

DE NOUVELLES MACROS SAS D'ANALYSE DES DONNÉES A L'INSEE : COMMENT RÉALISER UNE ANALYSE FACTORIELLE MULTIPLE

Brigitte GELEIN (), Olivier SAUTORY (**)*

() Insee, Direction Régionale de Poitou-Charentes*

*(**) Cepe - Insee*

L'analyse factorielle multiple (AFM) est une méthode d'analyse des données permettant de traiter conjointement plusieurs tableaux de données décrivant un même ensemble d'individus. Chaque tableau se rapporte à un thème particulier (analyse thématique) ou à une date (analyse des données évolutives).

Cette méthode proposée par Brigitte Escofier et Jérôme Pagès dans les années 80, est relativement peu utilisée. Deux nouvelles macros SAS vont être proposées prochainement aux statisticiens de l'Insee pour qu'ils puissent mettre en œuvre cette méthode :

- AFMULT permet de réaliser des AFM sur des variables quantitatives et qualitatives ;
- PLOTAFM permet d'obtenir des graphiques illustrant les résultats de l'analyse

1. Aspects théoriques de l'AFM

1.1 Principe de la méthode

L'analyse factorielle multiple (AFM) est une méthode permettant d'analyser des tableaux dans lesquels des individus sont décrits par plusieurs groupes de variables, quantitatives ou qualitatives. Ces groupes peuvent être constitués :

- de variables identiques mesurées sur les mêmes individus à différentes dates (AFM "temporelle") ;
- de variables provenant d'un seul tableau de données, mais qui sont regroupées par "thème" : chaque groupe correspond alors à un thème, donc à un sous-tableau (AFM "thématique").

La spécificité de l'AFM par rapport aux méthodes classiques (analyse en composantes principales ou analyse des correspondances multiples) est d'équilibrer l'influence des différents groupes, grâce à des pondérations attribuées à ces groupes. Ce point est important dans la mesure où un groupe de variables peut jouer un rôle prépondérant dans une analyse traditionnelle, pour deux raisons :

1. le nombre de variables du groupe : plus ce nombre est élevé, plus l'influence du groupe sur les résultats risque d'être forte ;

2. la structure du groupe : plus ses variables sont liées, plus il va influencer la détermination des premiers axes principaux de l'analyse globale.

L'AFM offre des représentations factorielles, construites à partir d'un système d'axes, où les individus sont décrits selon deux points de vue :

- globalement, en prenant en compte l'ensemble des variables ;
- au travers de chacun des thèmes pris séparément.

Chaque individu est donc représenté, dans un plan factoriel, par $1 + J$ points, où J est le nombre de groupes. On peut donc savoir si un individu se distingue des autres globalement ou bien sur un thème particulier.

De plus, l'étude est enrichie (par rapport à une analyse traditionnelle) par la possibilité de comparer les groupes de variables eux-mêmes.

L'AFM peut traiter en même temps des groupes de variables qualitatives et des groupes de variables quantitatives. Par contre, au sein d'un même groupe, les variables doivent être de même nature.

L'AFM se comporte comme une ACP pour les variables quantitatives et comme une ACM pour les variables qualitatives.

L'AFM équilibre l'influence des groupes de variables en donnant à chaque variable un poids : ces poids sont identiques au sein de chaque groupe, afin de conserver la structure interne de chacun des groupes. Le poids donné par l'AFM à chacune des variables d'un groupe est égal à l'inverse de l'inertie du premier axe principal de ce groupe (obtenue par des ACP ou des ACM réalisées sur les différents groupes de façon séparée). On a donc une normalisation des nuages représentant les individus pour chaque groupe.

Cette pondération rend possible la présence simultanée de variables continues et nominales parmi les éléments actifs. Elle autorise également la présence simultanée de variables continues centrées réduites et de variables continues non réduites.

Les éléments théoriques qui suivent portent, dans un premier temps, sur l'application de la méthode à des variables quantitatives. Sont ensuite traités les variables qualitatives et les éléments supplémentaires.

1.2 Notations

Les données portent sur I individus décrits par K variables formant J groupes. À chaque groupe correspond un tableau X_j , $j = 1 \dots J$. Mais l'ensemble des J tableaux initiaux peut être traité comme un unique tableau structuré en sous-tableaux. En effet, les J tableaux portent sur les mêmes individus et peuvent, par juxtaposition, former un tableau croisant les I individus et les K variables.

On note X le tableau complet, I l'ensemble des individus, K l'ensemble des variables, J l'ensemble des sous-tableaux, K_j l'ensemble des variables du groupe j . Les symboles I, J, K ou K_j désignent donc à la fois l'ensemble et son cardinal.

On considère les trois espaces suivants :

- \mathbb{R}^K dans lequel est situé le nuage des individus
- \mathbb{R}^I dans lequel est situé le nuage des variables

- \mathbb{R}^{I^2} dans lequel est situé le nuage des groupes de variables.

On associe à chaque individu i un poids p_i , avec $\sum_i p_i = 1$, et on note D la matrice diagonale (I, I) contenant ces poids.

On associe à chaque variable \mathbf{n}_k appartenant au groupe j un poids m_k^j , et on note M_j la matrice diagonale des poids des variables du groupe j . Les variables numériques ont souvent un poids égal à 1.

1.3 Étape préliminaire : les ACP séparées

On commence par effectuer J analyses en composantes principales (ACP) sur chacun de tableaux définis par les groupes de variables.

Ces ACP reviennent à déterminer les valeurs et vecteurs propres des matrices $X_j M_j X_j' D = W_j D$.

On note I_j^j la première valeur propre de l'ACP du groupe j . L'inverse de cette valeur propre est le poids donné au groupe j dans l'analyse : la variable \mathbf{n}_k appartenant au groupe j est donc munie du poids $m_k = m_k^j / I_j^j$.

Cette pondération permet d'équilibrer le rôle des groupes.

1.4 Les individus

Dans l'espace \mathbb{R}^K , on cherche :

- une représentation du nuage N_j des individus caractérisés par les K variables,
- une représentation superposée des J nuages N_j^j ($j = 1 \dots J$) d'individus caractérisés chacun par un groupe de variables.

\mathbb{R}^K peut se décomposer en somme directe de J sous-espaces orthogonaux deux à deux isomorphes aux espaces \mathbb{R}^{K^j} engendrés chacun par un des J groupes de variables : $\mathbb{R}^K = \oplus \mathbb{R}^{K^j}$

Les coordonnées des points de N_j^j sont contenues dans le tableau X_j . On définit le tableau \tilde{X}_j , de dimensions (I, K) , en complétant X_j par des zéros pour les variables n'appartenant pas au groupe j :

$$\tilde{C}_j = \begin{array}{|c|c|c|} \hline & & \\ \hline 0 & C_j & 0 \\ \hline & & \\ \hline \end{array}$$

Un individu i , décrit par les variables du groupe j , est alors représenté par un point de \mathbb{R}^K , noté i^j .

Dans \mathbb{R}^K , on note N_i^* le nuage des centres de gravité, notés i^* , des J points i^j , représentant le même individu i dans les N_i^j (J individus partiels). N_i^* se déduit de N_i^j par une homothétie de rapport $1/J$: il constitue un nuage moyen pour les N_i^j .

On effectue une ACP du tableau X , l'objectif étant de projeter le nuage N_i (et donc N_i^*) sur un sous-espace de faible dimension tel que le nuage projeté ressemble le plus possible au nuage initial N_i (respectivement N_i^*). Dans cette ACP, chaque variable \mathbf{n}_k est munie du poids m_k défini précédemment.

Les J nuages N_i^j étant également tous situés dans \mathbb{R}^K , il est possible de les projeter sur un même sous-espace de \mathbb{R}^K . Ce sous-espace doit permettre de comparer les positions d'un même individu dans les différents nuages. La projection sur ce sous-espace doit permettre d'obtenir :

- une bonne représentation de chaque N_i^j
- une ressemblance entre les nuages N_i^j : éviter que les symétries, rotations ou homothéties ne masquent les ressemblances. Il ne serait en effet pas possible de comparer les positions d'un même individu dans les différents nuages si ces représentations étaient très différentes.

On montre que la représentation optimale est celle obtenue par l'ACP pondérée du tableau X . La représentation superposée des J nuages définis par chaque groupe de variables s'obtient en traitant en éléments supplémentaires de cette ACP les individus partiels. On projette donc les nuages N_i^j sur les axes factoriels de N_i .

1.5 Les variables

Dans l'espace \mathbb{R}^I , on veut représenter :

- les K variables initiales,
- les composantes principales obtenues par des ACP séparées de chacun des groupes.

La représentation des variables s'obtient par l'ACP du tableau X . Duale de l'image de N_i obtenue dans \mathbb{R}^K , elle offre une aide à l'interprétation de la représentation du nuage des individus. Elle constitue également une représentation optimale des corrélations entre variables. Elle permet donc une comparaison fine des groupes de variables.

La pondération des groupes par I/I_i^j est telle que :

- aucun groupe ne détermine à lui seul le premier axe (sauf situation de symétrie)
- un groupe contribue à la détermination de nombreux axes s'il est composé de nombreuses variables indépendantes (il est de grande "dimensionnalité").

On obtient la représentation des composantes principales de chaque groupe en les introduisant en éléments supplémentaires dans l'ACP de X . On peut ainsi repérer des facteurs communs à plusieurs groupes de variables.

1.6 Les groupes de variables

Dans \mathbb{R}^{I^2} , on cherche à comparer globalement les groupes. Pour les représenter, on va construire un nuage de J points, noté N_j . Un groupe j est représenté par la matrice $(I, I) X_j M_j X_j' D = W_j D$, ou plus précisément, en faisant apparaître explicitement la pondération de l'AFM, $W_j D / I_j^j$. Cette matrice de I^2 scalaires est considérée comme un élément de l'espace vectoriel \mathbb{R}^{I^2} , muni du produit scalaire de Hilbert-Schmidt $\langle A, B \rangle = Tr(AB)$.

On obtient ainsi une mesure de liaison entre les groupes j et ℓ :

$$L(K_j, K_\ell) = \left\langle \frac{W_j D}{I_j^j}, \frac{W_\ell D}{I_\ell^\ell} \right\rangle$$

Pour comparer les groupes ainsi représentés dans \mathbb{R}^{I^2} , on décrit les proximités entre les $W_j D$ en les projetant sur un espace de faible dimension de \mathbb{R}^{I^2} . On cherche, dans \mathbb{R}^{I^2} , un repère orthonormé qui ajuste au mieux les $W_j D$ et dont les axes sont des éléments symétriques de rang 1. On montre que ces éléments sont de la forme $z_s z_s' D$, où les z_s sont les composantes principales issues de l'ACP de X. Ils maximisent la somme des projections des $W_j D$ sur les axes. Cette somme est égale à l'inertie dans \mathbb{R}^I des K variables projetées sur z_s : $\sum_j \langle W_j D, z_s z_s' D \rangle$.

La suite des éléments recherchés est associée à celle des composantes principales z_s . L'orthonormalité des z_s dans \mathbb{R}^I est équivalente à celle des $z_s z_s' D$, dans \mathbb{R}^{I^2} . La coordonnée de $W_j D$ sur $z_s z_s' D$ peut aussi s'interpréter comme la contribution du groupe j à l'inertie de z_s .

1.7 Variables qualitatives et tableaux mixtes

L'AFM peut s'appliquer aux tableaux mixtes, dans lesquels certains groupes sont constitués de variables qualitatives, grâce à l'équivalence suivante entre ACM et ACP : les résultats d'une ACM peuvent être obtenus en réalisant une ACP normée des indicatrices relatives aux modalités des variables, à condition d'affecter à ces indicatrices des poids adéquats.

Pour réaliser une AFM sur des tableaux mixtes, on commence donc par construire les indicatrices associées aux modalités des variables qualitatives, qui seront ensuite correctement pondérées dans l'analyse.

1.8 Éléments supplémentaires

Les individus, affectés d'un poids nul, peuvent intervenir en tant qu'éléments supplémentaires. Sont alors calculées les projections de leurs points représentatifs dans N_j^* et dans les N_I^j .

Un groupe de variables j' peut également être traité en élément supplémentaire. On effectue alors : une normalisation du nuage associé semblable à celle des autres groupes, la projection des composantes principales du groupe sur les axes de l'analyse, et la projection de $W_j D$.

2. Les macros %AFMULT et %PLOTAFM

On présente ici quelques informations donnant une idée du fonctionnement de ces macros ; un guide d'utilisation complet est en cours de réalisation.

2.1 La macro %AFMULT

Les paramètres de la macro

| | |
|------------------|--------------------------------------------------------------------------------------------------------------------------|
| DATAACT | table contenant les individus actifs |
| DATASUP | table contenant les individus supplémentaires |
| ACTNUM | liste des variables actives numériques, regroupées par thèmes |
| ACTQUAL | liste des modalités des variables actives qualitatives, regroupées par variables, et par thèmes |
| NBMODACT | liste des nombres de modalités des variables actives qualitatives |
| LIBVARACT | liste de libellés des variables actives qualitatives |
| SUPNUM | liste des variables supplémentaires numériques, regroupées par thèmes |
| SUPQUAL | liste des modalités des variables supplémentaires qualitatives, regroupées par variables, et par thèmes |
| NBMODSUP | liste des nombres de modalités des variables supplémentaires qualitatives |
| LIBVARSUP | liste de libellés des variables supplémentaires qualitatives |
| ID | variable identifiant les individus |
| DIM | nombre d'axes factoriels retenus pour les aides à l'interprétation |
| INDTOT | édition des aides à l'interprétation pour tous les individus |
| NINDCONTR | nombre d'individus édités dans les tableaux donnant les n individus ayant la plus forte contribution à l'inertie (intra) |
| FEN | utilisation du fenêtrage |
| CORR | édition de la matrice des corrélations entre les variables actives |

Quelques règles sur l'écriture des paramètres

Dans les paramètres ACTNUM et SUPNUM, les variables numériques sont regroupées par thèmes, les thèmes étant séparés par des /.

Exemple : ACTNUM = A B C / H J / R S T

Les variables A, B et C constituent le 1^{er} thème, H et J le 2^{ème} thème, etc.

Dans les paramètres ACTQUAL et SUPQUAL, les modalités des variables qualitatives (i.e. les variables indicatrices 1/0 associées à ces modalités) sont regroupées par variables et par thèmes, les thèmes étant séparés par des /.

Exemple : ACTQUAL = D1 D2 D3 E1 E2 / L1 L2 / M1 M2 M3 M4 N1 N2 P1 P2 P3

D1, D2, D3, E1, E2 sont les modalités des variables du 1^{er} thème, L1, L2 celles du 2^{ème} thème, etc.

Dans les paramètres NBMODACT et NBMODSUP, les nombres de modalités des variables qualitatives sont regroupés par thèmes, séparés par des /. Il doit y avoir correspondance avec l'ordre des modalités dans les paramètres ACTQUAL et SUPQUAL.

Exemple : NBMODACT = 3 2 / 2 / 4 2 3

D1, D2, D3 sont les 3 modalités d'une première variable, E1, E2 sont les 2 modalités d'une deuxième variable, ces deux variables constituant le 1^{er} thème, etc.

Dans les paramètres LIBVARACT et LIBVARSUP, les libellés des variables qualitatives sont regroupés par thèmes, séparés par des /. Il doit y avoir correspondance avec l'ordre des nombres de modalités dans les paramètres NBMODACT et NBMODSUP.

Exemple : LIBVARACT = D E / L / M N P

Remarque : D, E, L etc. désignent des libellés utilisés dans les éditions ; ils ne font pas référence à des variables SAS.

Les paramètres FEN, CORR, INDTOT prennent les valeurs O (oui) ou N (non).

Compléments

L'utilisation du fenêtrage permet :

- des enchaînements d'écrans pour la saisie et la modification des paramètres
- des tests en interactif sur la saisie des paramètres.

L'utilisateur peut, s'il le souhaite, sauvegarder les paramètres saisis dans les différentes fenêtres.

Les résultats de l'analyse sont stockés dans plusieurs tables SAS.

2.2 La macro %PLOTAFM

L'utilisateur peut, s'il le souhaite, utiliser le module SAS-GRAPH.

Les graphiques peuvent représenter **différents objets** :

- les individus compromis actifs ou supplémentaires,
- les individus partiels correspondant aux individus sélectionnés,
- les variables actives ou supplémentaires,
- les groupes actifs ou supplémentaires.

Il est possible de **sélectionner** les objets à représenter en fonction de :

- leur contribution aux plans factoriels choisis,
- leur qualité de représentation sur ces plans.

On peut ainsi alléger les graphiques.

La **taille des points** peut être uniforme ou proportionnelle à l'un de ces deux critères :

- leur contribution aux plans factoriels choisis,
- leur qualité de représentation sur ces plans.

Ce type de représentation permet de visualiser rapidement les points bien représentés ou ceux ayant joué un rôle important dans la construction des axes.

On peut choisir des **représentations différentes** pour les individus partiels en fonction du type de données :

- temporelles (trajectoires avec ou sans superposition du point moyen)
- thématiques (oursins).

Les paramètres de la macro

| | |
|-------------|-------------------------------------------------------|
| IAM | individus actifs moyens (compromis) |
| FIAM | filtre sur les individus actifs moyens |
| ISM | individus supplémentaires moyens (compromis) |
| FISM | filtre sur les individus supplémentaires moyens |
| IP | individus partiels |
| VAQ | variables actives qualitatives |
| FVAQ | filtre sur les variables actives qualitatives |
| VAN | variables actives numériques |
| FVAN | filtre sur les variables actives numériques |
| VSQ | variables supplémentaires qualitatives |
| FVSQ | filtre sur les variables supplémentaires qualitatives |
| VSN | variables supplémentaires numériques |
| FVSN | filtre sur les variables supplémentaires numériques |
| GA | groupes actifs |
| GS | groupes supplémentaires |

PLANS plans factoriels à représenter

TYPER type de représentation choisie

TYPED type de données

FEN utilisation du fenêtrage

GRAPH utilisation du module SAS-GRAPH

Quelques règles sur l'écriture des paramètres

Les paramètres IAM, ISM, IP, VAQ, VAN, VSQ, VSN, GA, GS, FEN, GRAPH prennent les valeurs O (oui) ou N (non).

Les paramètres FIAM, FISM, FVAQ, FVAN, FVSQ s'écrivent sous la forme "C 5" ou "Q 7" pour indiquer, par exemple, que l'on sélectionne les 5 points ayant la plus forte contribution à l'inertie du plan, ou les 7 points ayant la meilleure qualité de représentation sur le plan.

Le paramètre PLANS s'écrit sous la forme "1-2 3-4" (on peut demander plusieurs plans en même temps).

Le paramètre TYPER prend les valeurs S (représentation simple), C (proportionnelle à la contribution), Q (proportionnelle à la qualité de représentation).

Le paramètre TYPED prend les valeurs THEM (données thématiques), TEMP (données temporelles), TEMPS (données temporelles + compromis).

3. Un exemple : les activités physiques et sportives

3.1 Le contexte - Les données

La Mission Statistique du Ministère de la Jeunesse et des Sports a décidé en 2001 d'élaborer une nomenclature d'activités physiques et sportives. Le but était, grâce à cet instrument, de constituer un cadre d'élaboration des comptes économiques du sport, qui concernent environ 250 disciplines. Cette nomenclature doit permettre d'obtenir une image simplifiée et révélatrice du domaine : elle contient donc un nombre restreint de postes (9 en l'occurrence).

La méthode statistique utilisée pour construire la NAPS est une technique de classification ascendante hiérarchique (CAH). Les unités statistiques sur lesquelles a porté la classification sont 34 familles d'activités, décrites par 12 critères, eux-mêmes regroupés en 5 thèmes. En amont de la CAH, on utilise une analyse factorielle multiple (AFM), pour deux raisons :

- cette méthode permet d'équilibrer le poids des thèmes dans l'analyse, et de leur donner la même importance ;
- en se limitant aux premiers facteurs de l'analyse, on effectue un lissage des données, qui permet d'éliminer une part de l'information que l'on peut considérer comme du "bruit".

Les variables en entrée de la CAH étaient donc en réalité, non pas les critères, mais les coordonnées des familles d'activités sur les premiers axes factoriels de l'analyse.

3.1.1 Les familles d'activités

Les activités physiques et sportives retenues dans l'analyse statistique proviennent de deux listes :

- les activités déclarées par les personnes dans une enquête sur les pratiques sportives (réalisées en 2000) ;
- une liste établie par la Direction des sports du Ministère concernant les disciplines organisées par une fédération agréée.

La comparaison de ces deux listes a conduit à retenir 34 familles d'activités, comportant un effectif suffisamment important pour que les données de l'enquête soient fiables, compte tenu de la taille de l'échantillon :

marche sportive, escalade - montagne - spéléo, natation - baignade, vélo, athlétisme, boules, bowling, gymnastique & yoga, sports de neige, sports de glace, football, rugby, basket, volley, handball et divers, tennis, tennis de table, badminton – squash, musculation, roller – skate, danse, équitation, voile & planche, glisse sur eau, aviron - canoë kayak, judo, autres arts martiaux, golf, sports de combat, sports mécaniques, sports aériens, escrime, tir, pêche.

3.1.2 Les critères

Les critères choisis doivent discriminer les sports d'un double point de vue économique et démographique, et doivent pouvoir être renseignés pour l'ensemble des familles d'activités. 12 critères ont pu ainsi être retenus.

2 critères qualitatifs

- type de pratique, avec 3 modalités non ordinales : individuel (1), duel (2), collectif (3)¹
- instrumentation, avec 4 modalités ordinales : pas du tout (0), faiblement (1), moyennement (2), fortement instrumenté (3).

10 critères quantitatifs

- effectif des pratiquants de 15 ans et plus (source : enquête pratiques sportives MJS/INSEP 2000)
- âge moyen des pratiquants de 15 ans et plus (source : enquête pratiques sportives MJS/INSEP 2000)
- part des femmes parmi les pratiquants de 15 ans et plus (source : enquête pratiques sportives MJS/INSEP 2000)
- part des jeunes de moins de 15 ans parmi les licenciés (source : recensement mené par le MJS auprès des fédérations sportives)
- pourcentage de licenciés parmi les 15 ans et plus (source : enquête pratiques sportives MJS/INSEP 2000 et recensement mené par le MJS auprès des fédérations sportives)
- effectifs des sportifs de haut niveau (source : base de données du MJS sur les sportifs de haut niveau)
- budget des fédérations unisport et multisport au prorata du nombre de licences connues pour chaque discipline (source : exercice 1999, direction des sports MJS)

¹ Les chiffres entre parenthèses représentent des codages de ces modalités.

- budget du secteur professionnel (source : exercice 1999, direction des sports MJS)
- médiatisation via les heures de télévision (année 2000, source : Conseil Supérieur de l'Audiovisuel)
- médiatisation via l'Équipe (année 2001, source : étude MJS).

Recodage des critères quantitatifs

Afin de réduire dans l'analyse l'importance des familles d'activités fortement "atypiques" selon un ou plusieurs critères (par exemple le vélo ou la natation pour l'effectif de pratiquants, le judo pour le pourcentage de licenciés, le football pour la médiatisation via la télévision), et pour tenir compte d'éventuelles liaisons non linéaires, il a été décidé de rendre *qualitatif* chaque critère quantitatif, par un découpage en quatre tranches homogènes, numérotées de 1 à 4 (de 0 à 3 pour le critère "budget professionnel").

3.1.3 Les thèmes

Ces 12 critères sont regroupés en 5 thèmes :

1. nature du sport : type de pratique, instrumentation
2. démographie : effectif (*), âge moyen (*), % femmes(*), % jeunes (**)
3. haute performance : budget professionnel, effectif de sportifs de haut niveau
4. degré d'organisation : % licenciés (*), budget des fédérations
5. médiatisation : temps d'antenne télé sur chaînes hertziennes, volume occupé dans l'Équipe

(*) champ : pratiquants de 15 à 75 ans

(**) parmi les titulaires d'une licence sportive

3.2. Quelques résultats

On trouvera dans les pages suivantes des extraits des sorties des macros %AFMULT et %PLOTAFM mises en œuvre sur ces données

Résultats des analyses séparées des groupes

L'examen des valeurs propres des analyses séparées permet de repérer les groupes multidimensionnels qui vont influencer le plus grand nombre de facteurs. Il permet également de comparer la pondération des différents groupes qui correspond à l'inverse de la racine carrée de la 1^{ère} valeur propre. Tous les groupes étant qualitatifs, les analyses sont ici des ACM.

Tableau résumé des valeurs propres des analyses partielles

| groupe | dim | val p1 | val p2 | val p3 | val p4 | val p5 | val p6 | val p7 | val p8 | val p9 | val p10 | val p11 | val p12 | pct1 | pct2 | pct3 | pct4 | pct5 | pct6 |
|-----------------|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|---------|---------|------|------|------|------|------|------|
| n° 1 - qual act | 5 | 0.89 | 0.74 | 0.50 | 0.26 | 0.11 | | | | | | | | 35.4 | 29.8 | 20.0 | 10.2 | 4.6 | |
| n° 2 - qual act | 12 | 0.54 | 0.49 | 0.39 | 0.29 | 0.26 | 0.22 | 0.22 | 0.21 | 0.15 | 0.10 | 0.08 | 0.04 | 18.1 | 16.4 | 13.1 | 9.7 | 8.5 | 7.4 |
| n° 3 - qual act | 6 | 0.72 | 0.63 | 0.54 | 0.46 | 0.37 | 0.28 | | | | | | | 24.0 | 21.0 | 18.1 | 15.2 | 12.3 | 9.3 |
| n° 4 - qual act | 6 | 0.78 | 0.68 | 0.52 | 0.48 | 0.32 | 0.22 | | | | | | | 26.1 | 22.8 | 17.2 | 16.1 | 10.5 | 7.2 |
| n° 5 - qual act | 6 | 0.90 | 0.78 | 0.64 | 0.36 | 0.22 | 0.10 | | | | | | | 30.1 | 25.9 | 21.5 | 11.9 | 7.4 | 3.2 |

| groupe | pct9 | pct10 | pct11 | pct12 | pctcum1 | pctcum2 | pctcum3 | pctcum4 | pctcum5 | pctcum6 | pctcum7 | pctcum8 | pctcum9 | pctcum10 | pctcum11 | pctcum12 |
|-----------------|------|-------|-------|-------|---------|---------|---------|---------|---------|---------|---------|---------|---------|----------|----------|----------|
| n° 1 - qual act | | | | | 35.4 | 65.2 | 85.2 | 95.4 | 100.0 | | | | | | | |
| n° 2 - qual act | 5.0 | 3.5 | 2.6 | 1.3 | 18.1 | 34.6 | 47.6 | 57.3 | 65.8 | 73.2 | 80.5 | 87.5 | 92.5 | 96.0 | 98.7 | 100.0 |
| n° 3 - qual act | | | | | 24.0 | 45.0 | 63.1 | 78.3 | 90.7 | 100.0 | | | | | | |
| n° 4 - qual act | | | | | 26.1 | 48.9 | 66.1 | 82.3 | 92.8 | 100.0 | | | | | | |
| n° 5 - qual act | | | | | 30.1 | 56.0 | 77.5 | 89.4 | 96.8 | 100.0 | | | | | | |

Valeurs propres de l'AFM

L'examen des valeurs propres de l'AFM et le choix des axes à interpréter se fait selon les mêmes règles qu'une analyse factorielle simple. On notera une seule particularité pour l'AFM : la pondération des groupes conduit à une inertie maximum pour n'importe quel axe égale au nombre total des groupes. Cette valeur n'est atteinte pour un axe que si ce dernier est confondu avec le 1^{er} facteur de l'ACP de chaque groupe.

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

| N° | Valeur propre | Di ff. | Pct | Pct cumulé | |
|----|---------------|--------|------|------------|--------|
| 1 | 3.09 | | 15.7 | 15.7 | |
| 2 | 1.88 | 1.22 | 9.5 | 25.3 | !***** |
| 3 | 1.71 | 0.16 | 8.7 | 34.0 | !***** |
| 4 | 1.70 | 0.02 | 8.6 | 42.6 | !***** |
| 5 | 1.32 | 0.38 | 6.7 | 49.4 | !***** |
| 6 | 1.23 | 0.09 | 6.3 | 55.6 | !***** |
| 7 | 1.17 | 0.06 | 6.0 | 61.6 | !***** |
| 8 | 1.02 | 0.15 | 5.2 | 66.8 | !***** |
| 9 | 0.96 | 0.06 | 4.9 | 71.7 | !***** |
| 10 | 0.72 | 0.23 | 3.7 | 75.3 | !***** |
| 11 | 0.71 | 0.01 | 3.6 | 79.0 | !***** |

Rapport inertie inter/inertie totale :

Le rapport (inertie inter)/(inertie totale) permet de mesurer l'intérêt d'une représentation simultanée de l'ensemble des nuages d'individus induits par chaque groupe de variables (nuages partiels)..

Rapports inertie inter/inertie totale

| axe 1 | axe 2 | axe 3 |
|-------|-------|-------|
| 0.67 | 0.50 | 0.48 |

Matrice de liaison entre les groupes : coefficients Lg

Le coefficient Lg de liaison entre deux groupes j et j' est la somme, pondérée par les poids des variables de l'AFM, des carrés des covariances entre chaque colonne (indicatrice) du groupe j et chaque colonne du groupe j'.

Il vaut 0 si chaque colonne du groupe j est non corrélée avec chaque colonne du groupe j'. Une valeur élevée de ce coefficient signifie que les groupes ont en commun des axes d'inertie importante.

Le coefficient Lg d'un groupe avec lui-même est la somme des carrés des valeurs propres du groupe (en tenant compte de la pondération de l'AFM) : il mesure la "multidimensionnalité" du groupe ; sa racine carrée statistique notée Ng (les Ng sont les termes diagonaux de la matrice).

Coefficients Lg de liaison entre groupes

| Groupe | 1 | 2 | 3 | 4 | 5 |
|-----------------|------|------|------|------|------|
| n° 1 - qual act | 2.13 | 0.40 | 0.57 | 0.51 | 0.39 |
| n° 2 - qual act | 0.40 | 3.46 | 0.77 | 0.72 | 0.81 |
| n° 3 - qual act | 0.57 | 0.77 | 3.15 | 0.86 | 1.16 |
| n° 4 - qual act | 0.51 | 0.72 | 0.86 | 2.81 | 0.78 |
| n° 5 - qual act | 0.39 | 0.81 | 1.16 | 0.78 | 2.48 |

Matrice de liaison entre les groupes : coefficients RV

$$RV(j,j') = Lg(j,j') / [Ng(j) * Ng(j')]$$

RV(j,j') est toujours compris entre 0 et 1. Il prend la valeur 1 si les nuages d'individus induits par les deux groupes sont homothétiques.

Coefficients RV de liaison entre groupes

| Groupe | 1 | 2 | 3 | 4 | 5 |
|-----------------|------|------|------|------|------|
| n° 1 - qual act | 1.00 | 0.15 | 0.22 | 0.21 | 0.17 |
| n° 2 - qual act | 0.15 | 1.00 | 0.23 | 0.23 | 0.28 |
| n° 3 - qual act | 0.22 | 0.23 | 1.00 | 0.29 | 0.41 |
| n° 4 - qual act | 0.21 | 0.23 | 0.29 | 1.00 | 0.30 |
| n° 5 - qual act | 0.17 | 0.28 | 0.41 | 0.30 | 1.00 |

Représentation simultanée des individus pour les J groupes de variables

Chaque individu est représenté par autant de points qu'il y a de groupes actifs : les individus partiels. À ceux-ci s'ajoute un point représentant l'individu moyen (centre de gravité des individus partiels correspondants).

La contribution est égale au rapport inertie projetée du point/ inertie totale de l'axe pour les individus moyens, et à la contribution à l'inertie intra pour les individus partiels.

Le cosinus carré est le rapport inertie projetée du point/inertie totale du point.

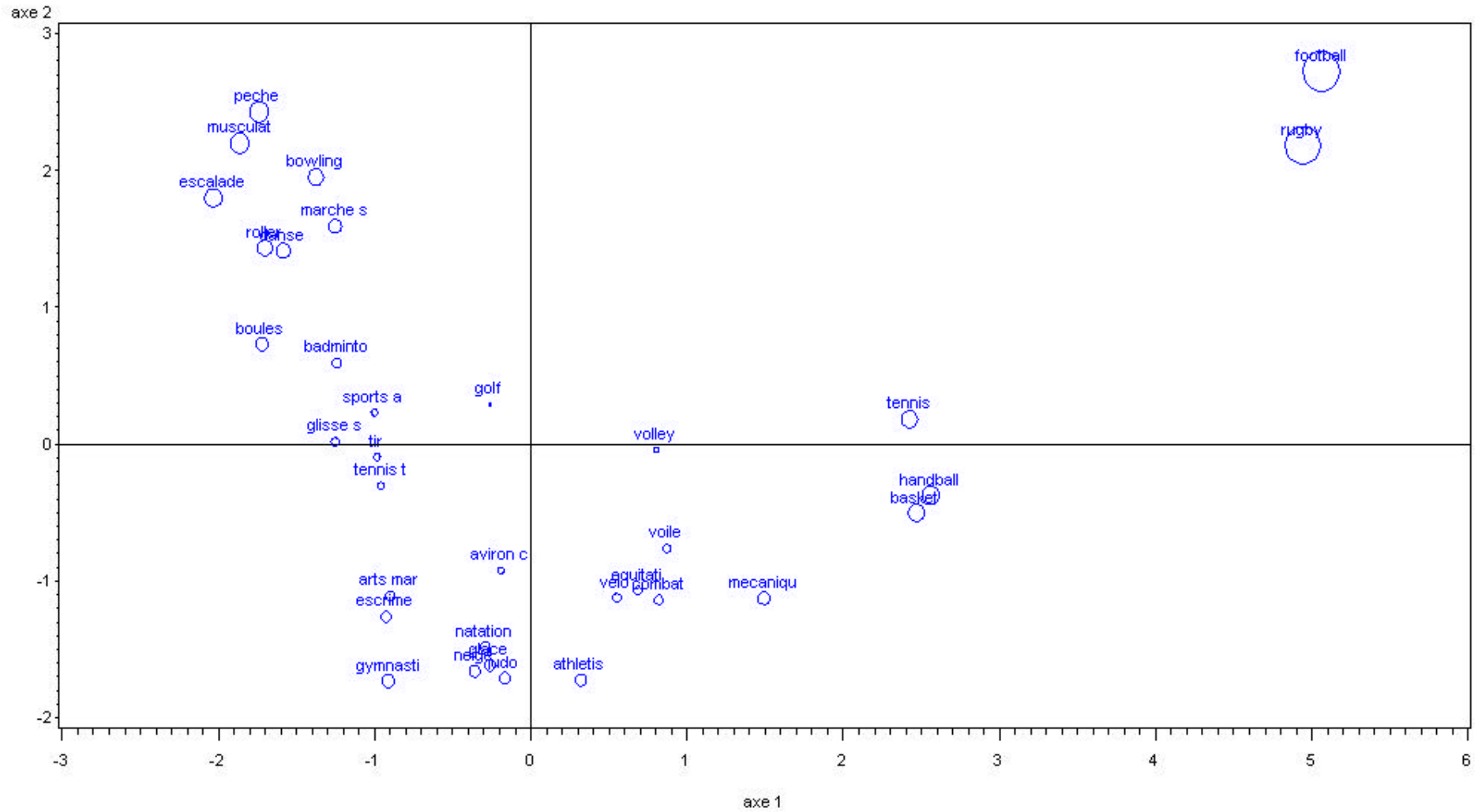
Coordonnées et aides à l'interprétation des individus actifs

| id | groupe | disto | coord1 | coord2 | coord3 | ctr1 | ctr2 | ctr3 | cos1 | cos2 | cos3 |
|----------|--------|-------|--------|--------|--------|------|------|------|------|------|------|
| marche s | | 13.09 | -1.25 | 1.59 | 1.07 | 1.5 | 4.0 | 2.0 | 12.0 | 19.4 | 8.8 |
| | 1 | -13.6 | 0.88 | 1.04 | -0.25 | 1.8 | 0.1 | 0.5 | -5.7 | -8.0 | -0.5 |
| | 2 | 136.7 | -1.81 | 1.10 | 0.94 | 0.1 | 0.1 | 0.0 | 2.4 | 0.9 | 0.7 |
| | 3 | 61.22 | -2.95 | 3.10 | -0.57 | 1.1 | 0.7 | 0.8 | 14.2 | 15.7 | 0.5 |
| | 4 | 88.57 | 0.06 | -0.69 | 3.98 | 0.7 | 1.6 | 2.6 | 0.0 | 0.5 | 17.9 |
| | 5 | 54.33 | -2.45 | 3.41 | 1.26 | 0.6 | 1.0 | 0.0 | 11.1 | 21.4 | 2.9 |
| escalade | | 11.22 | -2.03 | 1.80 | -0.09 | 3.9 | 5.1 | 0.0 | 36.8 | 28.8 | 0.1 |
| | 1 | -23.0 | -1.08 | -0.06 | -1.39 | 0.3 | 1.1 | 0.5 | -5.1 | -0.0 | -8.5 |
| | 2 | 133.0 | -1.38 | 1.19 | 2.03 | 0.2 | 0.1 | 1.4 | 1.4 | 1.1 | 3.1 |
| | 3 | 61.22 | -2.95 | 3.10 | -0.57 | 0.3 | 0.5 | 0.1 | 14.2 | 15.7 | 0.5 |
| | 4 | 54.98 | -2.30 | 1.35 | -1.77 | 0.0 | 0.1 | 0.9 | 9.6 | 3.3 | 5.7 |
| | 5 | 54.33 | -2.45 | 3.41 | 1.26 | 0.1 | 0.8 | 0.6 | 11.1 | 21.4 | 2.9 |
| natation | | 14.53 | -0.29 | -1.48 | -1.76 | 0.1 | 3.4 | 5.3 | 0.6 | 15.1 | 21.3 |
| | 1 | 19.36 | 0.88 | 1.04 | -0.25 | 0.5 | 2.0 | 0.7 | 4.0 | 5.6 | 0.3 |
| | 2 | 160.1 | -0.91 | -3.20 | -1.73 | 0.1 | 0.9 | 0.0 | 0.5 | 6.4 | 1.9 |
| | 3 | 55.85 | -0.24 | -0.34 | -1.90 | 0.0 | 0.4 | 0.0 | 0.1 | 0.2 | 6.5 |
| | 4 | 70.04 | -0.33 | -2.10 | -3.36 | 0.0 | 0.1 | 0.8 | 0.2 | 6.3 | 16.2 |
| | 5 | 57.90 | -0.87 | -2.80 | -1.55 | 0.1 | 0.5 | 0.0 | 1.3 | 13.5 | 4.2 |

...

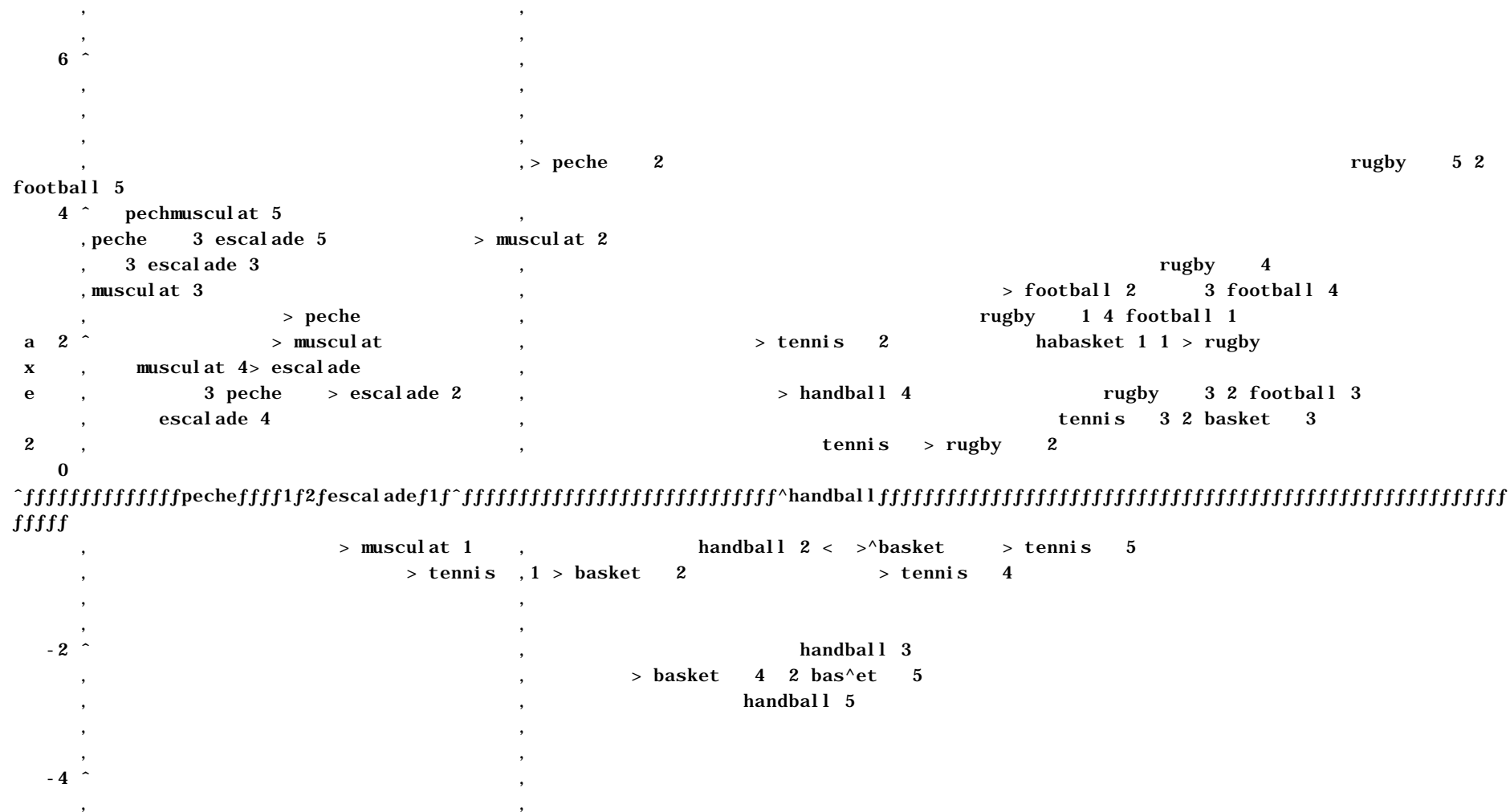
Représentation des individus compromis actifs

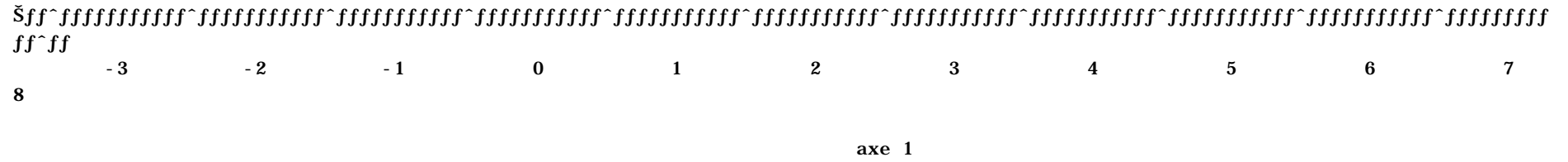
proportionnelle à la contribution à l'inertie expliquée par le plan



Représentation des individus compromis et partiels actifs
 8 individus actifs ayant la plus forte contribution à l'inertie expliquée par le plan

Plot of coord2*coord1\$_idind_. Symbol points to label.



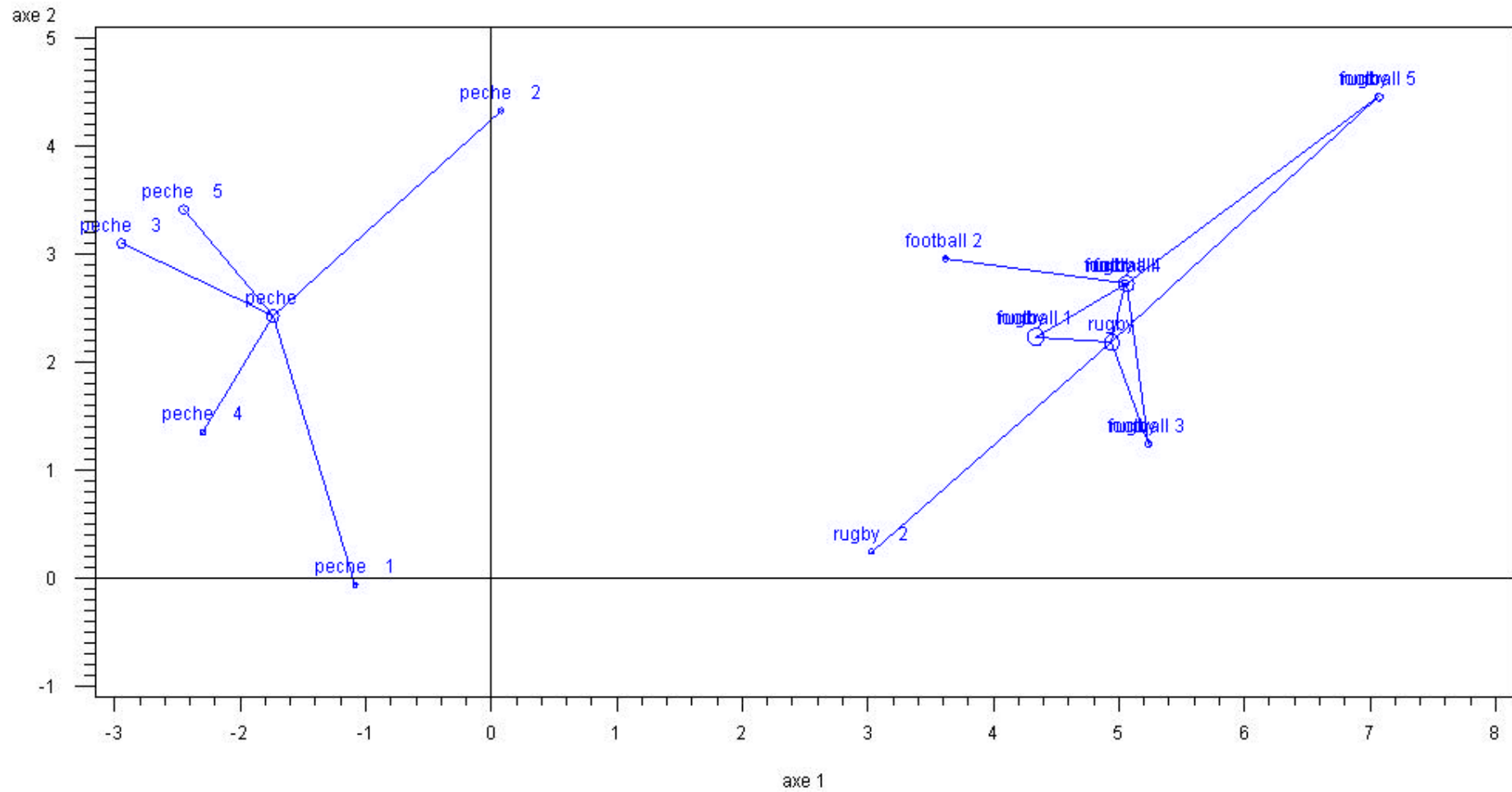


NOTE: 156 obs had missing values. 10 label characters hidden.

Représentation des individus compromis et partiels actifs

proportionnelle à la qualité de représentation sur le plan

3 individus actifs ayant la plus forte contribution à l'inertie expliquée par le plan



Individus ayant les plus fortes contributions à l'inertie de l'axe 1

| id | ctr1 | cumul |
|----------|------|-------|
| football | 24.4 | 24.4 |
| rugby | 23.2 | 47.6 |
| handball | 6.2 | 53.8 |

Individus ayant les plus fortes contributions à l'inertie de l'axe 2

| id | ctr2 | cumul |
|----------|------|-------|
| football | 11.6 | 11.6 |
| peche | 9.2 | 20.9 |
| musculat | 7.6 | 28.4 |

Individus ayant les plus fortes contributions à l'inertie de l'axe 3

| id | ctr3 | cumul |
|----------|------|-------|
| volley | 16.4 | 16.4 |
| voile | 10.9 | 27.2 |
| gymnasti | 10.7 | 37.9 |

Individus ayant les plus fortes inerties intra sur l'axe 1

| id | ctr1 | cumul |
|----------|------|-------|
| equitati | 9.6 | 9.6 |
| velo | 8.4 | 18.1 |
| volley | 7.5 | 25.6 |

Individus ayant les plus fortes inerties intra sur l'axe 2

| id | ctr2 | cumul |
|----------|------|-------|
| equitati | 6.1 | 6.1 |
| handball | 5.2 | 11.3 |

...

Aides à l'interprétation des variables

La contribution est égale à l'inertie projetée du point-variable divisée par l'inertie projetée de l'ensemble des points-variables actives.

La somme des contributions des colonnes (indicatrices) d'un groupe est égale à la contribution du groupe.

Le cosinus carré est égal au rapport : inertie projetée/inertie totale du point. Il mesure la qualité de représentation d'un groupe (en tant que point unique dans l'espace des groupes) par un facteur.

Coordonnées et aides à l'interprétation des variables actives

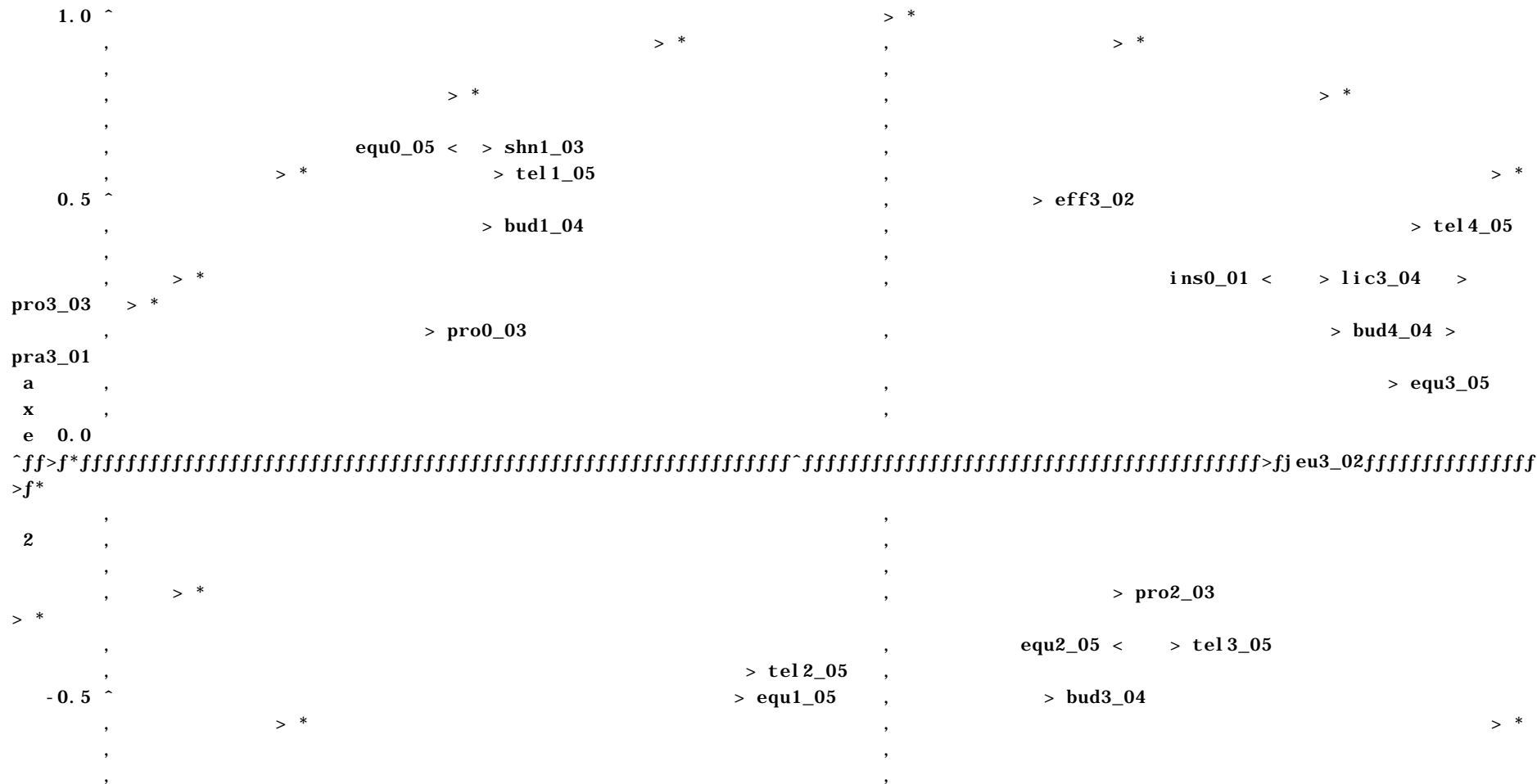
| variable | coord1 | coord2 | coord3 | ctr1 | ctr2 | ctr3 | cos1 | cos2 | cos3 |
|----------|--------|--------|--------|------|------|------|------|------|------|
| pra1_01 | -0.45 | 0.00 | -0.25 | 1.4 | 0.0 | 0.8 | 20.5 | 0.0 | 6.0 |
| pra2_01 | -0.11 | -0.20 | 0.16 | 0.2 | 1.0 | 0.7 | 1.1 | 4.1 | 2.6 |
| pra3_01 | 0.75 | 0.24 | 0.14 | 8.7 | 1.5 | 0.6 | 55.9 | 5.8 | 2.0 |
| ins0_01 | 0.50 | 0.28 | 0.04 | 3.5 | 1.8 | 0.0 | 25.1 | 7.8 | 0.2 |
| ins1_01 | -0.29 | -0.11 | -0.02 | 1.1 | 0.2 | 0.0 | 8.6 | 1.1 | 0.1 |
| ins2_01 | -0.24 | -0.02 | -0.33 | 0.7 | 0.0 | 2.4 | 5.6 | 0.1 | 11.2 |
| ins3_01 | 0.11 | -0.18 | 0.47 | 0.2 | 0.9 | 6.5 | 1.1 | 3.3 | 22.3 |
| groupe 1 | | | | 15.7 | 5.4 | 10.9 | 17.2 | 3.6 | 6.6 |
| eff1_02 | 0.08 | -0.20 | 0.20 | 0.1 | 0.6 | 0.6 | 0.7 | 4.2 | 3.8 |
| eff2_02 | -0.15 | 0.20 | 0.37 | 0.2 | 0.7 | 2.5 | 2.1 | 3.8 | 13.4 |
| eff3_02 | 0.20 | 0.50 | -0.22 | 0.5 | 5.4 | 1.2 | 4.1 | 25.2 | 4.9 |
| eff4_02 | -0.11 | -0.39 | -0.50 | 0.1 | 3.1 | 5.6 | 1.1 | 15.5 | 25.2 |
| age1_02 | 0.42 | -0.01 | -0.10 | 1.9 | 0.0 | 0.2 | 17.9 | 0.0 | 1.0 |
| age2_02 | -0.09 | -0.13 | 0.07 | 0.1 | 0.3 | 0.1 | 0.9 | 1.7 | 0.5 |
| age3_02 | -0.11 | 0.04 | 0.22 | 0.1 | 0.0 | 1.0 | 1.2 | 0.2 | 4.9 |
| age4_02 | -0.26 | 0.11 | -0.20 | 0.8 | 0.2 | 0.8 | 6.8 | 1.2 | 3.8 |
| jeu1_02 | -0.32 | 0.35 | 0.18 | 1.1 | 2.1 | 0.6 | 10.4 | 12.3 | 3.3 |
| jeu2_02 | -0.25 | 0.12 | 0.00 | 0.8 | 0.3 | 0.0 | 6.2 | 1.4 | 0.0 |
| jeu3_02 | 0.62 | -0.02 | 0.03 | 4.3 | 0.0 | 0.0 | 39.0 | 0.1 | 0.1 |
| jeu4_02 | -0.08 | -0.44 | -0.22 | 0.1 | 3.5 | 1.0 | 0.6 | 19.4 | 4.9 |
| fem1_02 | 0.29 | 0.25 | -0.01 | 1.0 | 1.2 | 0.0 | 8.4 | 6.2 | 0.0 |

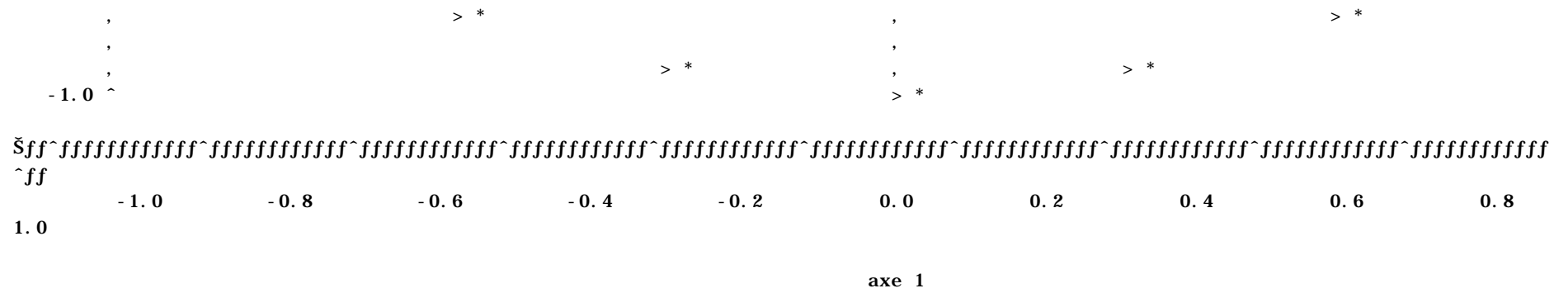
| | | | | | | | | | | | | |
|----------|--|-------|-------|-------|--|------|------|------|--|-----|-----|------|
| fem2_02 | | -0.09 | 0.06 | 0.17 | | 0.1 | 0.1 | 0.4 | | 0.7 | 0.4 | 2.8 |
| fem3_02 | | -0.08 | -0.23 | 0.04 | | 0.1 | 1.0 | 0.0 | | 0.6 | 5.1 | 0.2 |
| fem4_02 | | -0.16 | -0.15 | -0.34 | | 0.4 | 0.5 | 2.9 | | 2.6 | 2.1 | 11.8 |
| groupe 2 | | | | | | 11.5 | 19.0 | 16.9 | | 6.4 | 6.5 | 5.3 |

...

Représentation des variables actives
20 variables actives ayant la plus forte contribution à l'inertie expliquée par le plan

Plot of coord2*coord1\$idvar_. Symbol points to label.
 Plot of cos*sin\$idvar_. Symbol points to label.

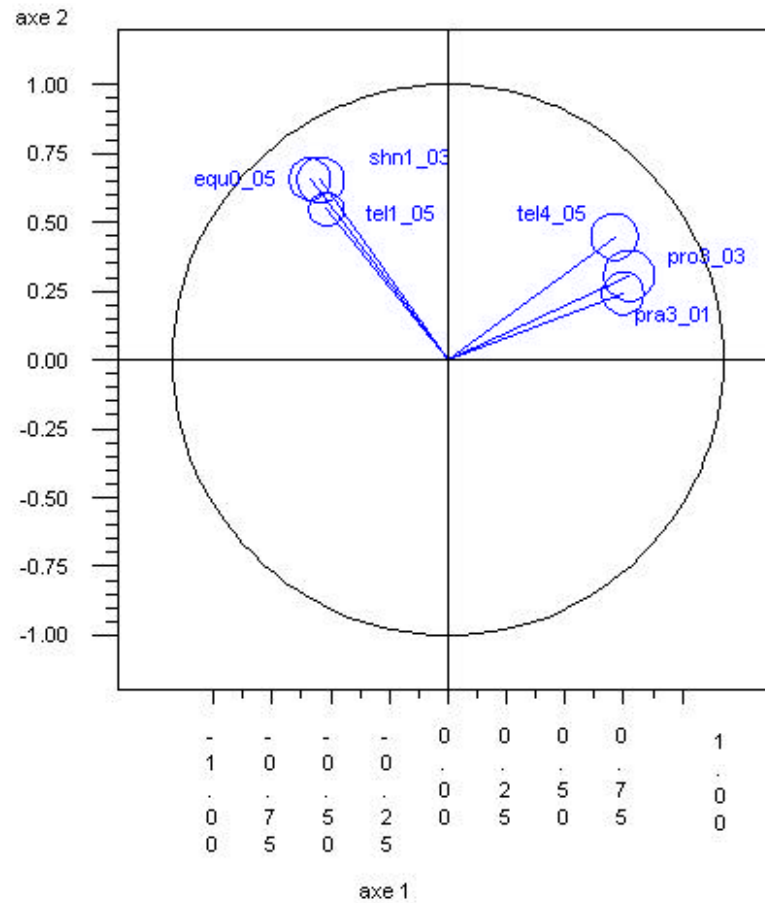




Représentation des variables actives

proportionnelle à la contribution à l'inertie expliquée par le plan

6 variables ayant la plus forte contribution à l'inertie expliquée par le plan



Aides à l'interprétation des groupes

Les résultats correspondent à l'analyse dans l'espace des groupes de variables.

Les coordonnées sont toujours comprises entre 0 et 1. La coordonnée d'un groupe sur un axe donné s'interprète, dans l'espace des variables, comme l'inertie projetée de l'ensemble des variables du groupe sur l'axe.

La contribution est égale à la coordonnée divisée par la somme des coordonnées des groupes actifs.

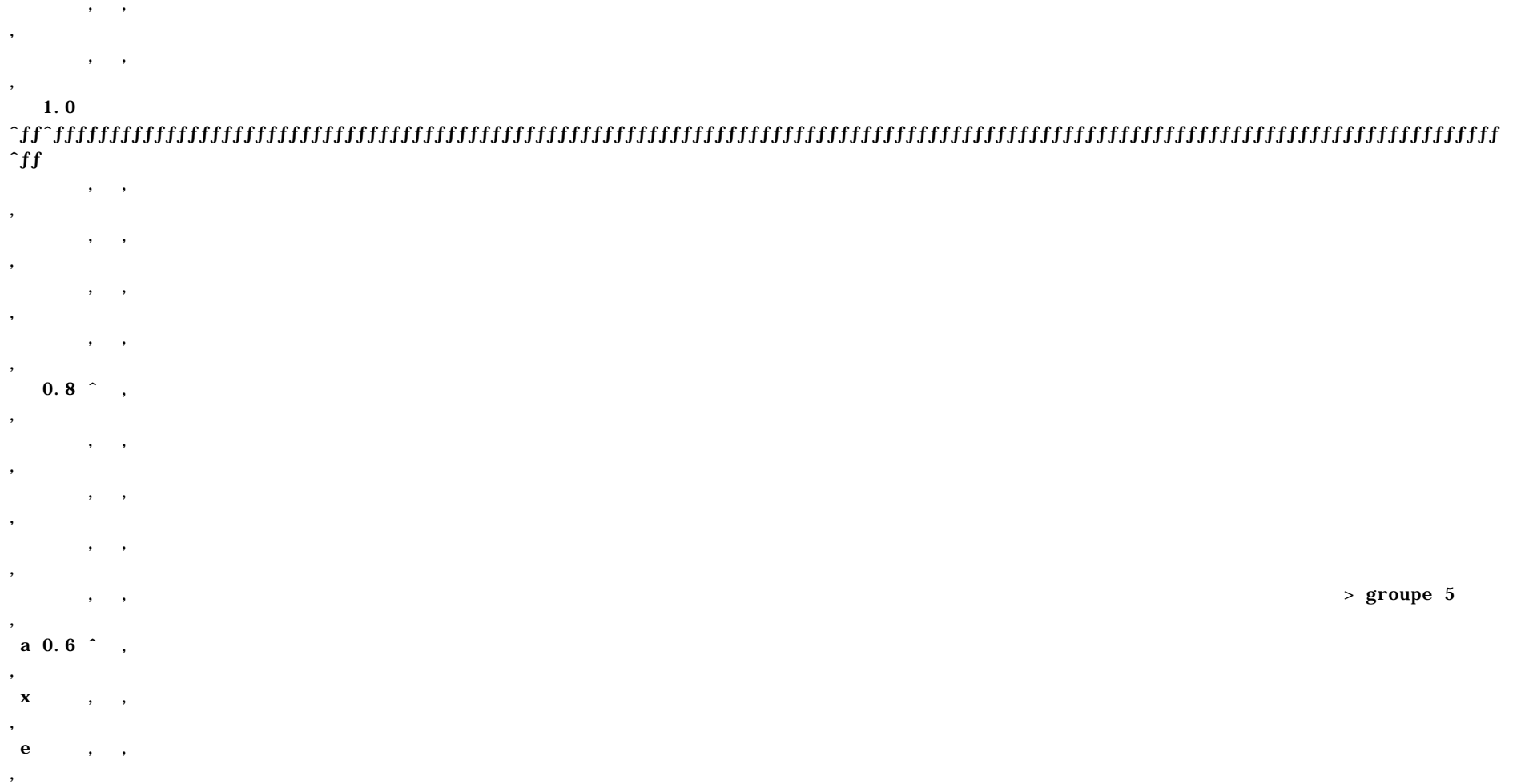
Le cosinus carré est égal au rapport : inertie projetée/inertie totale. Il mesure la qualité de représentation d'un groupe (en tant que point unique dans l'espace des groupes) par un facteur.

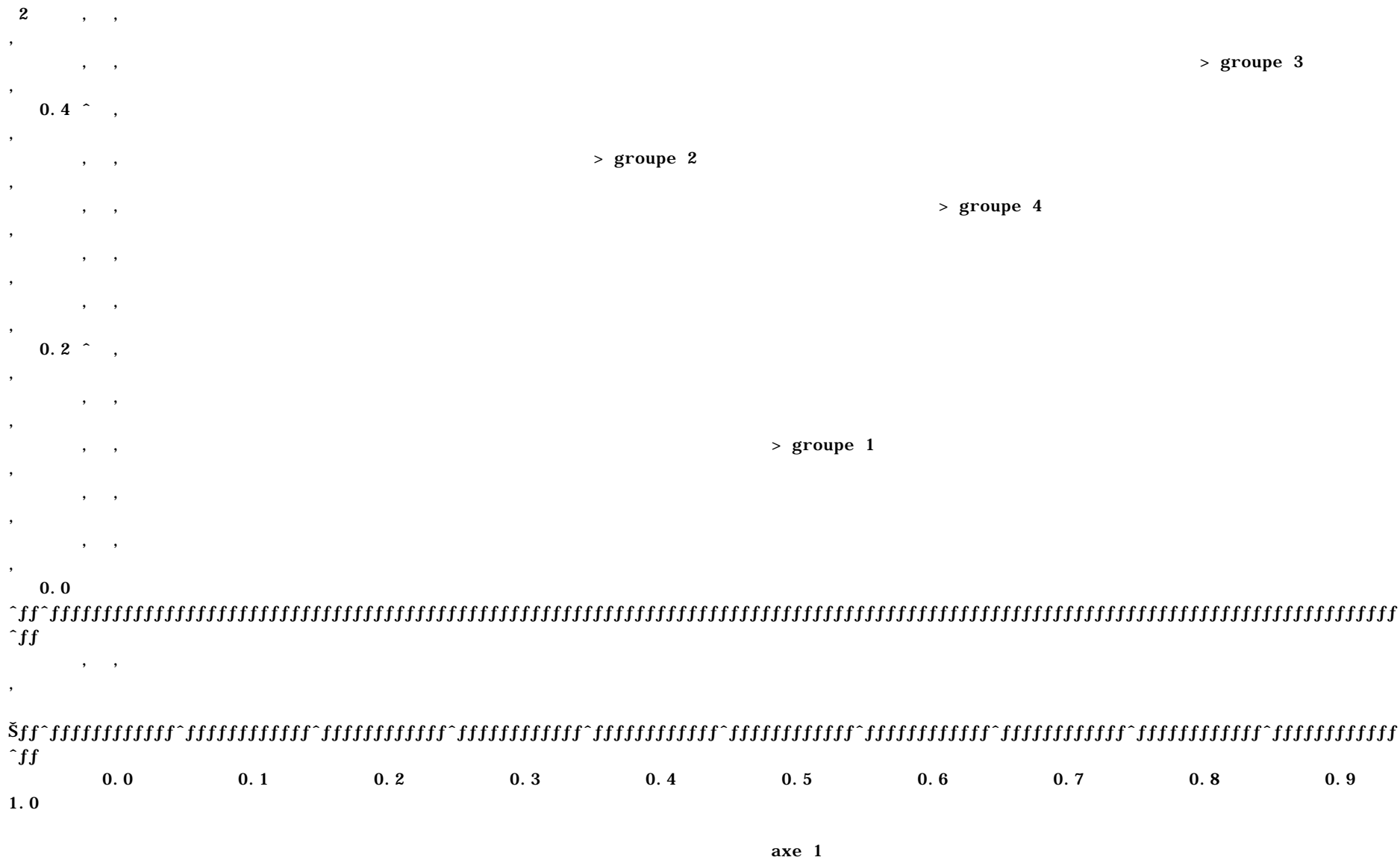
Coordonnées et aides à l'interprétation des groupes actifs

| groupe | di sto | | coord1 | coord2 | coord3 | | ctr1 | ctr2 | ctr3 | | cos1 | cos2 | cos3 |
|----------|--------|--|--------|--------|--------|--|-------|-------|-------|--|------|------|------|
| groupe 1 | 2.13 | | 0.49 | 0.10 | 0.19 | | 15.7 | 5.4 | 10.9 | | 11.2 | 0.5 | 1.6 |
| groupe 2 | 3.46 | | 0.35 | 0.36 | 0.29 | | 11.5 | 19.0 | 16.9 | | 3.6 | 3.7 | 2.4 |
| groupe 3 | 3.15 | | 0.80 | 0.44 | 0.57 | | 25.9 | 23.4 | 33.3 | | 20.4 | 6.1 | 10.3 |
| groupe 4 | 2.81 | | 0.61 | 0.32 | 0.48 | | 19.7 | 17.3 | 28.2 | | 13.2 | 3.7 | 8.3 |
| groupe 5 | 2.48 | | 0.84 | 0.66 | 0.18 | | 27.2 | 34.9 | 10.7 | | 28.5 | 17.3 | 1.4 |
| Ensemble | | | | | | | 100.0 | 100.0 | 100.0 | | 14.8 | 6.2 | 5.1 |

Représentation des groupes actifs

Plot of coord2*coord1\$nomvar_. Symbol points to label.





Bibliographie

[1] Escofier B., Pagès J. (1998), *Analyses factorielles simples et multiples*, Dunod.

[2] Groupe d'étude et de réflexion interrégional - GERI - (1996), *L'analyse des données évolutives : méthodes et applications*, Technip.

[3] Lebart L., Morineau A., Piron M. (2000), *Statistique exploratoire multidimensionnelle*, 3^{ème} édition, Dunod.

