

L'ANALYSE DES DONNÉES AVEC SAS V6

Olivier Sautory

1. L'offre SAS

Le module statistique du logiciel SAS (SAS/STAT) contient un certain nombre de procédures permettant de pratiquer de l'analyse des données "à la française".

1.1. Les procédures d'analyse factorielle

1.1.1. La procédure PRINCOMP

PRINCOMP réalise une analyse en composantes principales (ACP). La procédure édite les valeurs propres, les vecteurs propres, et les composantes principales, i.e. les coordonnées des individus sur les axes principaux. La procédure ne calcule pas d'aides à l'interprétation, et n'édite aucune information sur les variables.

PRINCOMP peut créer deux tables SAS :

- une table contenant les résultats de l'analyse : moyennes et écarts-types des variables, matrice des corrélations, valeurs propres et vecteurs propres ;
- une table contenant, outre les données de la table en entrée, les valeurs des composantes principales.

1.1.2. La procédure FACTOR

Comme PRINCOMP, FACTOR permet d'effectuer des ACP, mais elle calcule de plus les corrélations entre les variables et les composantes principales.

FACTOR offre également la possibilité à l'utilisateur averti de réaliser des analyses factorielles au sens anglo-saxon du terme. Dans les ouvrages anglo-saxons – et dans SAS –, l'analyse factorielle désigne une méthode proposée par Spearman (1904) dans un contexte psychométrique, et présentant certaines analogies avec l'ACP ; elle s'en distingue par une modélisation a priori des effets des variables (les "facteurs communs" et les "facteurs spécifiques", qui sont... inobservables) ainsi que par les "rotations" que

l'on peut opérer sur les axes factoriels. Ces techniques sont peu utilisées par les statisticiens français.

FACTOR peut créer deux tables SAS :

- une table contenant les résultats de l'analyse : moyennes et écarts-types des variables, matrice des corrélations, valeurs propres et vecteurs propres, corrélations entre variables et composantes principales... ;
- une table contenant, outre les données de la table en entrée, les valeurs des composantes principales.

1.1.3. La procédure CORRESP

La procédure CORRESP permet de réaliser une analyse factorielle des correspondances (AFC) ou une analyse des correspondances multiples (ACM)... au sens français du terme : l'analyse des correspondances est une "popular data analysis method in France and Japan"¹ qui trouve donc dans cette procédure, introduite dans la version 6 du logiciel, un moyen de conquérir des terres plus ou moins inconnues.

Contrairement aux deux procédures précédentes, CORRESP peut traiter des points supplémentaires (individus, variables, modalités, selon les cas), et édite des aides à l'interprétation.

Dans le cas d'une AFC (analyse de la liaison entre deux variables qualitatives A et B), les données en entrée de la procédure peuvent se présenter sous l'une des formes suivantes :

- **tableau individus x variables**, i.e. table SAS où les lignes ("observations" dans le langage SAS) sont les individus, et les colonnes ("variables" dans le langage SAS) sont les variables A et B : c'est la procédure qui construit le tableau de contingence croisant A et B qui sera soumis à l'analyse ;
- **tableau de contingence**, ou tableau croisé, i.e. table SAS où les lignes sont les modalités de la variable A, les colonnes les modalités de la variable B ;
- table SAS créée en sortie d'une PROC FREQ.

Note : dans les versions SAS antérieures, la procédure CORRESP ("procédure utilisateur" écrite, et disponible, à l'INSEE) n'acceptait les données en entrée que sous la forme d'un tableau de contingence.

1. (SAS/STAT User's guide, volume 1, page 616).

Dans le cas d'une ACM (analyse des liaisons entre M variables qualitatives A, B, C...), les données en entrée de la procédure peuvent se présenter sous l'une des formes suivantes :

- **tableau individus x variables**, i.e. table SAS où les lignes sont les individus, et les colonnes sont les variables A, B, C... : c'est la procédure qui construit le tableau de Burt croisant A, B, C... qui sera soumis à l'analyse (à noter que, dans ce cas, la procédure ne donne aucun résultat concernant les individus). On peut également demander à la procédure de construire le tableau disjonctif complet, et d'analyser ce tableau, ce qui permet d'obtenir les résultats relatifs aux individus ;
- **tableau disjonctif complet**, i.e. table SAS où les lignes sont les individus, et les colonnes sont les modalités des variables qualitatives A1 A2... B1 B2... C1 C2... (tableau de 1 et de 0) ;
- **tableau de Burt**, i.e. table SAS où les lignes sont les modalités des variables qualitatives, et les colonnes également.

Note : dans les versions SAS antérieures, la procédure CORRESP (INSEE) n'acceptait les données en entrée que sous la forme d'un tableau disjonctif complet.

La procédure édite les valeurs propres, les coordonnées des points, les contributions aux inerties des axes (CTR), les qualités de représentation des points sur les axes (CO2 et QLT), des indices indiquant les points expliquant le mieux l'inertie des axes (BEST) ; elle édite également, à la demande, les tableaux des effectifs théoriques sous l'hypothèse d'indépendance, des écarts à l'indépendance, des contributions au chi-deux du tableau.

CORRESP peut créer deux tables SAS :

- une table contenant le tableau de données analysé (tableau de contingence, tableau disjonctif complet ou tableau de Burt), les tableaux des profils-lignes et des profils-colonnes, les tableaux des écarts à l'indépendance, des contributions au chi-deux ;
- une table contenant les résultats de l'analyse : valeurs propres, coordonnées, aides à l'interprétation.

Pour une présentation détaillée de la procédure CORRESP, on peut se reporter au document cité en référence [5].

1.2. Les procédures de classification

1.2.1. La procédure CLUSTER

CLUSTER effectue une classification ascendante hiérarchique (CAH), à partir d'un tableau observations x variables, ou bien d'un tableau de distances. Pas moins de 11 stratégies d'agrégation (= modes de calcul des "distances" entre classes) sont proposées, parmi lesquelles on trouve :

- la méthode de WARD, ou minimisation de la variance intra-classe de la partition ;
- la méthode CENTROID, ou distance entre les centres de gravité des classes ;
- la méthode SINGLE linkage, ou minimum des distances entre les points des classes ;
- la méthode COMPLETE linkage, ou maximum des distances entre les points des classes ;
- la méthode AVERAGE linkage, ou moyenne des distances entre les points des classes ;
- la méthode DENSITY linkage, fondée sur une estimation non paramétrique de densité de probabilité.

La procédure édite un "historique" du processus de classification, i.e. des informations sur les dernières agrégations : les classes jointes, l'effectif des classes, les distances entre classes, divers indicateurs statistiques dépendant de la méthode choisie.

La procédure peut créer une table qui peut être utilisée en entrée de la procédure TREE (voir § 1.2.4.).

1.2.2. La procédure FASTCLUS

FASTCLUS réalise une classification non hiérarchique, i.e. construit une partition en un nombre de classes fixé a priori, selon la méthode des centres mobiles. L'utilisation de cette procédure, plutôt que CLUSTER, est recommandée lorsque l'on travaille sur un grand nombre d'observations (par exemple de l'ordre de plusieurs dizaines de milliers, ou plus), car FASTCLUS est dans un tel cas beaucoup plus rapide.

La procédure édite une description rapide des classes obtenues : effectif, distance maximale entre un point de la classe et son centre, distance entre le centre de la classe et le centre le plus proche, moyenne et écart-type de chaque variable dans la classe...

FASTCLUS peut créer deux tables SAS :

- une table donnant des informations pour chaque classe de la partition (du genre de celles qui sont éditées) ;
- une table analogue à la table en entrée de la procédure, enrichie de deux variables :
 - une variable donnant le numéro de la classe à laquelle appartient l'observation ;
 - une variable donnant la distance entre l'observation et le centre de la classe.

1.2.3. La procédure VARCLUS

VARCLUS réalise une classification (hiérarchique ou non hiérarchique) sur un ensemble de **variables** : chaque classe de variables est représentée par une variable synthétique (la variable "moyenne" ou la 1^{ère} composante principale), et les classes sont choisies de façon à maximiser la somme des parts de variance expliquées par ces variables synthétiques.

VARCLUS peut créer deux tables SAS :

- une table contenant des statistiques relatives aux différentes classes ;
- une table pouvant être utilisée en entrée de la procédure TREE (dans le cas où on a réalisé une classification hiérarchique).

1.2.4. La procédure TREE

TREE permet de tracer l'arbre de classification, ou dendrogramme, représentant les résultats d'une classification ascendante hiérarchique : elle utilise la table créée par la procédure CLUSTER (dans le cas d'une CAH sur individus) ou par la procédure VARCLUS (dans le cas d'une CAH sur variables). L'arbre produit par SAS est un arbre **américain** : assez différent des arbres français, il peut dans un premier temps surprendre quelque peu l'utilisateur familiarisé avec les logiciels d'analyse des données français. Mais avec un peu d'habitude, et de recul¹, on s'y adapte.

Cette procédure permet également de créer une table SAS dans laquelle, à chaque observation de l'ensemble soumis à la classification, est attribué le numéro de la classe

1. En effet, cet arbre se visualise bien "en négatif", en s'éloignant du listing d'une distance de l'ordre de 50 cm...

à laquelle elle appartient, pour une partition en un nombre de classes choisi par l'utilisateur.

1.3. Les procédures d'analyse discriminante

1.3.1. La procédure DISCRIM

DISCRIM réalise une analyse discriminante (sur variables numériques), avec un nombre de classes quelconque. Elle permet la mise en œuvre de méthodes :

- paramétriques (hypothèse de normalité des distributions des variables explicatives) :
 - méthode linéaire ;
 - méthode quadratique ;
- non paramétriques :
 - méthode du noyau (estimation locale de densité) ;
 - méthode des plus proches voisins.

La procédure évalue la méthode utilisée par des estimations des taux d'erreur de classement. Elle peut également effectuer une analyse canonique discriminante (comme CANDISC).

La procédure peut créer plusieurs tables SAS :

- une table contenant, outre les données de la table en entrée, les probabilités a posteriori d'affectation dans chaque classe, ou l'estimation de densité de chaque classe (selon la méthode utilisée), ainsi que la classe d'affectation ;
- une table contenant les résultats de la discrimination : différentes statistiques, fonctions discriminantes, etc.

1.3.2. La procédure STEPDISC

STEPDISC réalise une sélection de variables numériques en vue d'une analyse discriminante, par une méthode pas à pas :

- ascendante ;

- ascendante avec "remise en cause", i.e. avec possibilité de "sortir" une variable précédemment entrée (méthode "stepwise") ;
- descendante.

Le critère utilisé pour entrer, ou sortir, une variable, est le niveau de signification d'un test F d'une analyse de covariance, qui permet de juger de l'intérêt de sélectionner, ou de maintenir, une variable.

1.3.3. La procédure CANDISC

CANDISC réalise une analyse canonique discriminante, i.e. une analyse factorielle discriminante (au sens utilisé dans les ouvrages d'analyse des données français).

La procédure peut créer deux tables SAS :

- une table contenant les résultats de l'analyse : statistiques des variables dans les classes, composantes des variables canoniques (ou variables discriminantes)... ;
- une table contenant, outre les données de la table en entrée, les valeurs des variables discriminantes, ce qui permet de représenter, à l'aide d'une procédure PLOT, les observations dans les "plans discriminants".

1.4. Commentaires

L'offre SAS en analyse des données est donc assez importante. Si elle est relativement complète dans le domaine de l'analyse discriminante, elle laisse néanmoins sur sa faim l'utilisateur habitué à utiliser les logiciels d'analyse des données français, tels que ADDAD, SPAD, STAT-ITCF... En effet, les procédures SAS ne fournissent pas les aides à l'interprétation¹ familières aux statisticiens français, ni les graphiques de plans factoriels. Il faut toutefois souligner que ces procédures peuvent être utilement complétées grâce à l'utilisation de procédures usuelles de SAS, telles que MEANS, FREQ... pour les aides à l'interprétation d'une classification, PLOT (en version 6.07) pour les graphiques.

Plusieurs voies ont été utilisées pour enrichir l'offre en analyse des données au sein de l'univers SAS.

1. Sauf CORRESP, mais sous une forme un peu lourde.

1.4.1. Les procédures-utilisateurs

Jusqu'à la version 5 du logiciel, un utilisateur un peu expérimenté pouvait assez aisément écrire des procédures spécifiques adaptées à des méthodes statistiques non prises en compte par le logiciel. Ainsi, l'INSEE avait écrit au début des années 80 des procédures de noms (significatifs) PRINCIP, CORRESP, GRAPHE, CAH, qui présentaient les résultats de ces différentes méthodes (à peu près) comme dans les logiciels d'analyse des données. Le passage à la version 6 de SAS a rendu malheureusement caduques ces procédures, qui devaient être intégralement réécrites à l'aide de l'outil SAS/TOOLKIT, dont la disponibilité n'est que récente. Si diverses raisons ont conduit l'institut à ne pas envisager la reconversion de ces procédures, il n'en reste pas moins que la possibilité d'écrire en SAS des procédures d'analyse des données à la française existe ; toutefois, à notre connaissance, personne ne s'est encore engagé dans cette voie.

1.4.2. Les macros-SAS

Le langage macro de SAS permet d'"habiller" les procédures d'analyse des données disponibles dans SAS, par des enchaînements d'étapes DATA et de procédures, de façon à compléter ces procédures, en termes de statistiques calculées et en termes de présentation des résultats. Cette possibilité a été utilisée depuis longtemps par de nombreux statisticiens. Une macro-SAS offre l'avantage d'être, en principe, à peu près indépendante de la version SAS utilisée, et à peu près portable sur tout système. En revanche, un programme écrit en macro-langage n'est pas toujours optimum quant au temps de calcul.

C'est cette voie qui a été choisie par les statisticiens de l'INSEE : un certain nombre de macros ont été écrites en 1993, et sont présentées au § 2.

1.4.3. Les interfaces avec les logiciels

Le module SAS/TOOLKIT permet de réaliser l'interfaçage entre SAS et un autre logiciel, l'objectif étant d'accéder aux programmes d'un logiciel plus complet que SAS, tout en restant dans l'univers SAS, en particulier en travaillant directement sur des tables SAS. Des interfaces ont été écrites entre SAS et ADDAD, et entre SAS et SPAD, et sont présentées au § 3.

2. Les macros-SAS de l'INSEE

2.1. Les macros d'analyse factorielle

2.1.1. La macro ACP

La macro SAS ACP réalise une analyse en composantes principales "à la française" : elle utilise la procédure SAS PRINCOMP, mais les résultats sont édités sous la forme classique rencontrée dans les principaux logiciels d'analyse des données français (en particulier la macro édite des aides à l'interprétation). De plus, cette macro permet d'introduire des "points supplémentaires", qu'il s'agisse d'individus ou de variables.

Cette macro permet de définir des barycentres d'individus, et produit des tests de significativité sur ces barycentres.

Les résultats de l'analyse peuvent être stockés dans une table SAS, et réutilisés par la suite par une autre procédure, ou par la macro PLOTACP.

2.1.2. La macro AIDEAFC

La macro SAS AIDEAFC édite les aides à l'interprétation des résultats d'une analyse factorielle des correspondances (AFC) sous la forme classique rencontrée dans les principaux logiciels français d'analyse des données.

Cette macro n'effectue pas une AFC : elle doit être précédée d'une procédure CORRESP. La macro utilise la table-OUTC créée par cette procédure, et qui contient les résultats de l'AFC. On peut demander les aides à l'interprétation pour les d premiers axes (au plus), où d est le paramètre spécifié dans l'option DIMENS de la PROC CORRESP.

2.1.3. La macro AIDEACM

La macro SAS AIDEACM édite les aides à l'interprétation des résultats d'une analyse des correspondances multiples (ACM) sous la forme classique rencontrée dans les principaux logiciels français d'analyse des données. En particulier, elle calcule des valeurs-tests qui permettent de juger de la significativité des modalités supplémentaires.

Cette macro n'effectue pas une ACM : elle doit être précédée d'une procédure CORRESP. La macro utilise la table-OUTC créée par cette procédure, et qui contient

les résultats de l'ACM. On peut demander les aides à l'interprétation pour les d premiers axes (au plus), où d est le paramètre spécifié dans l'option DIMENS de la PROC CORRESP.

2.1.4. La macro PLOTACP

La macro SAS PLOTACP produit des représentations graphiques associées à une analyse en composantes principales réalisée à partir de la macro ACP. Elle utilise la procédure SAS PLOT. Les graphiques peuvent représenter les individus actifs, les individus supplémentaires (y compris les barycentres d'individus), les variables actives, les variables supplémentaires, et les "axes unitaires", i.e. les projections des axes de départ de l'espace des individus (axes représentant les variables actives de l'analyse) sur le plan factoriel considéré.

Cette macro utilise la table en sortie créée par une macro ACP.

2.1.5. La macro PLOTCOR

La macro SAS PLOTCOR produit des représentations graphiques associées à une analyse factorielle des correspondances (AFC) ou à une analyse des correspondances multiples (ACM) réalisée à partir de la procédure CORRESP. Elle utilise la procédure SAS PLOT.

Dans le cadre d'une AFC, les graphiques peuvent représenter les modalités-lignes actives, les modalités-lignes supplémentaires, les modalités-colonnes actives et les modalités-colonnes supplémentaires sur le plan factoriel considéré.

Dans le cadre d'une ACM, les graphiques peuvent représenter les individus actifs, les individus supplémentaires, les modalités actives et les modalités supplémentaires sur le plan factoriel considéré.

2.2. Les macros de classification ascendante hiérarchique

2.2.1. La macro CAHNUM

La macro SAS CAHNUM réalise une classification ascendante hiérarchique, à l'aide des procédures CLUSTER et TREE de SAS, sur les tableaux du type : individus (en lignes) X variables **numériques** (en colonnes).

La macro réalise une classification sur les **individus**. Les variables peuvent être réduites (i.e. divisées par leurs écart-types) ou non. On peut utiliser comme stratégie d'agrégation la méthode de WARD (maximisation de l'inertie interclasse du nuage des individus), ou toute autre méthode proposée dans la procédure CLUSTER.

La macro édite le tableau des nœuds issu de la classification, et à la demande l'arbre de classification. Les résultats de la classification sont stockés dans une table SAS, et peuvent être réutilisés par la suite par les macros PARTNUM, DESNUM..., si l'on souhaite réaliser et analyser une partition de l'ensemble des individus.

2.2.2. La macro CAHQUAL

La macro SAS CAHQUAL permet de réaliser une classification ascendante hiérarchique, à l'aide des procédures CLUSTER et TREE de SAS, sur les types de tableaux suivants :

- tableau de contingence ;
- tableaux de contingence juxtaposés ;
- tableau disjonctif complet.

La macro réalise une classification sur les **observations** (i.e. les lignes du tableau). La distance utilisée pour mesurer les proximités entre les observations est la **distance du chi-deux**. On peut utiliser comme stratégie d'agrégation la méthode de WARD (fondée sur la décomposition de l'inertie du nuage des individus), ou toute autre méthode proposée dans la procédure CLUSTER.

La macro édite le tableau des nœuds issu de la classification, et à la demande l'arbre de classification. Les résultats de la classification sont stockés dans une table SAS, et peuvent être réutilisés par la suite par les macros PARTQUAL, DESQUAL..., si l'on souhaite réaliser et analyser une partition de l'ensemble des observations.

2.2.3. La macro PARTNUM

La macro SAS PARTNUM permet de réaliser une partition d'un ensemble d'individus à partir des résultats d'une classification ascendante hiérarchique effectuée à l'aide de la macro CAHNUM (i.e. sur un tableau individus x variables numériques).

Elle édite des aides à l'interprétation de la partition : décomposition de l'inertie, des distances entre les centres des classes et l'origine, des distances entre les centres des

classes... Elle permet également d'obtenir la composition des classes, et la description des classes à l'aide des variables numériques analysées.

Cette macro doit être précédée de l'appel de la macro CAHNUM. Elle utilise la table-TABN créée par CAHNUM, qui contient les résultats de la classification.

2.2.4. La macro PARTQUAL

La macro SAS PARTQUAL permet de réaliser une partition à partir des résultats d'une classification ascendante hiérarchique effectuée à l'aide de la macro CAHQUAL sur un tableau de contingence, sur des tableaux de contingence juxtaposés, ou sur un tableau disjonctif complet. Dans les deux premiers cas, les unités partitionnées sont les modalités de la variable qualitative définissant les lignes du tableau ; dans le dernier cas, les unités partitionnées sont les individus.

La macro édite des aides à l'interprétation de la partition : décomposition de l'inertie, des distances entre les centres des classes et l'origine, des distances entre les centres des classes... Elle permet également d'obtenir la composition des classes, et la description des classes à l'aide des modalités de la (des) variable (s) analysée (s).

Cette macro doit être précédée de l'appel de la macro CAHQUAL. Elle utilise la table-TABN créée par CAHQUAL, qui contient les résultats de la classification.

2.2.5. La macro DESNUM

La macro SAS DESNUM permet de décrire la partition d'un ensemble d'observations à l'aide d'une ou plusieurs variables numériques "explicatives". Elle opère sur une table SAS contenant (au moins) :

- une "variable de classe" : cette variable (numérique ou caractère) à p modalités définit une partition de la population en p classes, une classe étant constituée par l'ensemble des observations prenant une même modalité ;
- les variables numériques explicatives.

La macro édite un tableau donnant les moyennes et les écarts-types des variables dans chaque classe, ainsi que des tests statistiques permettant de caractériser les classes de la partition par ces variables.

Note : la macro DESNUM produit le même type de sorties que celles de la macro PARTNUM lorsque le paramètre DESC de PARTNUM vaut OUI. Cette macro peut

donc être utilisée comme aide à la description d'une partition lorsque l'on désire l'expliquer avec des variables "supplémentaires" n'ayant pas pris part à la classification.

2.2.6. La macro DESQUAL

La macro SAS DESQUAL permet de décrire la partition d'un ensemble d'observations à l'aide d'une ou plusieurs variables qualitatives "explicatives". Elle opère sur une table SAS contenant (au moins) :

- une "variable de classe" : cette variable (numérique ou caractère) à p modalités définit une partition de la population en p classes, une classe étant constituée par l'ensemble des observations prenant une même modalité ;
- pour chaque variable qualitative¹ à q modalités, q variables numériques représentant les q modalités.

La macro édite les tableaux de contingence croisant la variable de classe avec chaque variables qualitative, ainsi que des tableaux et des tests statistiques permettant de caractériser les classes de la partition par les modalités des variables explicatives.

Note : La macro DESQUAL produit le même type de sorties que celles de la macro PARTQUAL lorsque le paramètre DESC de PARTQUAL vaut OUI. Cette macro peut donc être utilisée comme aide à la description d'une partition lorsque l'on désire l'expliquer avec des variables "supplémentaires" n'ayant pas pris part à la classification.

Les descriptions complètes de ces macros sont données dans les documents [2] et [3] cités en référence.

3. Les interfaces entre SAS et des logiciels d'analyse des données

3.1. L'interface SPAD•N – SAS

Le logiciel SPAD•N² est un logiciel spécialisé dans le traitement d'enquêtes. Orienté vers les techniques d'analyse de données, l'accent y est mis sur les aides à l'interprétation et la complémentarité entre les approches analyse factorielle et classification.

1. La variable qualitative peut ne pas être une variable SAS de la table.

2. SPADN est une marque déposée du Centre International de Statistique et d'Informatique Appliquées

La réalisation d'une interface entre SPAD•N et SAS est une co-production CISIA, CNRS Strasbourg, EDF Etudes et Recherches, INRETS et INSEE (voir la référence [1]). Cette interface, la procédure SPADN, permet d'accéder à un sous-ensemble du logiciel SPAD•N, à travers un certain nombre d'enchaînements. Un enchaînement est composé d'une analyse et de programmes permettant d'affiner les résultats de l'analyse.

Les enchaînements proposés actuellement sont les suivants :

ANACORMU	analyse des correspondances multiples, avec description des axes, graphiques factoriels, classification sur coordonnées factorielles
ANACORBI	analyse des correspondances binaires (AFC), avec description de axes, graphiques factoriels, classification sur coordonnées factorielles
ANACOPRI	analyse en composantes principales, avec description des axes, graphiques factoriels, classification sur coordonnées factorielles
ANADEMOD	description des modalités d'une variable qualitative
ANADESCO	description d'une variable continue
ANADIS2E	recherche exploratoire de modèles pour une analyse discriminante linéaire à 2 groupes
ANADIS2G	analyse discriminante linéaire à 2 groupes et scoring.
ANATYTRA	analyse typologique de trajectoires.

Les résultats des différentes analyses peuvent être sauvegardés dans des tables SAS ; les fichiers de sortie de SPAD•N sont également disponibles pour être repris par d'autres logiciels pouvant les récupérer.

3.2. L'interface ADDAD-SAS

Le logiciel ADDAD¹ est un ensemble de programmes d'analyse des données écrits en Fortran, incluant des programmes d'aides à l'interprétation, couvrant les domaines suivants :

1. Association pour le Développement et la Diffusion de l'Analyse de Données

- analyse factorielle :
 - analyse factorielle des correspondances ;
 - analyse en composantes principales ;
 - analyse des correspondances par sous-tableaux.
- classification :
 - classification ascendante hiérarchique ;
 - classification non hiérarchique (nuées dynamiques, boules optimisées) ;
- régression (linéaire, par voisinage) ;
- analyse discriminante linéaire.

Ces programmes sont écrits de façon modulaire : la réalisation d'une analyse (CAH par exemple), nécessite l'enchaînement de plusieurs programmes.

L'interface ADDAD-SAS "permet d'évoluer dans le monde SAS en profitant des avantages des programmes ADDAD" [4]. Cette interface, la procédure ADDADSAS, permet d'utiliser les programmes ADDAD sur des tables SAS, et de créer éventuellement de nouvelles tables.

BIBLIOGRAPHIE

- [1] AUGENDRE, H. et LADIRAY, D. (1993). SAS et l'analyse des données "à la française": une interface entre SAS et SPAD• N. *Actes du Club SAS 93, Toulouse, 13-15 octobre 1993.*
- [2] ISNARD, M. et SAUTORY, O. (1993). Les macros SAS d'analyse factorielle. *Document de travail de la Direction des statistiques démographiques et sociales, n° F 9303 Insee.*
- [3] ISNARD, M. et SAUTORY, O. (1993). Les macros SAS de classification ascendante hiérarchique. Documentation provisoire. *Document de travail de la Direction des statistiques démographiques et sociales Insee.*
- [4] LEBEAUX, M.-O. (1993). Pourquoi un interface entre ADDAD et SAS? Intégration avec SAS/TOOLKIT. *Actes des XXVèmes journées de Statistique, Vannes, 24-28 mai 1993.*
- [5] SAUTORY, O. (1993). La procédure CORRESP (version SAS 6.07). *Document de travail de la Direction des statistiques démographiques et sociales, n° F 9116 (3ème version) Insee.*