	97.55	*****
8	0.2201	2.45	100.00	*****
9	0.0000	0.00	100.00	*

3.2 Situation des matrices de corrélations les unes par rapport aux autres

PRODUITS SCALAIRES ENTRE MATRICES DE CORRELATIONS

	1	2	3	4
1	23.433			
2	21.023	20.573		
3	20.329	19.362	20.534	
4	17.062	16.476	17.175	17.293

DISTANCES EUCLIDIENNES ENTRE MATRICES DE CORRELATIONS

	1	2	3	4
1	0.000			
2	1.400	0.000		
3	1.819	1.544	0.000	
4	2.569	2.217	1.864	0.000

INTERPRETATION DES DISTANCES ENTRE MATRICES DE CORRELATIONS

CONTRIBUTIONS DES VARIABLES EN POURCENTAGE

DECOMPOSITION DE LA DISTANCE AU CARRE ENTRE LA MATRICE DE CORRELATIONS 1 ET LES AUTRES

		MATRICE			
	COLONNE	MOYENNE	2	3	4
EAGR		9.2	16.1	6.5	5.1
OAGR		5.2	6.5	4.7	4.5
INDU		18.8	11.0	22.2	23.3
LIBE		15.2	2.8	23.5	19.2
CADR		9.6	16.4	6.9	5.4
EMPL		12.9	17.8	10.7	10.2
OVR		10.1	7.6	8.2	14.4
SERV		6.9	5.0	6.6	9.2
CLER		12.1	16.8	10.8	8.6

DECOMPOSITION DE LA DISTANCE AU CARRE ENTRE LA MATRICE DE CORRELATIONS 2 ET LES AUTRES

		MATRICE			
	COLONNE	MOYENNE	1	3	4
EAGR		9.4	16.1	6.5	5.7
OAGR		5.9	6.5	6.1	5.2
INDU		14.3	11.0	16.5	15.4
LIBE		19.8	2.8	29.1	27.4
CADR		10.4	16.4	8.9	6.0
EMPL		10.7	17.8	7.7	6.6
OVR		10.5	7.6	8.7	15.2
SERV		5.6	5.0	3.1	8.6
CLER		13.3	16.8	13.4	9.9

DECOMPOSITION DE LA DISTANCE AU CARRE ENTRE LA MATRICE DE CORRELATIONS 3 ET LES AUTRES

		MATRICE		
COLONNE	MOYENNE	1	2	4
EAGR	5.0	6.5	6.5	2.1
OAGR	5.4	4.7	6.1	5.4
INDU	17.8	22.2	16.5	14.7
LIBE	22.9	23.5	29.1	16.0
CADR	7.0	6.9	8.9	5.1
EMPL	9.3	10.7	7.7	9.4
OUVR	10.4	8.2	8.7	14.2
SERV	7.2	6.6	3.1	12.0
CLER	15.1	10.8	13.4	21.1

DECOMPOSITION DE LA DISTANCE AU CARRE ENTRE LA MATRICE DE CORRELATIONS 4 ET LES AUTRES

		MATRICE		
COLONNE	MOYENNE	1	2	3
EAGR	4.3	5.1	5.7	2.1
OAGR	5.0	4.5	5.2	5.4
INDU	17.8	23.3	15.4	14.7
LIBE	20.9	19.2	27.4	16.0
CADR	5.5	5.4	6.0	5.1
EMPL	8.8	10.2	6.6	9.4
OUVR	14.6	14.4	15.2	14.2
SERV	9.9	9.2	8.6	12.0
CLER	13.2	8.6	9.9	21.1

On remarque sur le tableau des distances entre les matrices de corrélations que l'écart se creuse avec le temps. L'examen des contributions des variables, au carré de la distance entre deux matrices de corrélations, met en évidence que ces écarts sont dus principalement aux catégories socioprofessionnelles "professions libérales", "industriels et commerçants" et, dans une moindre mesure, "ouvriers" et "clergé, armée, police". Cela signifie que, sur cet ensemble de cantons, les corrélations de ces catégories socioprofessionnelles avec les autres, varient beaucoup d'un recensement à l'autre (elles changent même de signe ...).

On retrouve cette information d'une manière plus synthétique, en analysant la décomposition de la somme des distances au carré, selon les variables.

LA SOMME DES DISTANCES AU CARRE SE DECOMPOSE SELON LES VARIABLES :

EAGR = 6.07 %
 OAGR = 5.17 %
 INDU = 18.32 %
 LIBE = 20.75 %
 CADR = 7.04 %
 EMPL = 9.77 %
 OUVR = 12.44 %
 SERV = 8.11 %
 CLER = 12.33 %

3.3 Etude du "nuage des matrices de corrélations" en prenant comme origine la matrice compromis

Pour obtenir une représentation graphique de la configuration de l'interstructure dans laquelle chaque point représente une matrice de corrélations, on prend comme origine la matrice de corrélations compromis. Et on fournit les informations nécessaires pour comprendre comment sont situés les points de ce nuage les uns par rapport aux autres.

CONSTRUCTION DE LA MATRICE DE CORRELATIONS COMPROMIS

= MOYENNE PONDEREE DES MATRICES DE CORRELATIONS PAR LES COEFFICIENTS

1 = 0.269
 2 = 0.255
 3 = 0.254
 4 = 0.222

MATRICE DE CORRELATIONS COMPROMIS

	1	2	3	4	5	6	7	8	9
1	1.000								
2	0.107	1.000							
3	-0.340	-0.263	1.000						
4	-0.459	-0.272	0.216	1.000					
5	-0.360	-0.485	0.180	0.697	1.000				
6	-0.683	-0.233	0.087	0.482	0.483	1.000			
7	-0.724	-0.605	0.209	0.262	0.316	0.505	1.000		
8	-0.470	-0.216	0.254	0.511	0.520	0.492	0.255	1.000	
9	0.071	-0.270	-0.012	0.221	0.200	-0.045	-0.006	0.098	1.000

PRODUITS SCALAIRES ET DISTANCES ENTRE LA MATRICE COMPROMIS ET LES DIFFERENTES MATRICES DE CORRELATIONS

1 : PRODUIT SCALAIRE = 20.617 , DISTANCE = 1.192
 2 : PRODUIT SCALAIRE = 19.477 , DISTANCE = 0.917
 3 : PRODUIT SCALAIRE = 19.434 , DISTANCE = 0.941
 4 : PRODUIT SCALAIRE = 16.993 , DISTANCE = 1.590

La somme des distances au carré, entre la matrice compromis et chaque matrice de corrélations, est égale à l'inertie 5.6781 du "nuage des matrices de corrélations" par rapport à la matrice compromis. Ces distances au carré sont exprimées en pourcentages de l'inertie dans le tableau suivant.

DECOMPOSITION DE L'INERTIE SELON LES MATRICES DE CORRELATIONS

MATRICE DE CORRELATIONS 1 = 25.04 %
 MATRICE DE CORRELATIONS 2 = 14.81 %
 MATRICE DE CORRELATIONS 3 = 15.61 %
 MATRICE DE CORRELATIONS 4 = 44.53 %

Puisque le nuage contient cinq points, et qu'il est observé depuis un sixième point, il est logique d'obtenir exactement cinq axes principaux d'inertie. L'utilisateur a demandé que les coordonnées factorielles, contributions, et cosinus carrés soient édités sur trois axes (NEDIT = 3).

DECOMPOSITION DE L'INERTIE SELON LES AXES PRINCIPAUX

VALEURS PROPRES

TRACE DE LA MATRICE : 5.6781

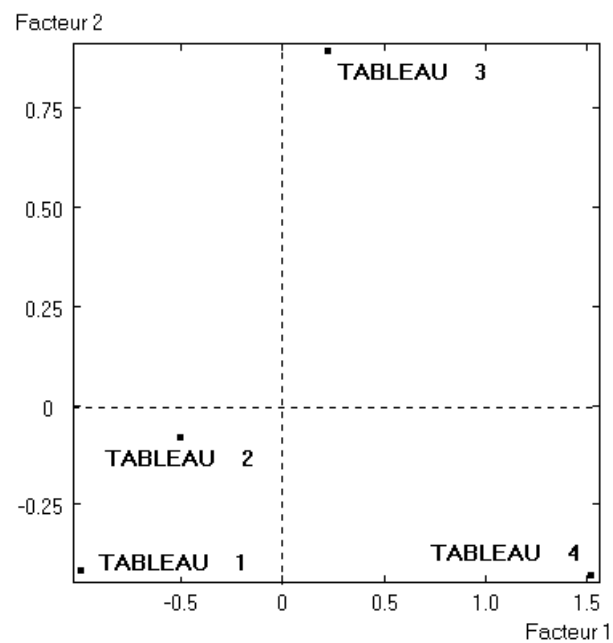
HISTOGRAMME DES 4 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	3.6427	64.15	64.15	*****
2	1.1624	20.47	84.63	*****
3	0.8729	15.37	100.00	*****
4	0.0000	0.00	100.00	*

COORDONNEES DES MATRICES DE CORRELATIONS SUR LES AXES PRINCIPAUX

		COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
matri	DISTO	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0
1	1.42	-1.00	-0.42	0.50	0.00	0.00	0.27	0.15	0.29	0.29	0.00	0.70	0.12	0.18	0.00	0.00
2	0.84	-0.50	-0.08	-0.76	0.00	0.00	0.07	0.01	0.67	0.26	0.00	0.30	0.01	0.69	0.00	0.00
3	0.89	0.22	0.89	0.19	0.00	0.00	0.01	0.69	0.04	0.26	0.00	0.06	0.90	0.04	0.00	0.00
4	2.53	1.53	-0.43	0.05	0.00	0.00	0.64	0.16	0.00	0.20	0.00	0.93	0.07	0.00	0.00	0.00

Les cartes factorielles de l'interstructure peuvent être obtenues par une procédure graphique à partir du fichier NGUS1 sur lequel on a stocké, à la demande de l'utilisateur, NAXTA coordonnées.



3.4 Représentation graphique de la matrice compromis

DECOMPOSITION DE L'INERTIE SELON LES AXES PRINCIPAUX

VALEURS PROPRES

TRACE DE LA MATRICE : 9.0000

HISTOGRAMME DES 9 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	3.7831	42.03	42.03	*****
2	1.3673	15.19	57.23	*****
3	1.1351	12.61	69.84	*****
4	0.9346	10.38	80.22	*****
5	0.6773	7.53	87.75	*****
6	0.5000	5.56	93.30	*****
7	0.3337	3.71	97.01	*****
8	0.2396	2.66	99.67	*****
9	0.0294	0.33	100.00	*

COORDONNEES DES VARIABLES SUR LES AXES PRINCIPAUX

VARIABLE	DISTO	COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
		1	2	0	0	0	1	2	0	0	0	1	2	0	0	0
EAGR	1.00	-0.77	0.49	0.00	0.00	0.00	0.16	0.17	0.00	0.00	0.00	0.60	0.24	0.00	0.00	0.00
OAGR	1.00	-0.56	-0.42	0.00	0.00	0.00	0.08	0.13	0.00	0.00	0.00	0.31	0.18	0.00	0.00	0.00
INDU	1.00	0.39	-0.07	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.16	0.00	0.00	0.00	0.00
LIBE	1.00	0.74	0.24	0.00	0.00	0.00	0.15	0.04	0.00	0.00	0.00	0.55	0.06	0.00	0.00	0.00
CADR	1.00	0.76	0.34	0.00	0.00	0.00	0.15	0.09	0.00	0.00	0.00	0.58	0.12	0.00	0.00	0.00
EMPL	1.00	0.76	-0.30	0.00	0.00	0.00	0.15	0.07	0.00	0.00	0.00	0.58	0.09	0.00	0.00	0.00
OUVR	1.00	0.71	-0.25	0.00	0.00	0.00	0.13	0.05	0.00	0.00	0.00	0.50	0.06	0.00	0.00	0.00
SERV	1.00	0.69	0.02	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.48	0.00	0.00	0.00	0.00
CLER	1.00	0.16	0.78	0.00	0.00	0.00	0.01	0.45	0.00	0.00	0.00	0.02	0.62	0.00	0.00	0.00

A partir de la matrice de corrélations compromis, on construit le nuage des variables dont les projections planes sur les axes principaux donnent les cercles de corrélations compromis.

L'utilisateur a demandé que soient retenus NAXCO axes principaux et que les coordonnées factorielles, contributions, et cosinus carrés soient édités sur deux axes (NAXCO = TOT, NEDCO = 2).

Les cercles de corrélations compromis peuvent être obtenus par une procédure graphique à partir du fichier NGUS sur lequel on a stocké, à la demande de l'utilisateur, NAXCO coordonnées.

