

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

BILOT ET ANALYSE DES CORRESPONDANCES
MULTIPLES : THÉORIE ET ANALYSE D'UN ENSEMBLE
DE DONNÉES

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

ANNIE LEMIEUX

OCTOBRE 2006

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 -Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

Je tiens à remercier ma directrice de mémoire, madame Pascale Rousseau, professeure au département de mathématiques et statistiques de l'Université du Québec à Montréal, pour son soutien, sa patience et ses conseils très appréciés. De plus, ce mémoire n'aurait pu être réalisé sans la merveilleuse participation de madame Suzanne Laberge et ses données qui ont été durement recensées.

Je voudrais également remercier le personnel du département de mathématiques, en particulier Gisèle Legault pour son soutien technique. En définitive, je voudrais remercier mes amis(es) et mon mari pour ses encouragements tout au long de ces deux années.

TABLE DES MATIÈRES

LISTE DES TABLEAUX	vii
LISTE DES FIGURES	viii
RÉSUMÉ	x
INTRODUCTION	1
CHAPITRE I	
ANALYSE PRÉLIMINAIRE DU CAS RÉEL	6
1.1 Première analyse	6
1.2 Ensemble de données	9
1.3 Transcription des programmes SPSS en SAS	9
1.3.1 Lecture des données originales	10
1.3.2 Création de variables dichotomiques	11
1.3.3 Élimination des données manquantes	12
1.3.4 Réduction de niveaux	12
1.3.5 Création de la variable summix	13
1.3.6 Transformation linéaire	14
1.3.7 Création des variables fclus et mclus	14
1.3.8 Ajustement des variables fclus et mclus	15
1.3.9 Assignation d'un individu à un cluster	16
1.3.10 Construction du Biplot	17
1.3.11 ANOVA	17
1.3.12 L'analyse des clusters	18
CHAPITRE II	
ANALYSE EN COMPOSANTES PRINCIPALES ET BILOT	19
2.1 Analyse en Composantes Principales (ACP)	19
2.2 Biplot	21
CHAPITRE III	

BIPLOT ET ANALYSE DES CORRESPONDANCES MULTIPLES	26
3.1 Analyse en Composante Multiple (ACM)	26
3.1.1 Matrice indicatrice \mathbf{G}	27
3.1.2 Propriétés de la matrice \mathbf{G}	29
3.1.3 Analyse des Correspondances Multiples	32
3.1.4 Centrage de \mathbf{X} et $ACM \equiv ACP$ sur \mathbf{X}	34
3.2 Biplot sur ACM	37
3.2.1 Définition des « <i>Category Level Point</i> »	37
3.3 Matrice de Burt	39
CHAPITRE IV	
APPLICATION DU BIPLLOT ASSOCIÉ À L'ACM	41
4.1 Programme SAS sur le Biplot	42
4.2 Validation des clusters des filles	43
4.2.1 Positionnement des points CLP	44
4.2.2 Élimination du cluster 1	46
4.2.3 Analyse du cluster 2	47
4.2.4 Analyse du cluster 3	48
4.2.5 Analyse du cluster 4	49
4.2.6 Conclusion chez les filles	49
4.3 Validation des clusters des garçons	50
4.3.1 Positionnement des points CLP	51
4.3.2 Élimination des clusters 1 et 4	51
4.3.3 Analyse du cluster 2	52
4.3.4 Analyse du cluster 3	53
4.3.5 Analyse du cluster 5	54
4.3.6 Conclusion chez les garçons	54
CONCLUSION	56
APPENDICE A	
DESCRIPTION DES VARIABLES DU CAS RÉEL	58
APPENDICE B	

LES PROGRAMMES	61
B.1 Programme 1 : Lecture des données originales	61
B.2 Programme 2 : Création de variables dichotomiques	65
B.3 Programme 3 : Élimination des données manquantes	69
B.4 Programme 4 : Réduction de niveaux	70
B.5 Programme 5 : Création de la variable <i>summix</i>	71
B.6 Programme 6 : Transformation linéaire	72
B.7 Programme 7 : Création des fclus et mclus	73
B.8 Programme 8 : Ajustement des variables fclus et mclus	74
B.9 Programme 9 : Assignation d'un individu à un cluster	75
B.10 Programme 11 : ANOVA	76
APPENDICE C	
RÉSULTATS DES PROGRAMMES SAS	78
C.1 Résultats programme 1	78
C.1.a Lecture des données originales avec label	78
C.1.b Tableaux des fréquences des variables	88
C.2 Résultats programme 2 : Données dichotomiques	109
C.3 Résultats programme 3	110
C.3.a Variables utilisées pour l'analyse	110
C.3.b Sortie des 47 variables dichotomiques	110
C.4 Résultats programme 4	112
C.4.a Réduction de niveaux chez les filles	112
C.4.b Réduction de niveaux chez les gars	112
C.5 Résultats programme 5 : Calcul de la variable <i>summix</i>	112
C.6 Résultats programme 6 : Sortie des transformations linéaires	113
C.7 Résultats programme 7 : Calcul des fclus pour les filles	114
C.8 Résultats programme 8 : Recalcul des fclus pour les filles selon <i>summix</i>	115
C.9 Résultats programme 9	116
C.9.a L'assignation des 20 premières filles	116

C.9.b	Descriptions statistiques des clusters filles	116
C.9.c	Descriptions statistiques des clusters garçons	117
C.9.d	Tableau croisé filles	118
C.9.e	Tableau croisé garçons	119
C.10	Résultats programme 11	120
C.10.a	Tableau anova filles	120
C.10.b	Tableau anova garçons	121
APPENDICE D		
	PROGRAMME BILOT	123
APPENDICE E		
	BILOTS DES FILLES	141
APPENDICE F		
	BILOTS DES GARÇONS	152
APPENDICE G		
	LES 47 VARIABLES DICHOTOMIQUES	163
G.1	Description des 47 variables	163
G.2	Clusters des filles	164
G.3	Clusters des garçons	164
G.4	Description des clusters des filles	165
G.5	Description des clusters des garçons	166
	BIBLIOGRAPHIE	168

LISTE DES TABLEAUX

3.1 Ensemble de données fictif	28
3.2 Tableau des variables qualitatives dichotomiques	28
A.1 Description des variables	58
G.1 Description des variables du cas réel	163
G.2 Description des clusters des filles	165
G.3 Description des clusters des garçons	166

LISTE DES FIGURES

E.1	Biplot des 470 filles identifiées par leur cluster et les 47 variables utilisées pour le clustering	142
E.2	Biplot des 470 filles identifiés par leur cluster	143
E.3	Biplot construit avec 470 filles et 47 variables : 47 variables affichées . .	144
E.4	Biplot construit avec 470 filles et 47 variables : 47 variables et cluster 1 affichés	145
E.5	Biplot construit avec 470 filles et 47 variables : 47 variables et cluster 2 affichés	146
E.6	Biplot construit avec 470 filles et 47 variables : 28 variables et cluster 2 affichés	147
E.7	Biplot construit avec 470 filles et 47 variables : 47 variables et cluster 3 affichés	148
E.8	Biplot construit avec 470 filles et 47 variables : 28 variables et cluster 3 affichés	149
E.9	Biplot construit avec 470 filles et 47 variables : 47 variables et cluster 4 affichés	150
E.10	Biplot construit avec 470 filles et 47 variables : 28 variables et cluster 3 affichés	151

F.1	Biplot des 403 garçons identifiés par leur cluster et les 47 variables utilisées pour le clustering	153
F.2	Biplot des 403 garçons identifiés par leur cluster	154
F.3	Biplot construit avec 403 garçons et 47 variables : 47 variables affichées	155
F.4	Biplot construit avec 403 garçons et 47 variables : 47 variables et cluster 4 affichés	156
F.5	Biplot construit avec 403 garçons et 47 variables : 47 variables et cluster 2 affichés	157
F.6	Biplot construit avec 403 garçons et 47 variables : 24 variables et cluster 2 affichés	158
F.7	Biplot construit avec 403 garçons et 47 variables : 47 variables et cluster 3 affichés	159
F.8	Biplot construit avec 403 garçons et 47 variables : 24 variables et cluster 3 affichés	160
F.9	Biplot construit avec 403 garçons et 47 variables : 47 variables et cluster 5 affichés	161
F.10	Biplot construit avec 403 garçons et 47 variables : 24 variables et cluster 5 affichés	162

RÉSUMÉ

Les graphiques biplots peuvent être associés à l'analyse des correspondances principales, à l'analyse canonique, aux modèles biadditifs, à l'analyse des correspondances ainsi qu'à l'analyse des correspondances multiples. Chacune de ces techniques nécessite des mathématiques algébriques particulières afin de tracer les graphiques biplots.

Dans le cadre de ce travail, on présente la théorie mathématique de l'analyse des correspondances multiples ainsi que celle du biplot associé à cette analyse. La méthodologie du biplot y est présentée à l'aide d'une analyse des correspondances multiples sur un ensemble de données; celui-ci fut recueilli à la suite d'une enquête faite par madame Suzanne Laberge sur la mixité au secondaire dans les cours d'éducation physique. Une première étude avait auparavant été faite avec le logiciel statistique SPSS. Ce travail présente tout d'abord ce premier rapport ainsi que la transcription des programmes SPSS en programmes SAS (logiciel utilisé à l'UQÀM). Puis, les résultats obtenus de la première étude de madame Laberge sont analysés à l'aide des graphiques biplots et les résultats que nous obtiendrons nous amèneront à vouloir améliorer cette étude.

INTRODUCTION

Le but de ce mémoire est de visualiser l'information qui a conduit aux conclusions d'une analyse de données faite dans un cas réel. Cette visualisation sera faite à l'aide du biplot associé à l'Analyse des Correspondances Multiples. Nous croyons que le biplot aidera à améliorer les conclusions de cette analyse et ouvrira de nouvelles voies pour soutirer davantage d'information pertinente. La problématique du cas réel est présentée, ainsi que les mathématiques de l'analyse des correspondances multiples et celles du biplot qui lui est associée.

Le cas réel étudié porte sur la mixité au secondaire qui est une question très d'actualité dans notre système scolaire. En effet, notre société voudrait pouvoir expliquer certains comportements d'étudiants(es) qui seraient, selon certains, en relation avec la mixité dans les écoles (c'est-à-dire, la présence simultanée des garçons et des filles aux mêmes activités physiques et sportives). C'est donc un sujet assez populaire des dernières années chez les sociologues québécois(es). En particulier, madame Suzanne Laberge, professeure au département de kinésiologie de l'Université de Montréal, a dirigé un mémoire sur ce sujet en 1996 en effectuant un sondage dans 11 polyvalentes du Québec auprès de 1122 étudiants(es) afin d'étudier leurs attitudes face à la mixité dans les activités physiques et sportives. Son objectif était de vérifier la présence ou l'absence de corrélations entre des variables socio-économiques et l'attitude face à la mixité dans le sport collectif. Les thèmes principaux du questionnaire étaient les suivants :

- La pratique d'activité physique dans les temps libres
- Les réactions et opinions face à la mixité dans les cours d'éducation physique
- Le rapport au corps, c'est-à-dire le rapport que les individus entretiennent avec leur corps
- Les conceptions relatives à la différence selon le sexe, c'est-à-dire l'importance

ou l'absence d'une distinction entre les filles et les garçons

- Les données socio-démographiques

Une analyse statistique sur les données recueillies a déjà été faite précédemment à partir de cette enquête à l'aide du logiciel SPSS et des conclusions avaient été émises. Cependant, nous croyons que l'utilisation de la technique d'analyse multidimensionnelle appelée *Biplot* peut nous aider à retirer davantage d'information de cet ensemble de données et nous aider à tirer des conclusions plus précises. Cette technique consiste en la représentation des modèles mathématiques par des modèles géométriques (graphiques) dans lesquels apparaît le positionnement des individus par des points ainsi que l'information donnée par les variables qui ont été observées. Le morphème *Bi* du terme *biplot* est une indication de cette double information. La réalisation de cette double représentation est essentiellement faite à l'aide de l'algèbre linéaire.

R. Gabriel a proposé cette technique en 1971 dans un contexte d'*Analyse en Composantes Principales* et de *Modèles Biadditifs*. La mathématique de cette technique a été développée et généralisée par John Gower à partir de 1998 dans les contextes de l'*Analyse en Composantes Principales*, du « *Multidimensional Scaling* », des *Modèles Biadditifs*, de l'*Analyse Canonique* et de l'*Analyse en Correspondance Multiple*. Une présentation de toutes ces idées est faite dans la monographie *BILOTS* (John Gower et David Hand, Chapman and Hall, 1996).

Les variables étudiées par madame Laberge dans son enquête sont du type qualitatif. La technique du *biplot* dans le contexte de l'*Analyse des Correspondances Multiples* devenait alors appropriée.

Remarquons que l'implantation sur ordinateur (avec le langage SAS ou S+) de la technique du *biplot*, dans les contextes de l'*Analyse en Composante Principales*, des *Modèles Biadditifs* et de l'*Analyse Canonique*, avait déjà été faite par messieurs Mohamed Nessim, Kamerane Meak et Sylvain Francoeur respectivement et madame Pascale Rousseau.

Afin d'utiliser les programmes informatiques cités ci-dessus, et de les compléter, nous devons continuer d'utiliser le langage SAS. Nous avons aussi à modifier le programme Biplot de l'Analyse en Composante Principale et l'utiliser dans un contexte d'*Analyse en Correspondance Multiple*.

Étant donné que les programmes informatiques utilisés par madame Laberge avaient été préalablement écrits en langage SPSS, nous avons d'abord à les transcrire en langage SAS. De cette façon, nous allons pouvoir continuer cette analyse à l'aide des programmes de SAS déjà implantés.

Remarquons que pour visualiser les graphiques biplots, ceux-ci sont faits dans l'espace à deux ou trois dimensions ; ceux-ci ne sont donc en général que des approximations aux modèles mathématiques multidimensionnels ; ainsi de l'information peut-être perdue. L'étude de l'information que l'on peut retrouver dans ces graphiques biplots (faits dans des cas réel) est à ses débuts.

Mon étude portera sur l'information que l'on peut retirer de ces graphiques biplots dans le cas réel de l'attitude face aux activités physiques et sportives des garçons et des filles en milieu pédagogique, niveau secondaire.

Cette étude est réalisée par :

- l'étude de la théorie mathématique du biplot associée à l'analyse des correspondances multiples
- l'écriture d'une extension au programme informatique en langage SAS qui implante la technique du biplot dans le contexte de l'*Analyse en Correspondances Multiples*
- la recherche de l'information pertinente qui peut être retrouvé dans ces graphiques

Remarquons que le questionnaire de madame Laberge était basé sur un choix de réponses multiples. L'élève devait faire un choix parmi les choix proposés qui correspondait le mieux à sa réalité. Le défi m'a tout de suite intéressé pour la simple raison que

c'est un sujet qui est très d'actualité et que le logiciel SAS est très de mise dans le milieu du travail en statistique. De plus, ce genre de rapport ressemble à ceux que je rencontrerai au cours de ma carrière en tant que statisticienne. Que ce soit en psychologie, en pharmacologie ou sociologie.

Les sujets qui seront abordés dans cet ouvrage sont :

- Les logiciels SAS et SPSS
- L'analyse multivariée
- L'analyse de données catégoriques
- Le clustering

L'analyse envisagée exploitera donc des sujets récents de la statistique, notamment le biplot associé à des données catégoriques et le clustering. Ces biplots permettent non seulement de visualiser la proximité des individus mais, aussi de caractériser cette proximité à l'aide des variables originales qui sont également représentées sur ce même graphique.

On pourra ainsi valider les clusters obtenus à l'aide de différentes techniques de clustering.

De plus, ce sujet m'a paru très pertinent pour deux raisons; premièrement, il permet à des chercheurs en sciences sociales d'avoir accès à des techniques statistiques récentes ainsi qu'à poursuivre leurs études en retirant davantage d'informations grâce à nos résultats et, deuxièmement, pour l'avancement des statistiques car, on applique ces derniers développements théoriques à des données réelles et on illustre comment utiliser ces développements. Remarquons que l'utilisation de cette technique dans des cas réels est à ses débuts.

Ce travail est divisé en quatre chapitres. Le premier présente la problématique de madame Laberge. On y présente un résumé du premier compte rendu qui avait été fait, la conversion du programme SPSS en SAS et de quelques résultats pertinents issus du nouveau programme SAS. Ce qui permettra à madame Laberge ainsi qu'à nous de bien situer l'étude préalable.

Le deuxième chapitre porte sur méthodologie utilisée dans les programme informatiques déjà implantés : le Biplot et l'Analyse en Composante Principale. À l'aide de l'algèbre linéaire, on démontre les principes de base du Biplot et de l'ACP.

Le troisième chapitre porte sur la méthodologie utilisée dans ce mémoire : Biplot sur données catégoriques et Analyse en Correspondances Multiples. Toujours avec l'algèbre linéaire, on y démontre comment pondérer les variables afin de tracer les graphiques biplots et comment on peut retirer de l'information pertinente.

Finalement, le dernier chapitre présente résultats biplots que l'on obtient et fait la critique de ceux-ci. On y parle de clustering et de la position des clusters dans le plan.

CHAPITRE I

ANALYSE PRÉLIMINAIRE DU CAS RÉEL

Dans ce premier chapitre, nous allons présenter le premier travail de madame Suzanne Laberge sur la mixité au secondaire. La problématique sur laquelle porte l'enquête faite par madame Laberge ayant été décrite d'une façon succincte dans l'introduction, nous allons présenter dans ce chapitre l'ensemble de données, les variables observées ainsi que l'analyse qui avait été fait en janvier 1996 sous la direction de madame Suzanne Laberge du département d'éducation physique de l'Université de Montréal. Puis, nous présenterons la conversion des programmes SPSS en langage SAS. Finalement, nous terminerons ce chapitre par la présentation de quelques résultats, de cette première analyse, obtenus avec les programmes réécrits en langage SAS.

1.1 Première analyse

Les objectifs généraux de la recherche sur la mixité de madame Laberge, comportaient deux volets; le premier consistait en une enquête auprès d'un échantillon représentatif d'adolescents(es) québécois(es) (volet quantitatif) et le second en une analyse de données rédigée sur un nombre plus restreint d'adolescents(es) afin d'identifier des relations entre la mixité et des variables socio-économiques. Le rôle de ce mémoire se retrouve dans ce deuxième volet. En effet, une double classification (regroupement) avait d'abord été élaborée : sous-regroupement des variables et regroupement des étudiants(es) en cluster. Notre rôle est de confirmer, d'améliorer ainsi que de porter une critique sur ces clusters d'étudiants(es) et regroupement de variables afin de vérifier

et de déterminer, s'il y a lieu, les relations sous-jacentes entre variables et clusters. Ces clusters et regroupements de variables serviront à :

- Connaître et résumer l'attitude des adolescents(es) face à la mixité dans les cours d'éducation physique au secondaire et savoir dans quelle mesure la mixité constitue un contexte d'apprentissage émancipatoire ou aliénant
- Établir des relations entre les sous-groupes de variables et les clusters d'étudiants(es) afin d'établir les caractéristiques des étudiants(es) quant à leur attitude face à la mixité dans les cours d'éducation physique
- Tenter de cerner les relations pouvant exister entre différentes variables influençant l'attitude des étudiants(es) face à la mixité et leur rapport avec leur corps

Conclusions quantitatives tirées

Plusieurs conclusions quantitatives ont été tirées par madame Laberge dans le rapport de janvier 1996. Afin de faire ressortir les idées principales de ce rapport ; nous avons extrait quelques unes de ces conclusions dont les suivantes.

- Une bonne majorité d'adolescentes (77p.100) et d'adolescents (82p.100) sont en faveur de la mixité dans les cours d'éducation physique. Toutefois, le fait que les garçons se montrent plus enthousiastes est peut-être indicateur du fait que certains y voient une occasion d'exhiber leur «supériorité» physique plus facilement parce qu'ils sont en présence de personnes qu'ils jugent plus faibles, i.e. des filles.
- Les adolescentes sont plus nombreuses que les adolescents à accorder une importance prioritaire à la bonne condition physique (statistiquement significatif à 78p.100 contre 69p.100).
- Plus d'adolescents que d'adolescentes (28p.100 contre 17p.100) manifestent un préjugé négatif quant à la «féminité» des filles habiles en sport. Ce préjugé, encore trop tenace, contribue sûrement en partie à la réticence des filles à

s'impliquer activement dans le domaine sportif. Notons par ailleurs que leur fermeture quant à la transformation de l'ordre de genre dans le domaine des activités physiques est encore plus prononcée à l'endroit de leur propre genre ; en effet, 44p.100 des adolescents (contre seulement 12p.100 chez les adolescentes) estiment que «les cours de danse et d'expressions corporelles, c'est pas fait pour les gars»

Sommes toutes, dans ce rapport on pouvait conclure facilement que les adolescents(es) sont en majorité en faveur de la mixité dans les cours d'éducation physique.

Bref aperçu de la méthodologie utilisée

Dans le cadre de l'analyse sur les sous-regroupements des variables, une analyse statistique avait été faite. Nous avons eu accès à tous les programmes informatiques SPSS de cette analyse (sauf 1 ou 2).

Ce que l'on sait est qu'on y avait tenté d'expliquer les différentes attitudes à l'aide des variables socio-économiques. Tout d'abord des regroupements de variables (définissant les attitudes face aux activités physiques et sportives) avaient été faits, puis une combinaison linéaire des variables d'un même groupe avait été calculée pour chaque individu formant ainsi de nouvelles variables. Chaque étudiant obtenait sa propre cote sur chacune de ces nouvelles variables. Ces regroupements de variables sont présentés aux tableaux G.1 et G.2 (respectivement filles et garçons) de l'annexe G. Ainsi, le nombre de variables avait été réduit à 5 pour les garçons et 4 pour les filles. Puis, à l'aide de ces nouvelles variables, on avait formé des clusters d'adolescents(es) ; chaque cluster devait contenir les étudiants qui se comportaient semblablement sur ces nouvelles variables et qui devaient se comporter le plus différemment possible des adolescents(es) d'un autre cluster. Le but était d'ensuite établir des relations entre ces clusters d'étudiants et des variables socio-économiques. Les clusters d'adolescents(es) avaient été construit à l'aide des 4 nouvelles variables pour les filles fclus1 à fclus4 (5 nouvelles variables pour les garçons, mclus1 à mclus5) de la façon suivante : un(e) adolescent(e) était

assigné(e) au cluster i si sa valeur pour la fclus i (mclus i) était maximale. Ainsi, un cluster d'adolescents(es) était constitué d'adolescents(es) ayant un score maximal sur l'ensemble des variables de ce cluster. Ainsi, une classification des individus y avait été développée d'une façon ad hoc.

Conclusion en suspens

Contrairement au premier volet quantitatif sur la mixité, madame Suzanne Laberge n'avait pas été satisfaite quant aux regroupements des adolescents(es) par clusters et les relations entre les variables socio-économiques et l'attitude des étudiants(es) face aux activités physiques et sportives. La relation ainsi établie avait été jugée trop faible mais, madame Laberge persistait à croire qu'une relation plus forte entre l'attitude face à la mixité dans le sport et certaines variables socio-économiques se devait d'exister.

Notre premier but est d'abord de visualiser, à l'aide de biplots, les relations qui existent entre ces clusters d'étudiants et les variables originales et aussi de visualiser les relations qui existent entre ces clusters et les nouvelles variables.

Nous espérons ainsi pouvoir visualiser les faiblesses de ces combinaisons de variables qui ont certainement un impact sur le manque de clarté de la relation que l'on croit qu'il devrait exister entre clusters d'étudiants et variables socio-économiques. Nous espérons ainsi, par la suite, améliorer l'analyse déjà faite.

1.2 Ensemble de données

L'ensemble de données est composé de 1122 étudiants(es) (588 filles et 534 garçons) et 68 variables. Les variables et leur définition sont présentés à l'annexe A.

1.3 Transcription des programmes SPSS en SAS

Dans cette section, nous allons revoir le cheminement du passage de SPSS en langage SAS. Cette section a été assez laborieuse car, il a fallu apprendre le langage

SPSS, décoder les programmes déjà écrits en SPSS et deviner (faute de commentaires) le but statistique des instructions. Donc, il nous a fallu plusieurs semaines à défricher ces programmes afin de les convertir en SAS.

Remarquons que tous les programmes SAS ainsi que leurs sorties statistiques (en partie ou en totalité selon le cas) dont nous ferons référence au cours de ce chapitre, sont présentés aux annexes B et C respectivement.

Notons que seuls les programmes relatifs aux filles, décrits dans cette section, sont présentés en annexe. La version relative aux garçons (légèrement différente de celle des filles) à été fournie à madame Laberge avec ce rapport.

Tous les noms des programmes SAS débutent par 1., 2.,...,11. Ce chiffre indique l'ordre d'exécution des programmes. J'utilise tout au long de mes programmes SAS la fonction **libname**. Cette fonction permet à SAS de garder en mémoire une étape **DATA** qui servira dans un programme ultérieur.

1.3.1 Lecture des données originales

Le premier programme, **1.q.mixidon** (annexe B.1) consiste

- à la création d'un **DATA** pour la lecture des données originales
- à l'attribution d'étiquettes aux variables avec l'instruction **Label**
- à l'attribution d'étiquettes aux valeurs des variables qualitatives avec **PROC FORMAT**

On obtient en sortie $n = 1122$ étudiants(es) et $p = 68$ variables. Les 21 premières lignes (individus) sont présentées au tableau C.1.a à l'annexe C.

Remarquons que la sortie de l'ensemble de données avec étiquette pour chaque variable est très volumineux et ne peut être présentée en annexe en entier.

Les fréquences d'observations pour chaque valeur des variables sont calculées avec **PROC FREQ** et sont données en annexe C au tableau C.1.b.

1.3.2 Création de variables dichotomiques

Le deuxième programme, **2.dicho** (annexe B.2), transforme les 18 variables qualitatives suivantes en $q_a = 75$ variables dichotomiques :

- prataps
- apspref
- apsfun
- apscomp
- apsform
- apscorps
- apsrelax
- risq
- mmalaise
- mcompare
- mmoqueri
- pdepasse
- douleur
- genedire
- nouritur
- santeapp
- ecoutcor
- mode

Pour chacune de ces variables ci-dessus, nous avons construit autant de variables dichotomiques qu'elle possède de niveaux. Donc, le nombre de variables dichotomiques n'est pas le même pour chaque variable.

À la fin de ce programme, j'ai donc $p + q_a = 68 + 75 = 143$ variables. Soient, les 68 variables originales ainsi que les 75 nouvelles variables dichotomiques.

Les 8 premières lignes des 18 variables souches et des 75 variables dichotomiques sont présentées au tableau C.2 en annexe C à titre d'exemple.

1.3.3 Élimination des données manquantes

Le troisième programme, **3.nonmiss** (annexe B.3), consiste à éliminer les individus avec des données manquantes.

Si SAS lit un point « . » au lieu d'une valeur numérique pour une variable alors, l'observation est éliminée. Il m'est donc resté 470 filles et 403 garçons après l'élimination des données manquantes.

Le tableau C.3.b en annexe C présente les résultats des 17 premiers individus des 47 variables dichotomiques conservées à cette étape.

1.3.4 Réduction de niveaux

Le quatrième programme, **4.quarantecinq** (annexe B.4), consiste à amalgamer des niveaux indicateurs de l'importance dans le choix de réponse pour les variables APSFUN, APSFORM, APSRELAX, MMALAISE ET MMOQUERI. En effet, nous avons jumelé les niveaux supérieurs ou jumelés les niveaux inférieurs pour ces variables. Par exemple, si deux individus avaient répondu, à la question sur l'importance du plaisir lors de la pratique des activités sportives (APSFUN), important et très important respectivement alors, maintenant ils seront considéré comme ayant la même cote sur cette variable. Même principe pour ceux dont on a jumelé les niveaux inférieurs.

Ces nouvelles variables portent les indices «45» ou «12» dépendamment si on a voulu regrouper les réponses «4 et 5» ou «1 et 2». La formation de ces nouvelles variables nous a permis de mieux visualiser le comportement des variables concernées en aillant formé des catégories de fort ou faible pour chacune. Donc, à la fin de ce programme on a toujours $n_f = 470$ filles et $n_g = 403$ garçons.

Remarquons qu'il y a 5 nouvelles variables amalgamées chez les garçons et 3

chez les filles. Donc, il nous reste chez les garçons $p_g = 47$ variables dichotomiques, les 5 nouvelles variables avec amalgames de niveaux, l'indicatrice mclus (l'identificateur numérique de chaque garçon, entre 1 et 1122) et le niveau scolaire de l'individu.

Chez les filles, les $p_f = 47$ variables dichotomiques, les 3 nouvelles variables avec amalgames de niveaux, l'indicatrice fclus (l'identificateur numérique de chaque fille, entre 1 et 1122) et le niveau de scolarité pour chacune d'elles.

Notons que les tableaux C.4.a et C.4.b présentent les différents amalgames qui ont été faits pour les filles et pour les garçons respectivement.

1.3.5 Création de la variable **summix**

Dans le sixième programme, **5.summix** (annexe B.5), on y construit une variable appelée **summix** qui est calculée sur les variables dichotomiques suivantes : mmalais1, mmalais4, mcompar1, mcompar4, mmoquer1, mmoquer4, pdepase1 et pdepase4.

Remarquons que ces questions portent sur les raisons pour lesquelles un individu aime ou n'aime pas la mixité. De plus, ces variables représentent les niveaux dichotomiques « tout à fait en accord » et « tout à fait en désaccord ». De cette façon, si un individu obtient 0 comme valeur à la nouvelle variable **summix**, c'est qu'il a répondu « plutôt en accord » ou « plutôt en désaccord » aux variables souches suivantes : mmalaise, mcompare, mmoqueri et pdepasse. Donc, cette étape vérifie les réponses des étudiants(es) aux variables mentionnées ci-dessus et fait une distinction s'ils (elles) ont ou non la valeur 0 à la nouvelle variable **summix**.

Notons que la distinction que nous faisons dans ce programme avec la variable **summix** sera considérée dans le calcul des clusters dans les programmes ultérieurs.

Nous avons donc à la fin de ce programme 1 variable de plus pour chaque sexe, soit 53 variables dichotomiques chez les filles et 55 variables chez les gars. Les 18 premiers individus sont présentés à l'annexe C au tableau C.5.

1.3.6 Transformation linéaire

Dans ce cinquième programme SAS, **6.nouvelledonnee** (annexe B.6), on y pondère chaque valeur sur chaque variable pour chacun des individus. En effet, pour chaque étudiant(e) et sur chacune des valeurs de ses variables dichotomiques, une transformation linéaire est effectuée. Par exemple, pour chaque étudiant(e), la valeur obtenue x à la variable `prataps1` est transformée de la façon suivante :

$$x \times (575 - 62) \div 575$$

Notons que la valeur 62 est celle qui varie selon la variable dichotomique.

A la fin de ce programme, nous avons toujours 403 garçons avec 52 variables dichotomiques et 470 filles avec 50 variables dichotomiques.

Remarquons que lorsque nous allons valider les clusters au neuvième programme, nous tracerons les graphiques biplots en utilisant les variables originales (0 ou 1).

Notons que les résultats des 7 premiers individus sont présentés en annexe C au tableau C.6.

1.3.7 Création des variables `fclus` et `mclus`

C'est dans ce septième programme, **7.cluster** (annexe B.7), que l'on calcule les valeurs de `mclus1` à `mclus5` (pour les garçons) et `fclus1` à `fclus4` (pour les filles) selon le regroupement de variables qu'avait décidé l'analyste.

Rappelons que les 47 variables ont été subdivisées en 4 groupes chez les filles et 5 groupes chez les garçons. La variable `fclus1` est une addition de 0 et de 1 de 19 des 47 variables dichotomiques, `fclus2` est une addition de 8 variables parmi les 47 variables dichotomiques, etc (idem chez les garçons).

Remarquons que ces nouvelles variables ont été calculées à partir des variables dichotomiques.

Les résumés des variables composant les regroupements des 47 variables sont présentés aux tableaux G.2 et G.2 de l'annexe G pour les filles et pour les garçons respectivement. De plus, pour faciliter la compréhension et l'analyse, les tableaux G.4 et G.5, de la même annexe, présentent les clusters selon les définitions de variables présentées au tableau G.1.

Le tableau C.7 en annexe C présente les calculs de ces nouvelles variables pour 20 filles.

1.3.8 Ajustement des variables *fclus* et *mclus*

Dans ce huitième programme, **8.clusterdeux** (annexe B.8), nous refaisons le calcul des variables *fclus_i* ou *mclus_j*, où $i = 1, \dots, 4$ et $j = 1, \dots, 5$, présentées dans le programme précédemment dans le cas où un individu avait obtenu la valeur 0 à la variable *summix* calculée au programme 5. En effet, il semble que l'analyste modifie les combinaisons linéaires pour obtenir *fclus_i* et *mclus_j*, pour les individus qui avaient répondu « plutôt en accord » ou « plutôt en désaccord » aux variables souches suivantes : *mmalaise*, *mcompare*, *mmoqueri* et *pdepasse*.

Nous présentons les nouveaux résultats des *fclus_i* au tableau C.8 en annexe C. Remarquons que nous pouvons visualiser les changements qui ont été faits en comparant avec le tableau C.7 de l'annexe C. Par exemple, on remarque que l'individu 5 a comme valeur de *summix* « 0 » alors, on note que sa valeur à la variable *fclus1* a passé de 0.10526 à 0.13333 .

Nous avons donc à la fin de ces deux derniers programmes, chez les filles, les 53 variables du sixième programme et 4 nouvelles qui représentent les clusters *fclus1* à *fclus4*. Donc, un total de 57 variables pour les filles.

Chez les garçons, nous avons à la fin du sixième programme, 55 variables, et dans ces deux derniers programme nous avons créé 5 nouvelles variables qui représentent les *mclus1* à *mclus5*, ce qui nous fait maintenant un total de 60 variables.

1.3.9 Assignment d'un individu à un cluster

C'est dans ce neuvième programme, **9.vmax** (annexe B.9), que nous assignons chaque étudiants(es) à son cluster. Cette neuvième étape observe, pour chaque individu, la valeur maximale parmi les $fclus_i$ ou $mclus_j$ et assigne l'adolescente au cluster i si $fclus_i = \max(fclus_1, fclus_2, fclus_3, fclus_4)$ (assigne l'adolescent au cluster j si $mclus_j = \max(mclus_1, mclus_2, mclus_3, mclus_4, mclus_5)$).

Donc, $fclus$ identifie le cluster d'assignation pour les filles (respectivement $mclus$ pour les garçons).

Nous présentons au tableau C.9.a de l'annexe C l'assignation des 20 premières filles à leur cluster.

Remarquons, qu'avec la commande **PROC MEANS**, nous avons calculé pour les clusters 1, 2, 3 et 4 (respectivement 1, 2, 3, 4 et 5 pour les garçons) les valeurs moyennes des variables $fclus_1, fclus_2, fclus_3$ et $fclus_4$ pour les filles (respectivement $mclus_1, mclus_2, mclus_3, mclus_4$ et $mclus_5$ pour les garçons). Par exemple, au tableau C.9.b de l'annexe C, on remarque que les filles appartenant au cluster 2 ($fclus=1$) ont eu en moyenne une valeur de 0.1262 à la variable $fclus_1$, une moyenne de 0.6223 à la variable $fclus_2$, la valeur moyenne de 0.3936 à la variable $fclus_3$ et 0.1556 à la variable $fclus_4$ (rappelons que chaque $fclus$ et $mclus$ est une variable construite par une addition de 0 et de 1 de certaines des 47 variables dichotomiques donc, il est possible de calculer une moyenne par personne sur ces variables). Effectivement, dans notre exemple, la moyenne la plus élevée est celle de la variable $fclus_2$ car, c'est de cette façon que nous avons attribué les individus à leur cluster (valeur maximal). Reste à savoir si cette valeur maximale est significativement plus élevée. Nous vérifierons dans le onzième programme si ces différences sont significatives à l'aide des tests d'ANOVA. Ces moyennes par cluster pour les filles et les garçons sont présentées aux tableaux C.9.b et C.9.c respectivement.

De plus, avec la commande **PROC FREQ**, on peut vérifier si pour chaque regroupement de variables, il y a une différence significative entre les niveaux secondaires

et la valeur maximale, c'est-à-dire si le fait d'être à un niveau secondaire ou un autre influencera l'attribution à un cluster. Donc, le test d'hypothèse porte sur l'hypothèse nul qu'il n'y a pas de relation entre le niveau secondaire et l'attribution au fclus ou mclus. Ce qui signifierait que le cluster d'un(e) étudiant(e) ne dépend pas de son niveau secondaire.

Nous remarquons, dans les tableaux croisés présentés aux tableaux C.9.d et C.9.e de l'annexe C, que les deux p-valeurs, filles et garçons respectivement 0.3656 et 0.5656, associées à la statistique χ^2 sont supérieures à $\alpha = 0.05$, ce qui signifie qu'il n'y a pas de relation globale significative entre le niveau de scolarité et l'appartenance à un cluster. Donc, on peut maintenant considérer tous les étudiants ensemble sans avoir à traiter les niveaux secondaire indépendamment pour l'analyse biplot, comme ce sera le cas pour le sexe de l'étudiant(e). En effet, il y a une différence significative entre les sexes, c'est pourquoi on a eu à traiter les filles et les garçons séparément.

1.3.10 Construction du Biplot

C'est dans ce dixième programme, **10.biplot** (annexe D), que nous construisons le programme permettant de visualiser les clusters et individus sur un espace à deux dimensions. Par contre, nous n'en ferons pas discussion dans cette section mais, nous y consacrerons plutôt deux chapitres, soient la théorie présentée au chapitre II et l'application aux données sur la mixité au chapitre III.

1.3.11 ANOVA

Dans ce dernier programme, **11.anova** (annexe B.10), nous testons l'hypothèse que dans chaque regroupement de variables, la valeur maximale qui constitue ce cluster, n'est significativement pas plus élevée que les autres valeurs des fclus ou mclus. Ce qui nous informerait que la raison d'être du cluster n'est pas valide car, sa valeur n'est pas significativement supérieure.

À l'aide des tableaux C.10.1 et C.10.2 en annexe C, on remarque que, autant chez

les filles que chez les gars, les p-valeurs sont significatives (toutes égales à 0.001) donc, toutes les valeurs maximales sont significativement plus grande que les autres. Donc, il ne semblerait pas y avoir de chance de croisement entre les clusters.

Maintenant que nous avons identifié les clusters trouvés antérieurement, nous allons proposer notre choix de méthodes que nous utiliserons pour poursuivre cette analyse et nous ferons l'exposé dans les chapitre II et III.

1.3.12 L'analyse des clusters

L'analyse de vérification de clusters est maintenant de mise. Les méthodes que nous allons présenter (aux fins de cette vérification) visent à visualiser l'ensemble de données sans privilégier une variable plutôt qu'une autre. Les méthodes les plus utilisées dans le traitement des enquêtes sont :

- L'Analyse en Composantes Principales (ACP)
- Le biplot sur ACP
- L'Analyse des Correspondances Multiples (ACM)
- Le biplot sur ACM

Le choix de l'une ou de l'autre de ces méthodes dépend des objectifs poursuivis et du type de données à analyser. Nous verrons :

- l'analyse en composantes et le biplot classique au chapitre II
- l'analyse en correspondance multiple et son biplot au chapitre III

CHAPITRE II

ANALYSE EN COMPOSANTES PRINCIPALES ET BILOT

Remarquons que les applications du biplot sont nombreuses et diversifiées. Pour chacune des méthodologies, comme l'analyse en composantes principales, le positionnement multidimensionnel (multidimensional scaling), les modèles biadditifs, l'analyse des correspondances multiples, on peut tracer un graphique biplot. Pour chacune de ces analyses, les mathématiques qui mènent au graphique biplot diffèrent. Dans ce deuxième chapitre, nous allons faire un survol de la notion importante du *Biplot* associé à l'*Analyse en Composante Principale (ACP)*. Celle-ci est décrite sans le détail des preuves mathématiques. Au chapitre 3, nous verrons avec le détail mathématique l'*Analyse des Correspondances Multiples (ACM)* et le *Biplot* qui lui est associé.

2.1 Analyse en Composantes Principales (ACP)

Dans cette section, nous allons donner un bref aperçu de l'ACP et du biplot qui lui est associé. Si l'on dispose d'un vecteur-observation $\mathbf{x}'_i = (x_{i1}, \dots, x_{ip})$ pour chaque individu $i = 1, \dots, n$ sur chaque variable p , alors on peut représenter chaque vecteur \mathbf{x}'_i par un point P_i dans l'espace p dimensionnel \mathbb{R}^p . On obtient ainsi un nuage de points, dans un espace de dimension p qui est en général > 3 donc, qui est impossible à visualiser. L'ACP va nous donner un sous-espace de dimension k où $1 \leq k \leq p$, tel que les projections des points P_i dans ce sous-espace retiennent le plus d'information possible. Le nuage de points ainsi projeté sera le plus dispersé possible c'est-à-dire, se rapprochera le plus possible de la dispersion du nuage de points $P_i \in \mathbb{R}^p$. Ce qui entraîne que les

distances entre les points sont les mieux préservées. L'ACP permet ainsi de réduire l'espace de représentation des individus \mathbb{R}^p à un sous-espace \mathbb{R}^k , $1 \leq k \leq p$.

On commence par translater les données pour que le centre du nuage de points soit à l'origine. Ensuite, une rotation des axes est effectuée de façon à ce que l'écart-type des coordonnées des points P_i sur le premier axe soit le plus grand possible, l'écart-type des coordonnées des points P_i sur le deuxième axe soit le deuxième plus grand possible, l'écart-type des coordonnées des points P_i sur le troisième axe soit le troisième plus grand possible, etc. La rotation des axes est celle qui est trouvée par la diagonalisation de la matrice de variance-covariance. Les vecteurs propres de la matrice de variance-covariance correspondent à la direction des axes obtenus après rotation. Le vecteur propre associé à la valeur propre la plus grande correspond au premier axe, le vecteur propre associé à la deuxième plus grande valeur propre correspond au deuxième axe, etc. Les coordonnées des points P_i sur le premier axe correspondent à la projection des points P_i sur ce premier axe : le nuage de points $P_i \in \mathbb{R}^p$ est ainsi projeté dans un sous-espace de dimension $k = 1$. Les coordonnées des points P_i sur les k premiers axes correspondent aux projections des points P_i sur les k premiers axes : le nuage de points $P_i \in \mathbb{R}^p$ est ainsi projeté dans un sous-espace de dimension k . Le nuage correspondant à ce sous-espace est la meilleure approximation au nuage de points $P_i \in \mathbb{R}^p$ au sens des moindres carrées.

Pour choisir la dimension de ce sous-espace, on porte en ordonnée la valeur de la valeur propre et en abscisse le rang de l'axe (1er, 2e, 3e, etc). On obtient une courbe ; si celle-ci décroît lentement pour le rang $k = d$, alors on conclut que l'on a obtenu toute l'information voulue pour le nuage de points $P_i \in \mathbb{R}^p$, $k = d$.

L'ACP s'applique lorsque les variables sont numériques. Si $k = 2$, elle permet de positionner les individus dans un plan à deux dimensions.

La technique du biplot permet de représenter les p variables originales dans ce sous-espace à k dimensions. Le biplot permet ainsi d'identifier les valeurs observées des variables pour chaque individu. On peut alors caractériser, à l'aide des variables, les

« patterns » observés chez les individus. Ainsi, si l'on souhaite analyser des points de vente en fonction de différents critères numériques (surface, personnel, nombre de pièces vendues...), le biplot permet d'obtenir une cartographie intéressante, qui regroupe les points de vente selon tous les critères retenus et qui permet ainsi de les catégoriser.

Si les variables sont qualitatives ordinales, c'est-à-dire les valeurs peuvent être ordonnées entre elles (échelles, fréquences...) alors, celles-ci peuvent être recodifiées afin de faire une ACP. Cette recodification doit être généralement préparée à l'avance.

Si les variables sont qualitatives, nous devons utiliser l'analyse des correspondances multiples. Nous présenterons dans le chapitre III, les idées de J.C Gower sur la façon qu'il a adapté l'ACP et son biplot à l'Analyse des Correspondances Multiples et le biplot pour un tel ensemble de données.

Remarquons que les axes du graphique donnés par l'ACP ne correspondent généralement pas à l'une ou l'autre des variables mais à une combinaison optimale de plusieurs variables (ex : revenu et niveau d'études peuvent participer ensemble à la formation d'un axe). Dans cette prochaine section, nous démontrerons comment définir les axes biplot qui représentent les variables originales.

2.2 Biplot

Le Biplot est la technique qui permet de tracer les p axes (que représentent les variables) dans un sous-espace de dimension moindre, $k \leq p$.

Soit $\mathbf{X} = (x_{ij})$ une matrice $n \times p$ de rang $p > n$. La valeur x_{ij} est la valeur observée sur la $j^{\text{ième}}$ variable ($j = 1, \dots, p$) pour l'individu i ($i = 1, \dots, n$). Nous allons démontrer qu'il est toujours possible d'associer un vecteur $\nu_i \in \mathbb{R}^k$ à la $i^{\text{ième}}$ ligne de \mathbf{X} et un vecteur $\mathbf{b}_j \in \mathbb{R}^k$ à sa $j^{\text{ième}}$ colonne de telle sorte que $x_{ij} = \nu_i' \mathbf{b}_j$. La représentation graphique des vecteurs ν_i et \mathbf{b}_j dans un espace \mathbb{R}^k donne la meilleure approximation à toute l'information sur la matrice \mathbf{X} au sens des moindres carrés. On veut évidemment projeter ces vecteurs $\nu_1, \dots, \nu_n, \mathbf{b}_1, \dots, \mathbf{b}_p$ dans un sous-ensemble de dimension $k < p$; si

$k = 2$, alors on peut facilement visualiser cette information.

Pour bien comprendre la construction de la représentation biplot, nous utiliserons la définition de décomposition en valeurs singulières pour identifier les vecteurs propres.

Décomposition en valeurs singulières

La décomposition en valeurs singulières de la matrice \mathbf{X} consiste à exprimer cette matrice par le produit de trois matrices de la façon suivante :

$$\mathbf{X} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}'$$

$n \times p$ $n \times p$ $p \times p$ $p \times p$

où

- $\mathbf{U}'\mathbf{U} = \mathbf{I}_p$ et $\mathbf{V}'\mathbf{V} = \mathbf{I}_p$, où \mathbf{I}_p est la matrice identité $p \times p$.
- $\mathbf{\Lambda}$ est une matrice diagonale à termes positifs $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$.

Si l'on suppose que le rang de \mathbf{X} est p alors, $\lambda_i > 0$, $i = 1, \dots, p$.

Notons $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ les colonnes de \mathbf{U} et $\{\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_p\}$ les colonnes de \mathbf{V} . Ces deux ensembles de vecteurs forment respectivement une base orthogonale de l'espace engendré par les colonnes de \mathbf{X} et par les lignes de \mathbf{X} .

Construction du biplot

Cette factorisation n'est pas unique mais, elle nous permet d'écrire \mathbf{X} comme combinaison linéaire des matrices \mathbf{U} , $\mathbf{\Lambda}$ et \mathbf{V}' . Voyons en détail cette décomposition.

$$\begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_p \end{pmatrix}_{(n \times p)} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_p \end{pmatrix}_{(p \times p)} \begin{pmatrix} \boldsymbol{\nu}'_1 \\ \vdots \\ \boldsymbol{\nu}'_p \end{pmatrix}_{(p \times p)}$$

ou

$$\begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}_{(n \times p)} \mathbf{I}_{p \times p} = \begin{pmatrix} \lambda_1 u_{11} & \lambda_2 u_{21} & \cdots & \lambda_p u_{p1} \\ \lambda_1 u_{12} & \lambda_2 u_{22} & \cdots & \lambda_p u_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 u_{1n} & \lambda_2 u_{2n} & \cdots & \lambda_p u_{pn} \end{pmatrix}_{(n \times p)} \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{p1} & v_{p2} & \cdots & v_{pp} \end{pmatrix}_{(p \times p)}$$

où \mathbf{I} est la matrice identité

Nous voyons ici que les vecteurs colonnes de la matrice \mathbf{V} (c'est-à-dire les vecteurs lignes de la matrice \mathbf{V}') représentent une base pour les lignes de la matrice \mathbf{X} . Considérons maintenant les vecteurs colonnes de la matrice \mathbf{V}'

$$\mathbf{V}'_{(p \times p)} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{p1} & v_{p2} & \cdots & v_{pp} \end{pmatrix}_{(p \times p)} = \left(\beta_1 \quad \beta_2 \quad \cdots \quad \beta_p \right)$$

où chaque β_j , où $j = 1, \dots, p$, est la j^e colonne du vecteur \mathbf{V}' , c'est-à-dire

$$\beta_j = \begin{pmatrix} v_{1j} \\ v_{2j} \\ \vdots \\ v_{pj} \end{pmatrix}_{(p \times 1)}$$

Nous pouvons prouver que les vecteurs β_j , $j = 1, \dots, p$, peuvent représenter les p directions des p axes biplot. En effet, ces axes biplots peuvent servir pour interpoler un point P dans le graphique ; ils peuvent servir aussi pour prédire les valeurs des variables qu'un point P du graphique aurait.

Nous allons maintenant considérer seulement les deux premières composantes principales qui sont données par les deux premiers vecteurs propres ν_1 et ν_2 . On aura que la meilleure approximation de \mathbf{X} au sens des moindres carrés est :

$$\mathbf{X}_{(n \times p)} \cong \begin{pmatrix} \lambda_1 u_{11} & \lambda_2 u_{21} \\ \vdots & \vdots \\ \lambda_1 u_{1n} & \lambda_2 u_{2n} \end{pmatrix}_{(n \times 2)} \begin{pmatrix} \nu_{11} & \cdots & \nu_{1p} \\ \nu_{21} & \cdots & \nu_{2p} \end{pmatrix}_{(2 \times p)} = \mathbf{A}_{(n \times 2)} \begin{pmatrix} \beta_1 & \cdots & \beta_1 \\ [2] & & [2] \end{pmatrix}_{(2 \times p)}$$

où les n vecteurs-lignes de la matrice \mathbf{A} sont les coordonnées des n points $P_i^* \in \mathbb{R}^2$;
 $\beta_j = \begin{pmatrix} \nu_{1j} \\ \nu_{2j} \end{pmatrix}_{[2]}$ est la direction du $j^{\text{ème}}$ axe biplot $\in \mathbb{R}^2$ (β_j est le vecteur des 2 premiers éléments de β_j).

Pour simplifier la notation, nous allons noter β_j par β_j .

Nous allons maintenant donner un sens géométrique aux vecteurs β_j . Soit $P_i \in \mathbb{R}^p$ le point dont le vecteur des coordonnées est \mathbf{x}_i et qui représente l'individu i . Et soit P_i^* le point qui représente l'individu dans \mathbb{R}^2 . On a alors que la coordonnée de P_i^* est $(\lambda_1 u_{1i}, \lambda_2 u_{2i})$ que nous noterons \mathbf{x}_i^* . On a donc

$$x_{ij} \cong x_{ij}^* = \mathbf{x}_i^* \beta_j$$

La projection de P_i^* sur un vecteur de direction β_j est :

$$pr(P_i^* \text{ sur } \vec{O\beta}_j) = \frac{\mathbf{x}_i^* \beta_j}{\|\beta_j\|^2}$$

aussi, l'expression permet d'écrire :

$$x_{ij} \cong x_{ij}^* = \frac{\mathbf{x}_i^* \beta_j}{\|\beta_j\|^2} \|\beta_j\|^2 = pr(\mathbf{x}_i^* \text{ sur } \beta_j) \|\beta_j\|^2$$

On a donc obtenu :

$$x_{ij} \cong \|\beta_j\|^2 pr(P_i^* \text{ sur } \vec{O\beta}_j)$$

La projection de P_i^* sur $\vec{O\beta}_j$ donne la valeur approximative à x_{ij} à un facteur $\|\beta_j\|^2$ constant. C'est-à-dire, une approximation à la valeur observée sur la variable j pour l'individu i . Les droites de directions β_j sont donc appelées *axe-biplot*; celles-ci représentent les axes des variables $j = 1, \dots, p$.

Alors, si l'on trace toutes les droites de directions $\beta_j, (j = 1, \dots, p)$ et que l'on indique sur ces droites une échelle de valeurs appropriées, alors en y projetant le point P_i^* sur un axe β_j , on y lit au pied de la projection la valeur $x_{ij}^* \cong x_{ij}$.

Sommes toutes, la projection orthogonale du point P_i^* sur chacun des axes biplot permet ainsi de retrouver approximativement les valeurs $(x_{i1}, x_{i2}, \dots, x_{ip})$ correspondant à P_i^* . Donc, on a réussi à prédire les valeurs $(x_{i1}, x_{i2}, \dots, x_{ip})$ à partir des projections de P_i^* sur chaque axes biplot β_j .

On a affirmé plus haut que lorsque $k = 2$, alors on a la meilleure approximation au sens des moindres carrés de la matrice $X_{n \times p}$ de rang p par une matrice de rang 2. Ce qui nous permet de représenter cette matrice dans \mathbb{R}^2 .

Et de cette façon, nous avons construit nos axes biplot et représenté nos individus. Ces axes sont possibles dans le cas où les variables sont continues, c'est-à-dire quantitatives. Dans cette analyse, madame Laberge a construit son questionnaire sur des variables qualitatives. Donc, on ne pourra pas représenter nos variables par des axes continues dans le plan, il faudra trouver le moyen qu'elles soient représentées par un seul point dans le plan. Dans son livre sur le biplot, Gower appelle ces points des CLP « Category Level Point » (on en fera la définition au chapitre IV). Et nous démontrerons que l'on peut se servir de ces points pour prédire les niveaux des variables qualitatives comme dans le biplot construit avec l'ACP.

CHAPITRE III

BIPLOT ET ANALYSE DES CORRESPONDANCES MULTIPLES

Dans le cadre de notre travail, on ne peut donc pas utiliser la technique de l'ACP et du biplot pour variables quantitatives car, les variables sont qualitatives. Nous allons donc utiliser des techniques spécifiques pour variables qualitatives. Une de ces techniques se nomme l'Analyse des Correspondances Multiples (ACM). Le but de l'ACM est le même que celui de l'ACP. Il s'agit de représenter graphiquement les individus par des points dans un sous-espace de manière à ce que le nuage de points ressemble le plus au nuage de points de l'espace original. On verra que l'on peut procéder exactement comme pour l'ACP, à ceci près qu'on ne mesure pas les distances entre les individus avec la métrique canonique (euclidienne), mais à l'aide de la «distance du Khi-deux». De plus, le biplot permet de représenter les niveaux de chaque variable qualitative par des points dans le même sous-espace. Nous allons donc vous exposer un bref aperçu de la théorie de Gower sur la construction du biplot pour données qualitatives.

3.1 Analyse en Composante Multiple (ACM)

L'Analyse en Correspondance Multiple (ACM) est une analyse sur des observations faites sur p variables catégoriques pour n individus. La technique de l'ACM permet de représenter graphiquement les vecteurs-observations $\mathbf{x}_i \in \mathbb{R}^p$ par des points P_i^* dans un plan contenu dans un sous-espace $\mathbb{R}^k \subseteq \mathbb{R}^p$ de telle sorte que si deux vecteurs \mathbf{x}_i et \mathbf{x}_j sont semblables (dissemblables) alors, les points P_i^* et P_j^* devraient être près (loin) l'un de l'autre. La métrique qui mesure la distance entre les points P_i^* et P_j^* est la distance

canonique euclidienne qui représentera la distance du Khi-deux dont on discutera dans ce chapitre. Si la distance du Khi-deux est petite (grande) alors, la distance euclidienne entre les points P_i^* et P_j^* devrait être petite (grande). La technique de l'ACM ressemble donc à celle de l'ACP en ce sens que le but est de représenter les individus par des points dans un sous-espace $\mathbb{R}^k \subseteq \mathbb{R}^p$.

John Gower a développé les idées du biplot sur l'ACM dans sa monographie. Chaque niveau de chaque variable qualitative est représenté par un point dans le même sous-espace que celui des individus. De plus, une région est assignée à chaque niveau de chaque variable. Si des points se trouvent dans une région alors, ces points devraient représenter des individus dont le choix de réponse est ce niveau.

On peut donc facilement utiliser le graphique afin de juger de la ressemblance entre individus, de visualiser les regroupements entre individus et de juger si les « patterns » se profilent. Le biplot permet de caractériser aisément à l'aide des variables originales les ressemblances et « patterns » observés sur le graphique. Son but est aussi de dégager des dimensions cachées contenues dans les réponses aux variables sélectionnées, pour faciliter l'interprétation de la matrice de données.

Nous notons dans la prochaine section, que l'algèbre linéaire qui sert à la construction des biplots avec l'ACM est différente de celle du biplot pour l'ACP. Et pour ce faire, nous allons vous proposer un exemple tout au long de la section afin d'illustrer la théorie.

3.1.1 Matrice indicatrice G

Considérons une matrice $n \times p$. Par exemple, $n = 6$ et $p = 3$ variables catégoriques. La première variable représente le sexe de la personne, la deuxième est la couleur de cheveux et la troisième est sa situation financière. Voir le tableau 3.1

Donc, nous notons que le sexe a deux niveaux et que la couleur de cheveux et la situation financière en ont quatre chacun. Maintenant, pour chacun de ces niveaux, nous allons nous définir des variables binaires (dichotomiques). Donc, deux variables

Tableau 3.1 Ensemble de données fictif

variables			
individu	sexe de l'individu	couleur de cheveux	situation financière
1	masculin	brun	riche
2	masculin	roux	pauvre
3	féminin	rousse	confortable
4	masculin	noir	faillite
5	féminin	blonde	pauvre
6	masculin	blond	pauvre

relatives au sexe, quatre à la couleur et quatre à la finance. Chacune de ces nouvelles variables correspond à un niveau d'une catégorie et prendra la valeur 1 si l'individu a ce score exactement et 0 sinon. On voit au tableau 3.2 la nouvelle matrice qui ne comprend maintenant que des valeurs numériques (0 ou 1).

Tableau 3.2 Tableau des variables qualitatives dichotomiques

variables			
individu	sexe de l'individu	couleur de cheveux	situation financière
1	10	0010	1000
2	10	1000	0010
3	01	1000	0100
4	10	0001	0001
5	01	0100	0010
6	10	0100	0010

La matrice qui correspond au tableau 3.2 est appelée matrice indicatrice \mathbf{G} . De plus, nous noterons \mathbf{G}_k la matrice indicatrice de la variable catégorique k . Alors, nous avons $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$. Soit L_k le nombre de niveau de la variable k et $L = L_1 + \dots + L_p$.

3.1.2 Propriétés de la matrice \mathbf{G}

Nous avons que :

$$\underset{(n \times L_k)}{\mathbf{G}_k} \underset{(L_k \times 1)}{\mathbf{1}_k} = \underset{(n \times 1)}{\mathbf{1}} \quad (3.1)$$

où \mathbf{G}_k (de dimension $(n \times L_k)$) est la matrice indicatrice pour la k^e variable-catégorique, $\mathbf{1}'_k = (1, \dots, 1)$ de dimension $(1 \times L_k)$ et $\mathbf{1}' = (1, \dots, 1)$ de dimension $(1 \times n)$.

Nous utiliserons fréquemment le vecteur $\mathbf{1}$ dont tous les éléments sont de valeur 1 ; la dimension de ce vecteur sera indiqué comme dans l'expression (3.1).

Exemple : 1^e variable qualitative $k = 1, L_k = 2$

$$\mathbf{G}_k \mathbf{1}_k = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Nous avons aussi :

$$\underset{n \times L}{\mathbf{G}} \underset{L \times 1}{\mathbf{1}} = p \underset{n \times 1}{\mathbf{1}} \quad (3.2)$$

car

$$\mathbf{G}\mathbf{1} = (\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_p) \underset{L \times 1}{\mathbf{1}}$$

$$\begin{aligned}
 \mathbf{G}\mathbf{1} &= (\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_p) \begin{pmatrix} \mathbf{1}_{L_1 \times 1} \\ \mathbf{1}_{L_2 \times 1} \\ \mathbf{1}_{L_3 \times 1} \\ \vdots \\ \mathbf{1}_{L_p \times 1} \end{pmatrix} \\
 &= \mathbf{G}_1 \mathbf{1}_{L_1 \times 1} + \mathbf{G}_2 \mathbf{1}_{L_2 \times 1} + \dots + \mathbf{G}_p \mathbf{1}_{L_p \times 1} \\
 &= \mathbf{1}_{n \times 1} + \mathbf{1}_{n \times 1} + \dots + \mathbf{1}_{n \times 1} = p \mathbf{1}_{n \times 1}
 \end{aligned}$$

Posons \mathbf{L}_k la matrice diagonale dont la diagonale contient les fréquences pour chaque niveau de la k^e variable-catégorique.

Posons \mathbf{L} la matrice diagonale dont la diagonale contient les fréquences pour toutes les variables.

On peut écrire :

$$diag(\mathbf{1}'\mathbf{G}_k) = \mathbf{L}_k \quad (3.3)$$

$$diag(\mathbf{1}'\mathbf{G}) = \mathbf{L}$$

Exemple $k = 1$,

on a :

$$\left(\begin{array}{cccc} 1 & 1 & 1 & \\ & 1 & 1 & 1 \end{array} \right) \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 2 \end{pmatrix}$$

donc :

$$\text{diag}(\mathbf{1}'\mathbf{G}_1) = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} = \mathbf{L}_1$$

Nous avons aussi que

$$\text{diag}(\mathbf{1}'\mathbf{G}) = \text{diag}(\mathbf{1}'(\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_p)) = \mathbf{L}$$

On a donc,

$$\begin{aligned} \mathbf{1}'_{1 \times n} \mathbf{G}_k &= \mathbf{1}'_{1 \times L_k} \mathbf{L}_k_{L_k \times L_k} \\ \mathbf{1}'_{1 \times n} \mathbf{G} &= \mathbf{1}'_{1 \times L} \mathbf{L}_{L \times L} \end{aligned} \quad (3.4)$$

Exemple avec $k = 1$

$$\mathbf{1}'_{1 \times L_1} \mathbf{L}_1_{L_1 \times L_1} = \mathbf{1}' \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 2 \end{pmatrix} = \mathbf{1}'_{1 \times n} \mathbf{G}_1$$

et on a

$$\begin{aligned} \mathbf{1}'\mathbf{G} &= \mathbf{1}'(\mathbf{G}_1, \dots, \mathbf{G}_p) = (\mathbf{1}'\mathbf{G}_1, \dots, \mathbf{1}'\mathbf{G}_p) \\ &= (\mathbf{1}'\mathbf{L}_1, \dots, \mathbf{1}'\mathbf{L}_p) = \mathbf{1}'\mathbf{L} \end{aligned}$$

Avec les expressions (3.1) et (3.2) on obtient :

$$\mathbf{1}'\mathbf{G}_k\mathbf{1}_k = \mathbf{1}'\mathbf{1} = n \quad (3.5)$$

$$\mathbf{1}'\mathbf{G}\mathbf{1} = p\mathbf{1}'\mathbf{1} = np$$

Les expressions (3.4) et (3.5) donnent :

$$\mathbf{1}'\mathbf{G}_k\mathbf{1}_k = \mathbf{1}'\mathbf{L}_n\mathbf{1} = n \quad (3.6)$$

$$\mathbf{1}'\mathbf{G}\mathbf{1} = \mathbf{1}'\mathbf{L}\mathbf{1} = np$$

Et prenons encore une fois $k = 1$ afin d'illustrer la première équation de (3.6) :

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 6$$

3.1.3 Analyse des Correspondances Multiples

Avec ces quelques propriétés, nous allons démontrer que l'ACM peut être considéré comme une simple ACP sur la matrice $\mathbf{X} = p^{-1/2}\mathbf{G}\mathbf{L}^{-1/2}$.

Soit l'individu i , avec coordonnée $x_i = \frac{1}{\sqrt{p}}\mathbf{g}_i\mathbf{L}^{-1/2}$ et l'individu j avec coordonnée $x_j = \frac{1}{\sqrt{p}}\mathbf{g}_j\mathbf{L}^{-1/2}$.

Où \mathbf{g}_i est la i^e ligne de la matrice \mathbf{G} (remarquons que x_i est un vecteur ligne).

On obtient la distance suivante :

$$(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)' = p^{-1/2}(\mathbf{g}_i \mathbf{L}^{-1/2} - \mathbf{g}_j \mathbf{L}^{-1/2})p^{-1/2}(\mathbf{g}_i \mathbf{L}^{-1/2} - \mathbf{g}_j \mathbf{L}^{-1/2})'$$

$$(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)' = p^{-1}(\mathbf{g}_i - \mathbf{g}_j)\mathbf{L}^{-1}(\mathbf{g}_i - \mathbf{g}_j)'$$

(\mathbf{g}_i est la i^e ligne de \mathbf{G} , \mathbf{x}_i la i^e ligne de \mathbf{X}).

On obtient donc que la distance d_{ij} entre les individus i et j satisfait l'équation suivante :

$$\begin{aligned} d_{ij}^2 &= \frac{1}{p}(\mathbf{g}_i - \mathbf{g}_j)\mathbf{L}^{-1}(\mathbf{g}_i - \mathbf{g}_j)' \\ &= \frac{1}{p} \sum_{r=1}^L \frac{1}{L_r} (g_{ir} - g_{jr})^2 \end{aligned} \quad (3.7)$$

où l'indice r parcourt tous les niveaux de toutes les variables.

Lorsque deux individus sont distincts l'un de l'autre uniquement sur la k^e variable, alors ils diffèrent de valeurs sur deux niveaux k_1 et k_2 de cette variable.

Supposons l_{k_1} et l_{k_2} les fréquences d'occurrence de ces deux niveaux alors, on peut écrire :

$$\begin{aligned} d_{ij}^2 &= \frac{1}{p} \left\{ 0 + \dots + 0 + \frac{1}{l_{k_1}} (1 - 0)^2 + \frac{1}{l_{k_2}} (0 - 1)^2 + 0 + \dots + 0 \right\} \\ &= \frac{1}{p} \left\{ 0 + \dots + 0 + \frac{1}{l_{k_1}} + \frac{1}{l_{k_2}} + 0 + \dots + 0 \right\} \end{aligned}$$

Ce qui indique que la contribution de cette variable au carré de la distance est :

$$= \frac{1}{p} \left\{ \frac{1}{l_{k_1}} + \frac{1}{l_{k_2}} \right\} \quad (3.8)$$

Le carré total de la distance est donc la somme des contributions de chaque variable qui sont donc du type de l'expression (3.8).

John Gower fait la remarque que lorsqu'on fait la somme sur toutes les contributions de chaque variable, cette expression devient une forme spéciale de la distance du chi-carré utilisée dans le cadre d'une analyse de correspondance. La distance de Khi-deux s'écrit comme suit :

$$d_{ij}^2 = \sum_{k=1}^c \frac{1}{c_k} \left(\frac{x_{ik}}{r_i} - \frac{x_{jk}}{r_j} \right)^2$$

Dans le cas d'une analyse en correspondance multiple, $x_{ik} = x_{jk} = 1, \forall i, j, k$, les r_i sont les fréquences pour la catégorie i de la variable ligne (ici $r_i = l_i$) et les c_k sont les fréquences de variable colonne. D'où l'expression spéciale présentés ci-haut.

Nous avons donc exprimé le problème de représentation graphique d'observations sur des variables catégoriques comme un problème typique d'une ACP. En effet, nous avons défini une matrice $\mathbf{X} = \frac{1}{\sqrt{p}} \mathbf{G} \mathbf{L}^{-1/2}$, calculé les distances euclidiennes entre deux individus et avons conclu que cette distance est analogue à la distance du Khi-deux. Nous pouvons représenter les vecteurs lignes de la matrice \mathbf{X} par des points P_i dans un espace \mathbb{R}^2 . Et, comme dans l'ACP, nous voulons représenter ces vecteurs lignes par des points P_i^* dans un sous-espace $\mathbb{R}^k \subseteq \mathbb{R}^L$ de telle sorte que le nuage de points P_i^* sera le plus près possible du nuage de points $P_i, i = 1, \dots, n$. De plus, nous avons démontré que le carré de la distance euclidienne entre deux vecteurs lignes de la matrice \mathbf{X} représente la distance du Khi-deux entre ces mêmes vecteurs lignes.

Mais, tout d'abord, nous allons vérifier quelques éléments essentiels afin de pouvoir bien utiliser la théorie présentée au chapitre 2. En effet, nous allons vérifier la nécessité de centrer notre matrice \mathbf{X} comme dans le cas d'une ACP et voir si les axes biplots sont appropriés dans notre cas, compte tenu que nos données ne sont pas continus.

3.1.4 Centrage de \mathbf{X} et $ACM \equiv ACP$ sur \mathbf{X}

Dans le cas de l'ACP, un des principes de bases est le centrage De la matrice de données. On veut que $1'\mathbf{X} = 0$ de telle sorte que le nuage de points soit centré à l'origine.

Alors, on aurait tendance à vouloir centrer aussi la matrice de données qualitatives à l'aide de cette opération :

$$(\mathbf{I} - \mathbf{N})\mathbf{X} = \mathbf{X} - \frac{\mathbf{1}}{n}\mathbf{1}'\mathbf{X}$$

où $\mathbf{N} = \frac{\mathbf{1}}{n}\mathbf{1}'$.

Cependant, nous allons démontrer qu'il est inutile de le faire à cause de résultats algébriques intéressants. Nous allons définir quelques identités sur \mathbf{X} afin d'en faire la démonstration. Nous allons évidemment utiliser les propriétés de \mathbf{G} données à la section 3.1.2.

Nous avons que :

$$\mathbf{X}(\mathbf{L}^{1/2}\mathbf{1}) = p^{-1/2}\mathbf{G}\mathbf{L}^{-1/2}(\mathbf{L}^{1/2}\mathbf{1}) = p^{-1/2}\mathbf{G}\mathbf{1} = p^{-1/2}p\mathbf{1} = p^{1/2}\mathbf{1}$$

$$\mathbf{1}'\mathbf{X} = \mathbf{1}'p^{-1/2}\mathbf{G}\mathbf{L}^{-1/2} = p^{-1/2}\mathbf{1}'\mathbf{G}\mathbf{L}^{-1/2} = p^{-1/2}\mathbf{1}'\mathbf{L}\mathbf{L}^{-1/2} = p^{-1/2}\mathbf{1}'\mathbf{L}^{1/2}$$

On obtient ainsi :

$$\mathbf{X} \frac{(\mathbf{L}^{1/2}\mathbf{1})}{\sqrt{np}} = \frac{p^{1/2}\mathbf{1}}{\sqrt{np}} = \frac{\mathbf{1}}{\sqrt{n}}$$

$$\frac{\mathbf{1}'}{\sqrt{n}}\mathbf{X} = \frac{p^{-1/2}\mathbf{1}'\mathbf{L}^{1/2}}{\sqrt{n}} = \frac{\mathbf{1}'\mathbf{L}^{1/2}}{\sqrt{np}}$$

On remarque que

$$\frac{(\mathbf{L}^{1/2}\mathbf{1})}{\sqrt{np}}$$

et

$$\frac{\mathbf{1}}{\sqrt{n}}$$

sont des vecteurs propres « 2 sided » de norme 1 de la matrice \mathbf{X} . D'après le théorème de Frobenius, il s'en suit qu'ils sont des vecteurs singuliers de \mathbf{X} correspondants à la plus grande valeur singulière $\sigma = 1$ de \mathbf{X} , compte tenu que \mathbf{X} est non négative. Donc, par la décomposition en valeur singulière vue au chapitre 2, section 2.2, sur la construction d'un biplot, on peut définir la matrice \mathbf{X} en fonction de ses valeurs singulières.

$$\mathbf{X} = \sum_{k=1}^L \sigma_k \mathbf{u}_k \mathbf{v}'_k = \sigma_1 \mathbf{u}_1 \mathbf{v}'_1 + \sum_{k=2}^L \sigma_k \mathbf{u}_k \mathbf{v}'_k = (\mathbf{1}) \left(\frac{\mathbf{1}}{\sqrt{np}} \mathbf{L}^{1/2} \mathbf{1} \right) \left(\frac{\mathbf{1}'}{\sqrt{n}} \right) + \sum_{k=2}^L \sigma_k \mathbf{u}_k \mathbf{v}'_k$$

Donc,

$$\mathbf{X} = \frac{\mathbf{1}}{n\sqrt{p}} \mathbf{1}\mathbf{1}'\mathbf{L}^{1/2} + \sum_{k=2}^L \sigma_k \mathbf{u}_k \mathbf{v}'_k$$

Mais,

$$\frac{\mathbf{1}\mathbf{1}'\mathbf{X}}{n} = \frac{\mathbf{1}\mathbf{1}'\mathbf{L}^{1/2}}{n\sqrt{p}}$$

Ainsi la matrice \mathbf{X} est automatiquement centrée avec les premiers vecteurs propres.

En effet,

$$\frac{\mathbf{1}}{n\sqrt{p}} \mathbf{1}\mathbf{1}'\mathbf{L}^{1/2} = \frac{\mathbf{1}}{n} \mathbf{1}\mathbf{1}'\mathbf{X} = \mathbf{N}\mathbf{X}$$

On a donc,

$$\mathbf{X} - \frac{\mathbf{1}\mathbf{1}'\mathbf{X}}{n} = \mathbf{X} - \frac{\mathbf{1}}{n\sqrt{p}} \mathbf{1}\mathbf{1}'\mathbf{L}^{1/2} = \sum_{k=2}^L \sigma_k \mathbf{u}_k \mathbf{v}'_k$$

$$\mathbf{X} = \frac{\mathbf{1}}{n\sqrt{p}} \mathbf{1}\mathbf{1}'\mathbf{L}^{1/2} + \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$$

où

$$\mathbf{1}'\mathbf{U} = 0$$

et

$$\mathbf{1}'\mathbf{L}^{1/2}\mathbf{V} = 0$$

Nous avons donc démontré qu'il suffit de faire la décomposition en valeur singulière de \mathbf{X} et d'utiliser les vecteurs propres v_k , $k = 2, \dots, L$, comme base. Nous projetons les vecteurs observateurs x_i sur cette base et nous obtenons une représentation des individus par des points P_i^* dans le sous-espace \mathbb{R}^r , $r \leq L$.

3.2 Biplot sur ACM

Maintenant que nous avons présenté l'ACM comme un cas particulier de l'ACP, nous voulons aussi représenter les variables catégoriques dans le même sous-espace des individus. Celles-là seront représentées par des points appelés CLP (Category Level Point); évidemment. Dans cette section, nous allons discuter de la construction des point CLP, des utilités possibles du biplot sur ACM ainsi qu'une brève introduction sur la **Matrice de Burt**.

3.2.1 Définition des « *Category Level Point* »

Tout d'abord, définissons ce qu'est un point CLP. Un CLP représentera simplement un des choix de réponse d'une des variables de notre expérience. Donc, dans notre exemple, on avait déjà défini que le sexe a deux niveaux, la couleur des cheveux en a quatre et la situation financière en a quatre aussi. Alors, il y aurait dans cet exemple dix PCN.

Remarquons que le vecteur-observation \mathbf{g}_i sur le i^e individu a été multiplié par $\frac{1}{\sqrt{p}}\mathbf{L}^{-1/2}$.

Remarquons aussi que le point qui représente la première catégorie de la première variable est $(1, 0, 0, \dots, 0)$.

Notons que nous n'avons pas représenté le vecteur \mathbf{g}_i par un point P_i ; c'est plutôt le vecteur $\frac{\mathbf{g}_i}{\sqrt{p}} \mathbf{L}^{-1/2}$ qui a été représenté. Alors, ce ne sera pas le vecteur $(1, 0, \dots, 0)$ qui sera représenté mais, $(1, 0, \dots, 0) \frac{\mathbf{L}^{-1/2}}{\sqrt{p}}$.

Ainsi, la projection de $(1, 0, \dots, 0) \frac{\mathbf{L}^{-1/2}}{\sqrt{p}}$ sur la base \mathbf{V} est $(1, 0, \dots, 0) \frac{\mathbf{L}^{-1/2}}{\sqrt{p}} \mathbf{V}$.

On fait de même pour chaque point qui représente une catégorie pour chaque variable. On obtient les coordonnées de ces points avec l'expression suivante :

$$\mathbf{Z} = \frac{1}{\sqrt{p}} \mathbf{L}^{-1/2} \mathbf{V}$$

où chaque ligne de \mathbf{Z} contient les coordonnées des points qui représentent les niveaux des catégories de chaque variable.

Si on considère les ρ premières colonnes de \mathbf{Z} , on obtient les projections des CLP en ρ dimensions connues dans le cas d'une ACP. Si on prend les deux premières colonnes de \mathbf{V} , par exemple, on obtient les coordonnées des CLP dans l'espace à deux dimensions. De plus, par les propriétés démontrées ci-haut, soient $\mathbf{1}' \mathbf{L}^{1/2} \mathbf{V} = 0$ et $\mathbf{1}' \mathbf{U} = 0$, on a qu'alors, $\mathbf{1}' \mathbf{L} \mathbf{Z} = 0$. Ce qui démontre que la somme des CLP, pondérés par les fréquences ($\mathbf{1}' \mathbf{L}$), est à l'origine. Ceci est analogue au fait que les axes biplots pour l'ACP passent par l'origine.

Donc, c'est fait, on a démontré que dans le cas d'une ACM, nous avons des points qui représentent nos variables et non des axes biplots comme dans le cas d'une ACP, vu au chapitre II. Montrons maintenant les applications du biplot, c'est-à-dire comment on utilise les points CLP.

Interpolation

Comme nous l'avons déjà mentionné, le biplot sert à interpoler et à prédire. Maintenant que nous avons identifié les coordonnées des points CLP dans l'espace à deux dimensions, il est possible pour nous de faire l'interpolation d'un nouveau vecteur \mathbf{g}^* , c'est-à-dire de représenter ce nouveau vecteur par un point P^* dans le graphique déjà

tracé à partir de la matrice de données. Évidemment, on pourrait refaire une analyse en y ajoutant cette nouvelle coordonnée mais, il y a plus rapide. En effet, pour chaque variable où l'individu a une valeur unitaire « 1 », il faudra faire l'addition des vecteurs qui correspondent aux coordonnées des PCN correspondants à ces variables. De cette façon, on trouve le point où l'individu se situe dans l'espace. En effet, nous remarquons que les coordonnées \mathbf{Z}_0 des points représentant les individus sont :

$$\mathbf{Z}_0 = \mathbf{X}\mathbf{V} = \frac{1}{\sqrt{p}}\mathbf{G}\mathbf{L}^{-1/2}\mathbf{V} = \mathbf{G}\mathbf{Z}$$

Les points P_i sont positionnés en additionnant vectoriellement les CLP des niveaux présents pour chaque individu.

Alors, pour interpoler un nouveau individu \mathbf{g}^* , on additionne vectoriellement les niveaux présents ainsi on obtient :

$$\mathbf{z}^* = \mathbf{g}^*\mathbf{Z}$$

Ce n'est qu'une façon d'interpoler, il en existe plusieurs que nous exploiterons pas dans ce le cadre de ce travail. Cependant, laissez-nous vous parler brièvement d'une autre façon d'exploiter la matrice de données \mathbf{X} et d'en élaborer quelques avantages et inconvénients.

3.3 Matrice de Burt

Jusqu'ici, nous avons exploité \mathbf{X} par la décomposition en valeurs singulières, soit $\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}'$. Mais, il est possible de vouloir décomposer \mathbf{X} d'une autre façon afin d'en diminuer la grandeur. Prenons la décomposition spectrale de \mathbf{X} , soit

$$\underset{p \times n}{\mathbf{X}'} \underset{n \times p}{\mathbf{X}} = \mathbf{V}\mathbf{\Sigma}\mathbf{U}'\mathbf{U}\mathbf{\Sigma}\mathbf{V}' = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}'$$

Comme on a souvent beaucoup plus d'individus que de variables, cette matrice devient considérablement plus petite et du coup elle possède quelques avantages face

à la décomposition spectrale. Peut-être un jour je me pencherai plus davantage sur ce sujet !

CHAPITRE IV

APPLICATION DU BIPLLOT ASSOCIÉ À L'ACM

Dans ce chapitre nous allons voir en détail le dixième programme SAS (10.biplot) que nous avons mentionné au chapitre I et qui est présenté en annexe D. Ce programme est celui qui permet de représenter, dans un même plan, les individus par des points-individus et les variables catégoriques par des points CLP. En fait, ce programme incorpore la théorie présentée au chapitre III : **Analyse en composante multiple (ACM) et Biplot sur ACM**

Ce quatrième chapitre est subdivisé en trois sections. La première section nous présente un court résumé du programme et de sa composition. La deuxième section est la présentation des résultats de l'analyse faite sur l'ensemble de données des filles avec le programme SAS. La troisième section est la présentation des résultats des garçons. Dans ces deux dernières sections, on y apporte une critique sur les clusters trouvés préliminairement, nous allons donc faire la validation des clusters obtenus à l'aide graphiques biplots. C'est donc à ce chapitre que nous vérifions si le regroupement de variables qui a conduit à l'établissement des clusters semble adéquat.

Remarquons que le programme Biplot, les graphiques biplots des filles et ceux des garçons sont présentés aux annexes D, E et F respectivement.

Notons que tous les graphiques portent le même titre en haut de page : **BIPLLOT SUR ESPACE DES INDIVIDUS ACM**. La légende de chaque graphique biplot est située au bas de chaque graphique.

De plus, notons que les noms de variables dans les graphiques biplots débutent par les lettres A, B, C, D ou E dépendamment si elles entraînent dans la combinaison des variables définissant les clusters 1, 2, 3, 4 ou 5 respectivement. Cette technique nous a permis de vite repérer les regroupements des variables afin de vérifier visuellement si les combinaisons de variables définissaient correctement les clusters.

Finalement, notons que les individus sont identifiés par le numéro du cluster auquel ils appartiennent, c'est-à-dire de 1 à 4 pour les filles et 1 à 5 pour les garçons. C'est donc de cette façon que nous allons valider visuellement les regroupements d'individus en fonction de leur assignation au cluster. Ça revient à vérifier si les 1 sont regroupés ensemble, les 2 ensemble, etc. De plus, les points CLP débutant par A devraient être près des «1», ceux débutant par B, près des «2», etc.

4.1 Programme SAS sur le Biplot

Le langage de programmation employé est **SAS**. La version originale de ce programme n'incorporait que l'ACP et le biplot qui lui est associé. J'ai modifié cette version afin d'incorporer l'ACM et le biplot qui lui est associé. Ce programme **SAS** comporte plusieurs étapes et il est présenté en entier en annexe. Cependant, voici les grandes lignes de son contenu :

- La **MACRO TRAITEMENT**, effectue les opérations arithmétiques (à l'aide de **PROC IML**) sur la matrice de données, la décomposition en valeur singulière, le calcul des coordonnées des individus ainsi que ceux des directions des axes biplots et des points CLP
- La **MACRO IMPRIMER** est celle qui produit certaines impressions
- La **MACRO BIPLINTR** est celle qui permet de choisir les axes de l'espace individu : $\text{prin1 vs prin2}, \dots, \text{prin1 vs prin}_j$, où $i \neq j = 1, \dots, n$
- L'étape **DATA BPLOT** fait la lecture des données

4.2 Validation des clusters des filles

La première étape de cette section est de vérifier visuellement si les individus se regroupent vraiment en cluster tels que construit par le précédent analyste. Remarquons aussi que pour les graphiques biplots on utilise le 2^e axe et le 3^e axe (prin2 et prin3) car, le 1^{er} axe sert à centrer le nuage de points. La figure E.1 de l'annexe E, représente le biplot des 470 filles et des 47 variables dichotomiques utilisées pour les regrouper en 4 clusters. Chaque adolescente est identifiée par le numéro de son cluster d'appartenance. Chaque variable est représentée par un point CLP et identifiée par le nom de la variable et précédée de la lettre A, B, C ou D selon que cette variable a été utilisé pour la construction du cluster 1 (lettre A), cluster 2 (lettre B), cluster 3 (lettre C) ou cluster 4 (lettre D). On remarque que les points CLP sont sur le pourtour du nuage de points-individus. Ce nuage de points est difficile d'interprétation. Toutefois, les individus qui ressortent du nuage de points, c'est-à-dire sur le pourtour du nuage de points, sont bien caractérisés par les variables qui définissent leur cluster respectif. Ainsi, selon le sens des aiguilles d'une montre et en partant à gauche en haut, on distingue les clusters 4, 2 et 3 ainsi que **certaines variables** dont les noms débutent par D, B et C. On peut déjà se poser des questions, à savoir si les autres variables débutant par D sont nécessaire pour définir le cluster 4 (idem pour les autres variables et autres clusters).

On y remarque difficilement des regroupements de clusters pour les individus au centre du nuage de points Ce qui suggère de faire un «zoom» sur ce graphique afin de mieux visualiser ce qui se passe au centre. Notons que ce «zoom» est présenté au graphique E.2.

Remarquons au graphique E.2 que le nuage de points-individus au centre du graphique est en fait mieux défini qu'à première vue. Effectivement, on y distingue assez bien les clusters.

Notons que tout le centre du graphique E.2 est composé d'individus du cluster 3.

Cependant, dans le cadran 2, on y observe une zone où les adolescentes apparte-

nant aux clusters 3 et 4 y sont mêlées.

De même pour le cadran 1, où se sont les filles des clusters 2 et 3 qui y sont regroupées.

En conclusion, le centre, les cadran 3 et le cadran 4 ne contiennent que des filles du cluster 3. Aussi, à cause des chevauchements d'individus dans les cadran 1 et 2, il serait tentant de croire que quelques variables auraient été mal définies dans les définitions de clusters. Dans les prochaines sections, on vérifiera cette hypothèse.

4.2.1 Positionnement des points CLP

Afin de mieux voir les positions des 47 variables dichotomiques, on a fait l'impression d'un graphique biplot (construit avec les 470 filles et les 47 variables) mais, n'affichant que les 47 variables (voir figure E.3 de l'annexe E).

Remarquons que ce sont surtout les variables du cluster 1 qui semblent attirer les individus à l'origine. Sur la figure E.3 de l'annexe E, on remarque que les variables appartenant au cluster 1 (celles dont le nom débute d'un A) sont toutes au centre du graphique. Donc, effectivement, elles contribuent à l'attraction des individus au centre.

Si \mathbf{z}_0 est le vecteur des coordonnées pour un individu dont le vecteur observation est \mathbf{g}_0 alors,

$$\mathbf{z}_0 = \mathbf{g}_0 \mathbf{Z}_{CLP}$$

C'est-à-dire, les coordonnées de \mathbf{z}_0 sont situées à la somme vectorielle des points \mathbf{Z}_{CLP} pour lesquelles les éléments de \mathbf{g}_0 sont 1 .

Voyons un exemple :

$$\mathbf{g}_0 = (1, 0, 1, 0, 0, 1, \dots)$$

$$\mathbf{Z}_{CLP} = \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_{47} \end{pmatrix}$$

$$\begin{aligned} \mathbf{z}_0 = \mathbf{g}_0 \mathbf{Z}_{CLP} &= (1, 0, 1, 0, 0, 1, \dots) \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_{47} \end{pmatrix} \\ &= \mathbf{z}_1 + \mathbf{z}_3 + \mathbf{z}_6 + \dots \end{aligned}$$

Ainsi, le point P_0 ayant comme vecteur de coordonnées \mathbf{z}_0 est obtenu à l'aide d'une somme vectorielle :

$$\vec{OP} = \vec{oz}_1 + \vec{z}_1 \vec{z}_3 + \vec{z}_3 \vec{z}_6 + \dots$$

Si les coordonnées de \mathbf{g}_0 sont 1 pour les points CLP sur le pourtour et 0 sur les points CLP au centre alors, les coordonnées de \mathbf{z}_0 sont sur le pourtour du nuage de points. Dans l'exemple ci-dessus, les variables qui influencent la coordonnée sont la 1^{iere}, la 3^e et la 6^e. Donc, si \mathbf{z}_1 , \mathbf{z}_3 et \mathbf{z}_6 sont situés sur le pourtour alors, la coordonnée de l'individu sera aussi sur le pourtour. De même si \mathbf{z}_1 , \mathbf{z}_3 et \mathbf{z}_6 sont au centre, l'individu aura sa coordonnée au centre.

Notons que si l'on a

$$\begin{aligned} \mathbf{z}_0 &= \frac{1}{\sqrt{p}} \mathbf{GL}^{-1/2} \mathbf{V} \\ &\cong \mathbf{z}_0^* = (z_{01}^*, z_{02}^*) = \frac{1}{\sqrt{p}} \mathbf{GL}^{-1/2} (\nu_1, \nu_2) \end{aligned}$$

où $\mathbf{z}_0^* \in \mathbb{R}^2$, $\mathbf{z}_0 \in \mathbb{R}^{47}$ et ν_1 , ν_2 sont les 2 premiers vecteurs singuliers de la décomposition en valeur singulière de la matrice \mathbf{X} .

Si ν_{2j} est petit alors, la j^e variable ne compte pas beaucoup dans le calcul de z_{01}^* (idem pour ν_{3j} et z_{02}^*), c'est-à-dire elle ne contient pas beaucoup d'information pour les dimensions prin2 et prin3 (les 2 premiers axes).

On peut donc conclure que les individus positionnés au centre ne nous donnent pas beaucoup d'information sur leur cluster d'appartenance ni sur les variables qui les définissent.

Maintenant, nous présentons l'analyse graphique des 4 clusters dans lesquels seul un cluster est représenté à la fois. (figures E.4 à E.10 de l'annexe E)

4.2.2 Élimination du cluster 1

Remarquons que le cluster 1 est celui dont il y a une seule observation qui se trouve dans le cadran 4. Ce cluster provoque une perte d'information pour tous les graphiques biplots car, ce cluster est constitué de 19 variables : les graphiques biplots subissent ainsi une réduction de l'espace de 47 dimensions à 2 dimensions. Ce qui est inutile, on devrait utiliser $47 - 19 = 28$ variables et on passerait de 28 dimensions à 2 dimensions ce qui est mieux et qui constituerait une moins grande perte d'information.

En particulier, on remarque entre autre que pour appartenir à ce cluster, il y a contradiction sur les réponses de certaines variables. Par exemple, une étudiante ne se sent pas mal à l'aise de **dire son malaise** en cours d'éducation physique afin d'attirer l'attention des autres; et aussi elle **ne dit pas ses malaises** car, elle n'aime pas se plaindre et a peur de la perception des autres face à son malaise. Ce qui, à notre avis, représente une contradiction en soi car, d'un côté on a une étudiante gênée qui **n'aime pas attirer l'attention** et de l'autre celle qui **cherche à attirer l'attention des autres**. De plus, ce cluster est aussi défini avec des variables qui indiquent qu'il y a la plus grande variété de sports pratiqués et qu'aussi ce cluster est défini à l'aide de

variables qui indique peu d'activités sportives et physiques.

Donc, afin de mieux pouvoir observer comment se comportaient les autres clusters, nous avons procédé à l'élimination du cluster 1. Ce qui a libéré le centre des graphiques et permis une meilleure analyse.

Remarquons que nous présentons aux figures E.6, E.8 et E.10 de l'annexe E, les graphiques biplots des clusters 2, 3 et 4 respectivement sans les variables du cluster 1 afin de visualiser s'il y a eu une amélioration depuis l'élimination des variables du cluster 1.

4.2.3 Analyse du cluster 2

Remarquons que le cluster 2 (figure E.6) est quelque peu plus net que le graphique où les variables du cluster 1 y étaient encore présentes (figure E.5). Effectivement, tous les individus se retrouvent maintenant dans le cadran 1.

Notons que ce cluster représente celui dont les étudiantes pratiquent parfois des activités physiques et sportives (*prataps2*) et préfèrent les activités collectives (*apspref1*). Ce sont aussi des filles qui s'arrêtent face à la douleur (*douleur1*), pour qui le plaisir dans les activités est très important (*apsfun1*) et qui n'ont pas d'attraction pour le risque (*risq2*). Donc, sommes toutes se sont des filles faiblement actives et pour qui la santé n'est pas important.

Les points CLP des variables du cluster 2 sont, mise à part *apspref1*, toutes dans les cadrans 1 et 4. La variable *apspref1* est celle qui attribue au cluster son attraction vers les sports collectifs et elle se retrouve dans le cadran 2.

Remarquons que seule la variable *apspref1* se retrouve complètement dans le cadran 2. Toutes les autres variables définissant ce cluster sont toutes du côté des cadrans 1 et 4 se distinguent de l'origine. Notons que cette variable est celle qui attribue à ce cluster sa préférence aux activités collectives. Les activités collectives sont des activités comme : ballon-chasseur, hockey, basketball, etc. Donc, il serait tentant de croire que

cette variable a été mal attribuée car, les filles de ce cluster ne semblent pas être très compétitives. De plus, remarquons qu'elle se trouve parmi les variables appartenant au cluster 4 (soient celles débutant d'un E). Nous verrons plus loin s'il serait possible qu'elle y soit mieux.

4.2.4 Analyse du cluster 3

Pour ce qui est du cluster 3, il est celui dont ses variables sont les plus dispersées autour de l'origine : donc les filles appartenant au cluster 3 sont plutôt placées près de l'origine. On peut donc penser que se sont les filles complètement différentes de celles des clusters 2 et 4.

Il est un cluster très peuplé, soit 341 filles sur un total de 470. Ce sont des filles plus sportives (*prataps3*) que celles du cluster 2 mais, moins que celles du cluster 4 (prochaine section). Elles font attention à leur alimentation (*nourtur3*), optent pour le confort avant la mode dans leur habillement (*mode1*) et choisissent des aliments bons pour la santé à la cafétéria (*santea3*).

Il n'y a aucune contradiction apparente dans la définition de ce cluster. Sommes toutes, ce cluster est celui des filles saines et pratiquant du sport quelquefois par semaine.

Notons que les variables *mode1*, *apsform1*, *nourtur3* et *santeapp3* sont très près les unes des autres dans les graphiques biplots avant et après l'élimination du cluster 1 (figures E.7 et E.8 respectivement de l'annexe E). Elles pourraient être représentées par une seule d'entre elle. Elles sont celles qui distinguent ce groupe par leur souci de la santé et de la forme.

Finalement, une variable semble être à l'écart dans ce cluster. La variable *risque1* est sûrement celle qui attire les filles de ce cluster vers le haut du graphique. Cette variable attribue le goût de risque à ce cluster. Peut-être ici aussi nous faisons face à une variable mal attribuée. Cette dernière cadre aussi avec celle du cluster 4 (prochaine section).

4.2.5 Analyse du cluster 4

Le cluster 4 est celui qui se retrouve en majorité dans le cadran 2. L'élimination du cluster 1 n'a pas donné un effet significatif sur ce cluster. En effet, les deux graphiques du cluster 4 représentant avant et après l'élimination du cluster 1 (respectivement les figures E.9 et E.10) sont sensiblement les mêmes.

Ce cluster est celui des filles actives et compétitives. Toutes les variables du cluster 4 ne semblent pas présentées de contradictions entre elles, elles sont toutes des variables représentant des adjectifs compétitifs. De plus, elles sont toutes près les unes des autres, ce qui est intéressant statistiquement. Cependant, il est un cluster très peu peuplé, soit 34 filles.

Notons que les variables *apspref1* (appartenant au cluster 1) et *risque1* (appartenant au cluster 3) sont des variables qui se retrouvent du côté du cluster 4 avec les variables de ce cluster. Effectivement, le goût du risque cadrerait plus avec cette définition de cluster plus compétitif ainsi que la préférence des activités collectives.

4.2.6 Conclusion chez les filles

Sommes toutes, il y aurait matière à redéfinir les définitions de clusters chez les filles. Premièrement, le cluster 1 est complètement inutile et possède peut-être des variables qui aideraient les autres clusters à mieux se définir.

Deuxièmement, les variables *apspref1* et *risque1* sont des variables qui semblent mieux cadrer avec le cluster 4 autant par leur définition que par leur position dans le plan.

Finalement, il serait possible de faire des regroupement de variables. Effectivement, on remarque que quelques fois, des variables semblent être très près les unes des autres et pourraient être que représentés d'une d'entre elles.

4.3 Validation des clusters des garçons

Voyons maintenant les graphiques biplots pour les garçons. Comme on l'a déjà mentionné, les garçons avaient été distribués selon 5 clusters. Comme première analyse visuelle, nous observons le graphique biplot des 47 variables dichotomiques et les 403 garçons (figure F.1 de l'annexe F). Comme chez les filles, la majorité des adolescents semblent aussi se positionner autour du centre des graphiques. Il est donc un peu difficile de cibler à première vue les regroupements de clusters. Cependant, nous remarquons le même phénomène que chez les filles : quelques garçons ressortent sur le pourtour du nuage de points accompagnés de quelques unes de leurs variables respectives. Remarquons les individus du cluster 3 sur le pourtour dans les cadrans 2 et 3 accompagnés de variables précédées de l'indice C qui les affectent à ce cluster. Même chose dans le cadran 4 pour les garçons du cluster 5 (variables précédées d'un E). Pour ce qui est des individus des clusters 2 et 4, à cette étape il est difficile de visualiser les tendances.

Notons que la figure F.2 est celle où l'on présente le «zoom» des individus afin mieux observer les regroupements d'individus.

Remarquons que ce «zoom» a permis de voir plus clairement le comportement du cluster 2. En effet, au graphique F.1 il était impossible de visualiser où il se situait exactement mais, au graphique F.2 on voit bien qu'il se situe surtout dans le cadran 1 et autour de l'origine. En effet, mis à part quelques individus, les adolescents du cluster 2 sont assez bien regroupés.

De plus, les adolescents du cluster 3 sont aussi assez bien définis sur le pourtour des cadrans 2 et 3.

Aussi, on voit nettement que le cluster 5 est situé sur le pourtour du cadran 4.

Finalement, on observe que le centre du graphique contient des individus appartenant aux clusters 2 et 3 mélangés. Donc, il est possible de croire que ce chevauchement de garçons serait causé par certaines variables qui auraient été mal attribuées lors des définitions de clusters du précédent analyste.

4.3.1 Positionnement des points CLP

Comme chez les filles et afin de mieux voir les positions des 47 variables dichotomiques, on a fait l'impression d'un graphique biplot (construit avec les 403 garçons et les 47 variables dichotomiques) mais, n'affichant que les 47 variables (voir figure F.3 de l'annexe F).

Notons que nous cherchons à identifier les variables au centre afin d'éliminer le ou les clusters auxquels elles appartiennent.

Remarquons que les variables situées au centre ne semblent pas appartenir au même cluster, c'est-à-dire que toutes les variables au centre du graphique F.3 débutent soit par un A, B, C ou D. Contrairement à chez les filles, il ne semble pas y avoir ici un regroupement de variables appartenant toutes à une même définition de cluster. Nous allons voir dans la prochaine section pourquoi nous allons quand même passer à l'élimination des clusters 1 et 4.

4.3.2 Élimination des clusters 1 et 4

Notons que le cluster 1 des garçons est absent car, il ne comportent aucun adolescent. Ce cluster est en effet un regroupement de variables contradictoires donc, il n'est pas surprenant que personne n'y adhère. Effectivement, il comprend les 2 variables suivantes :

- APSCOR1 : importance du corps dans les APS «très important»
- APSCOR5 : importance du corps dans les APS «très peu important»

Remarquons que comme on n'a pas accès au premier travail statistique qui avait été fait, on ne peut vérifier si cela avait été voulu ou pas.

Aussi, le cluster 4 est aussi très peu peuplé (figure F.4 de l'annexe F). Il comprend en fait 9 adolescents et ils sont dispersés les uns des autres dans les cadrans 2 et 3 donc, pas très convainquant. Ce cluster comprend par exemple les variables suivantes :

- MMALAISE12 : n'aime pas la mixité pcq se sent dévisagé «tout à fait ou plutôt en accord»
- MCOMPARE1 : n'aime pas la mixité pcq comparé aux autres «tout à fait en accord»
- MMOQUERI12 : n'aime pas la mixité pcq se sent moqué «tout à fait ou plutôt en accord»
- PDEPASSE1 : n'aime la mixité pcq se dépasse pour impressionner «tout à fait en accord»

C'est donc pour ces variables contradictoires entre elles, énumérées ci-dessus, que peu de garçons cadrent avec la définition de ce cluster.

C'est donc pour les raisons mentionnés dans cette section que nous allons procéder comme dans le cas des filles, nous éliminons les cluster 1 et 4 de l'analyse afin de mieux observer les clusters 2, 3 et 5 individuellement.

Notons que les graphiques, avant et après les éliminations des clusters 1 et 4, du cluster 2 (F.5 et F.6), du cluster 3 (F.7 et F.8) et ceux du cluster 5 (F.9 et F.10) seront analysés dans les prochaines sections.

4.3.3 Analyse du cluster 2

Sur le graphique de la représentation Biplot des garçons après les éliminations des clusters 1 et 4 (figure F.6 de l'annexe F) le cluster 2 est un peu mieux défini qu'au graphique avant l'élimination (figure F.5). Donc, l'élimination des clusters 1 et 4 a eu un effet bénéfique sur le regroupement de ce cluster. Le cluster 2 des garçons est similaire au cluster 3 des filles. Ce sont des garçons qui pratiquent le sport quelques fois par semaine (*prataps3*) et qui ont une attirance vers le risque (*risq1*). Ils prennent plaisir au sport (*apsfun1*) mais, ne sont toutefois pas des gens pour qui la santé passe avant tout (*nourtur2*). Ce cluster ne présente donc pas de contradiction de variables en soit.

Remarquons qu'ils sont le cluster le plus peuplé, 231 garçons. Donc, on peut

affirmer qu'ils représentent la majorité des garçons du total des 403 garçons.

Notons que, comme chez le cluster 3 des filles, la masse d'individus se situe au centre du graphique. Cette tendance nous porte à croire que ce cluster est différent des autres clusters.

De plus, l'attirance pour le risque est une variable, comme le cluster 3 des filles, qui définit ce cluster 2 des garçons. Cette variable, *risq1*, semble ici aussi ne pas être à sa place. Dans ce cluster, les garçons arrête face à la douleur dans les activités (*douleur2*) et sont attentifs aux malaises de leur corps (*ecoucor1*) donc, sûrement pas des gens qui aiment le risque.

Finalement, remarquons que cette variable du risque semble se retrouver du côté du cluster 3. Nous verrons dans la prochaine section, qu'effectivement, elle cadrerait mieux avec les variables de ce cluster plus actif.

4.3.4 Analyse du cluster 3

Pour le cluster 3 des garçons, ce qui ressort depuis l'élimination des cluster 1 et 4, est que la masse d'individus du cluster 3 s'est déplacé vers le cadran 3. Effectivement, avec la présence de toutes les 47 variables (figure F.7), les garçons appartenant au cluster 3 étaient dispersé un peu dans tous les cadrans, majoritairement dans le cadran 3 mais, on en retrouvait aussi dans les autres. Depuis l'élimination (figure F.8), ils sont surtout concentrés dans le cadran 3 et le cadran 2 et ne dépassent plus dans le cadran 1.

Remarquons que ce cluster comprend 99 garçons et est celui qui est le plus en faveur de la mixité. Ces adolescents pratiquent souvent des sports (*prataps4*) et ont une préférence pour les activités collectives (*apspref1*).

Aussi, la variable *risq1* qui attribue le goût du risque est sûrement mieux dans ce cluster où les garçons sont moins attentifs aux malaises corporels (*ecoutcor2*) et qui poussent plus lorsque la douleur se fait ressentir (*douleur3*).

De plus, notons que ce cluster n'est pas concentré à l'origine comme le cluster 2 ce qui permet de croire qu'il représente un cluster différent de celui-ci.

4.3.5 Analyse du cluster 5

Finalement, à l'opposé graphique du cluster 3, on retrouve le cluster 5. Le graphique du biplot du cluster 5 sans les variables des clusters 1 et 4 (figure F.10) nous démontre que ce cluster se retrouve en effet dans les cadrans 1 et 4 et n'est non plus centré à l'origine. Cependant, si l'on fait la comparaison avec le graphique lorsqu'il était en présence des 47 variables (figure F.9), on remarque que l'élimination des clusters 1 et 4 a fait que sa masse a faite un mouvement vertical en partant du cadran 4 vers le cadran 1.

Remarquons que ce cluster est composé de peu de variables, soit 6 variables, et qu'il est aussi peu peuplé, soit 64 garçons. Ils sont des garçons pour qui la santé est importante (*santea3*), peu compétitifs (*apscomp5*) et qui accorde une importance à leur forme physique (*apsform1*). Cette définition de cluster représente le contraire de celle du cluster 3 où les garçons sont moins attentionnés à leur corps et plus au sport. On remarque cette différence par le positionnement des clusters sur les graphiques biplots.

Aussi, les variables *santeapp3*, *apsform1* et *nourtur3* sont très près les unes des autres et se sont les mêmes que chez les filles. On pourrait donc, comme chez les filles, vouloir les représentées à l'aide d'une seule.

Notons que ce cluster ne comporte aucune variable qui les associe à une activité sportive ou physique de préférence.

4.3.6 Conclusion chez les garçons

Sommes toutes, les graphiques des garçons présentent quelques clusters pertinents mais, majoritairement les regroupements se font autour de l'origine ce qui nous porte à croire qu'une réévaluation de la classification est de mise. Tous les individus se surtout

regroupés au centre et il est assez difficile de tirer des conclusions car, ils représentent la moyenne et cela nous est pas très pertinent au sens statistique.

Les mêmes étapes que chez les filles pourraient être faites soient, la revue des clusters ici inutiles (clusters 1 et 4), la réassignation de quelques variables qui tendent vers un autre cluster et l'association de variables afin d'en réduire le nombre.

Finalement, on faudrait aussi refaire une sélection des variables. En effet, 47 variables semblent causer problème dans cette analyse, c'est un peu trop nombreux.

CONCLUSION

Ce mémoire avait comme but la validation de clusters à l'aide de biplots sur analyse des correspondances multiples. Nous y avons donc fait la discussion de la technique qui permet la construction du biplot sur données qualitatives. A partir d'une banque de données, obtenue de madame Laberge, d'une étude sur la mixité au secondaire, nous avons d'abord élaboré une analyse des correspondances multiples pour ensuite appliquer la théorie du biplot de John Gower sur cette analyse.

Tout d'abord, nous avons présenté les données et la première analyse qui avait été faite en janvier 1999. En effet, un premier travail statistique avait été réalisé et quelques conclusions en avaient été tirées. Donc, la première partie présente quelques faits pertinents sur cette analyse et la conversion de ces programmes statistiques, préliminairement faits en SPSS (logiciel statistique), en langage SAS, le logiciel statistique utilisé par madame Rousseau pour les analyses biplots. Nous avons principalement discuté de ces logiciels et fait la présentation des données sur les adolescents (es) recensés (es).

Dans le deuxième chapitre, nous avons fait la discussion des généralités de l'analyse en composante principale et le biplot sur données quantitatives. Nous avons principalement discuté d'analyse linéaire, de décomposition spectrale et de manipulation de données quantitatives.

La troisième partie est celle qui fait le complément du chapitre II afin d'utiliser les techniques qui y sont présentés sur les données de madame Laberge qui sont qualitatives. On y présente la théorie de John Gower appliquée sur ce type de données. On y parle principalement d'Analyse des Correspondances, Analyse des Correspondances Multiples (ACM) et de Biplot sur ACM et toujours sur une base algébrique.

Finalement, la dernière partie applique ces théories à la base de données. L'analyse biplot obtenue a permis de visualiser l'essentiel de l'information contenue dans le tableau de données sur la mixité afin de valider les clusters qui avaient été faits lors de la première analyse. On y discute du programme SAS créé afin de construire les biplots et de visualiser l'espace des individus et des variables. Aussi, on y discute de clustering et de dispersion des individus dans le plan.

De plus, il serait intéressant de pousser davantage cette analyse. Ce mémoire avait comme objectif de valider les clusters préliminairement établis par madame Larberge. Cependant, compte tenu que les clusters obtenus de la première analyse ne se dégageaient pas assez les uns des autres, nous croyons qu'ils manquaient de précisions quant à leur définition. C'est pour ce manque de netteté entre les clusters que nous estimons qu'il serait profitable de reprendre cette analyse du début avec **SAS** et de redéfinir les compositions de ceux-ci. Compte tenu du fait que les premières analyses n'ont pas été conservées, nous devons reprendre cette analyse sans savoir comment les clusters avaient été définis.

APPENDICE A

DESCRIPTION DES VARIABLES DU CAS RÉEL

Tableau A.1 Description des variables

Variable	Description
prataps1	pratique activités physiques ou sportives (APS) «peu souvent»
prataps2	pratique APS «parfois»
prataps3	pratique APS «qq fois par semaine»
prataps4	pratique APS «souvent»
apspref1	APS de préférences : «activités collectives»
apspref2	APS de préférences : «activités à deux»
apspref3	APS de préférences : «activités individuelles»
apspref4	APS de préférences : «plein» air»
apspref5	APS de préférences : «marche»
apspref6	APS de préférences : «aime pas faire activités»
apsfun1	importance du plaisir dans les APS «très important»
apsfun45	importance du plaisir dans les APS «peu ou très peu important»
apsform1	importance de la forme dans les APS «très important»
apsfor45	importance de la forme dans les APS «peu ou très peu important»
apscor1	importance du corps dans les APS «très important
apscor5	importance du corps dans les APS «très peu important
<i>suite à la prochaine page</i>	

<i>suite de la page précédente</i>	
Variable	Description
apscomp1	importance de la compétition dans APS «très important»
apscomp5	importance de la compétition dans APS «très peu important»
apsrlx12	importance de relaxer dans APS «très important ou important»
apsrlx5	importance de relaxer dans APS «très peu important»
genedir1	géné de dire les malaises dans les APS «oui à cause de la perception des autres»
genedir2	géné de dire les malaises dans les APS «oui pcq pas se plaindre»
genedir3	géné de dire les malaises dans les APS «non pcq normal de dire»
genedir4	géné de dire les malaises dans les APS «non pour attirer l'attention»
nourtur1	choix des aliments à la cafétéria «bourratif»
nourtur2	choix des aliments à la cafétéria «santé pas important»
nourtur3	choix des aliments à la cafétéria «bon pour la santé»
risq1	attirance pour le risque «oui»
risq2	attirance pour le risque «non»
douleur1	réaction face à la douleur dans l'intensité «arrête et prudent»
douleur2	réaction face à la douleur dans l'intensité «arrête et content»
douleur3	réaction face à la douleur dans l'intensité «pousse plus»
ecoucor1	attention aux malaises corporels «oui»
ecoucor2	attention aux malaises corporels «non»
mmalais12	aime pas la mixité pcq se sent dévisagé «tout à fait ou plutôt en accord»
mmalais4	aime pas la mixité pcq se sent dévisagé «tout à fait en désaccord»
mcompar1	aime pas la mixité pcq compare aux autres «tout à fait en accord»
mcompar4	aime pas la mixité pcq compare aux autres «tout à fait en désaccord»
mmoquer12	aime pas la mixité pcq autre sexe se moque «tout à fait ou plutôt en accord»
mmoquer4	aime pas la mixité pcq autre sexe se moque «tout à fait en désaccord»
pdepase1	aime la mixité pcq se dépasse pour impressionner «tout à fait en accord»
pdepase4	aime la mixité pcq se dépasse pour impressionner «tout à fait en désaccord»
<i>suite à la prochaine page</i>	

<i>suite de la page précédente</i>	
Variable	Description
santeap1	santé vs apparence physique «corps pas important»
santeap2	santé vs apparence physique «plus apparence physique»
santeap3	santé vs apparence physique «plus santé»
mode1	confort d'abord
mode2	mode d'abord

APPENDICE B

LES PROGRAMMES

B.1 Programme 1 : Lecture des données originales

```
libname q 'C:\saslibrary';

DATA q.mixidon ;

    infile 'C:\Documents and Settings\Anie\My Documents\mixidon_txt.htm' firstobs=1;

    /*indentification des variables*/

input   fclus $ 1-4 region 5-6 sexe 7 prataps 8 apspref 9 apsfun 11 apscomp 12 apsform 13
        apscorps 14 apsrelax 15 prefsexe 16 compet 17 risq 18 eptime 19-20 epmoy 21-22
        habil 23 newspo 24 epmix 25 mplate 26 pconnait 27 mmalaise 28 pmeieur 29
        mcompare 30 pequil 31 mmoqueri 32 pdepasse 33 mcomplex 34 ptetgars 35 prefmix 36
        collmix 37 deumix 38 aquamix 39 combmix 40 expremix 41 plenamix 42 indimix 43
        sexeprof 44 raison 45 ephequip 46 ephdefou 47 ephtech 48 ephcomp 49 ephlimit 50
        ephrelax 51 ephform 52 douleur 53 ecoutcor 54 genedire 55 nouritur 56 mode 57
        santeapp 58 sexediff 59 diffcmpt 60 diffpsy 61 diffmode 62 diffapp 63
        bonpfemi 64 expremga 65 garsppop 66 fillppop 67 nivsec 68 age 69-70
        famille 71 occuper 72 occumere 73 scolpere 74 scolmere 75 nivecono 76;

    /*explication des noms des variables*/

label   prataps='pratique activités physiques ou sportives(APS)'
        apspref='APS de préférence'
        apsfun='importance plaisir APS'
        apscomp='importance compétition APS'
```

apsform='importance forme APS'
 apscorps='importance corps APS'
 apsrelax='importance relaxer APS'
 prefsexe='préférence sexe partenaire APS'
 compet='attirance pour compétition APS'
 risq='attirance pour risque APS'
 epnote='note moyenne des étapes en Education physique(EPH)'
 epmoy='moyenne de classe EPH'
 habil='perception niveau habileté APS'
 newspo='réaction face à un nouveau sport'
 epmix='a eu cours EPH mixte au secondaire'
 Mplate='aime pas mixité pcq plate'
 Pconnait='aime mixité pcq connaître autre sexe'
 Mmalaise='aime pas mixité pcq se sent dévisagé'
 Pmeieur='aime mixité pcq meilleur que autre sexe'
 Mcompare='aime pas mixité pcq compare autre sexe'
 Pequil='aime mixité pcq travaille plus'
 Mmoqueri='aime pas mixité pcq autre sexe se moque'
 Pdepasse='aime mixité pcq se dépasser pour impressionner'
 Mcomplex='aime pas mixité pcq se sent moins bon que autre sexe'
 Ptetgars='aime mixité pcq travail tet'
 prefmix='pour ou contre la mixité'
 collmix='pour ou contre la mixité act collective'
 deuxmix='pour ou contre mixité act à deux'
 aquamix='pour ou contre mixité act aquatique'
 combmix='pour ou contre mixité act combat'
 expremix='pour ou contre mixité act expression'
 plenamix='pour ou contre mixité act plein air'
 indimix='pour ou contre act individuelle'
 sexeprof='préférence professeur de même sexe'
 raison='raison de préférence de prof même sexe'
 ephequip='EPH sert pour équipe'
 ephdefou='EPH sert pour défouler'
 ephtech='EPH sert pour technique'
 ephcomp='EPH sert pour compétition'
 ephlimit='EPH sert pour découvrir limites'
 ephrelax='EPH sert pour se détendre'
 ephform='EPH sert pour la forme'
 douleur='intensité max réaction face à la douleur'
 ecoutcor='attention malaises corporels'
 genedire='gène de dire malaises'
 nouritur='choix aliments cafétéria'

```

mode='importance dans habillage'
santeapp='sante vs apparence physique'
sexediff='importance différences entre sexe'
diffcmt='comportement différent selon sexe'
diffpsy='psychologie différente selon sexe'
diffmode='habillement différent selon sexe'
diffapp='apparence différente selon sexe'
bonPfemi='fille bonne sport effort féminine'
expreMga='gars pas fait pour danser'
garsPpop='gars bon sport plus populaire avec filles'
fillPpop='fille bonne sport plus populaire avec gars'
nivsec='niveau du secondaire'
famille='situation familiale'
occupere='occupation du père'
occumere='occupation de la mère'
scolpere='scolarité du père'
scolmere='scolarité de la mère'
nivecono='perception niveau économique'
;

/*légende des variables*/ /*on ne peut mettre le même nom que la
variable, c'est pourquoi j'ai rajouté une lettre à la fin*/

PROC FORMAT;

value regionn 10='Bas St-Laurent' 20='Saguenay' 30='Québec' 40='Trois-Rivière'
50='Estrie' 61='MtlNord' 62='MtlSud' 63='Ile Mtl' 70='Outaouais'
80='Abitibi' 90='Côte Nord';
value sexee 1='fille' 2='garçon';
value pratapss 1='peu souvent' 2='parfois' 3='qq fois semaine' 4='souvent';
value apspreff 1='act coll' 2='act à deux' 3='act indiv' 4='plein air' 5='marche'
6='aime pas act' 7='pas de préférence';
value aps 1='très important' 2='important' 3='moyen' 4='peu important'
5='très peu important';
value prefsex 1='même sexe même si moin bon' 2='même sexe même niveau'
3='autre sexe même niveau' 4='indifférent sexe même niveau'
5='autre sexe même si moin bon';
value compett 1='oui caractère' 2='oui jeu Pint' 3='oui dépasse' 4='non conflit'
5='non violence' 6='non décourage';
value ree 1='oui' 2='non';/*risq epmix ecoutcor*/
value sexedif 1='non' 2='oui';
value habill 1='très bon' 2='moyen' 3='faible';

```

```

value newspoo 1='joue beaucoup' 2='apprendre' 3='joue peu';
value epmixx 1='oui' 2='non';
value M 1='tout à fait en accord' 2='plutôt en accord' 3='plutôt en désaccord'
4='tout à fait en désaccord';
value prefmixx 1='tout à fait en faveur' 2='plutôt en faveur' 3='plutôt contre'
4='tout à fait contre' 5='ça dépend du prof' 6='pas opinion';
value mix 1='mixte' 2='séparé' 3='pas opinion';
value sexepro 1='jamais eu prof même sexe' 2='jamais eu prof autre sexe'
3='préfère femme' 4='préfère homme' 5='indifférent';
value raisonn 0='affinité de genre' 1='autre sexe plus motivant'
2='pref homme car plus relax' 3='pref homme car plus sportif'
4='pref femme car plus gentile' 5='pref femme car plus motivante'
6='pref pas homme car plus macho' 7='pref pas homme car plus sévère'
8='pref pas homme car plus pro fille' 9='pref pas femme car moins relax';
value eph 1='ppp important' 2='pp important' 3='p important' 4='neutre' 5='m important'
6='mm important' 7='mmm important';
value douleurr 1='arrête et prudent' 2='arrête et content' 3='pousse plus';
value ecoutco 1='oui' 2='non';
value genedir 1='oui perception des autres' 2='oui pcq pas se plaindre'
3='non pcq normal de dire' 4='non pour attention des autres';
value nouritu 1='bourratif' 2='santé pas important' 3='bon pour la santé';
value modee 1='confort d'abord' 2='mode d'abord';
value santea 1='corps pas important' 2='plus apparence physique' 3='plus santé';
value famillee 1='père et mère' 2='mère' 3='père' 4='autre';
value occ 0='chômeur étudiant' 1='maison' 2='ouvrier spécialisé' 3='col blanc'
4='technicien agent' 5='cadre supérieur' 6='professionnel'
7='cadre moyen' 8='petit proprio' 9='ouvrier non spécialisé';
value scol 1='primaire' 2='secondaire' 3='technique' 4='collégial' 5='université'
6='ne sais pas';
value nivcono 1='riche' 2='moyen' 3='pauvre';

RUN;

/*Il faut associer chaque variable avec la légende appropriée */

PROC PRINT label noobs;

format region regionn. sexe sexee. prataps pratapss. apspref apspreff. apsfun aps. apscorp.
apsform aps. apscorps aps. apsrrelax aps. prefsexe prefsex. compet compett. risq ree.
epnote ree. epmoy ree. habil habill. newspoo newspoo. epmix epmixx. mplate M. pconnait M.
mmalaise M. pmeieur M. mcompare M. pequil M. mmoqueri M. pdepasse M. mcomplex M.
ptetgars M. prefmix prefmixx. collmix mix. deuxmix mix. aquamix mix. combmix mix.

```

```

exremix mix. plenamix mix. indimix mix. sexeprof sexeupro. raison raisonn.
ephequip eph. ephdefou eph. ephtech eph. ephcomp eph. ephlimit eph. ephrelax eph.
ephform eph. douleur douleurr. ecoutcor ecoutco. genedire genedir. nouritur nouritu.
mode modee. santeapp santea. sexediff sexedif. diffcmpt M. diffpsy M. diffmode M.
diffapp M. bonpfemi M. expremga M. garsppop M. fillppop M. famille familiee.
occupere occ. occumere occ. scolpere scol. scolmere scol. nivecono nivcono.;

```

```
RUN;
```

```
PROC FREQ;
```

```
RUN;
```

B.2 Programme 2 : Création de variables dichotomiques

```
libname q 'C:\saslibrary'; libname r 'C:\saslibrary';
```

```
DATA r.dicho ;
```

```

set q.mixidon;

array A{4} prataps1-prataps4;
array B{7} apspref1-apspref7;
array C{5} apsfun1-apsfun5;
array D{5} apscomp1-apscomp5;
array E{5} apsform1-apsform5;
array F{5} apscor1-apscom5;
array G{5} apsrlx1-apsrlx5;
array H{2} risq1-risq2;
array J{5} mmalais1-mmalais5;
array K{5} mcompar1-mcompar5;
array L{5} mmoquer1-mmoquer5;
array M{5} pdepase1-pdepase5;
array NN{3} douleur1-douleur3;
array O{4} genedir1-genedir4;
array P{3} nourtur1-nourtur3;
array Q{3} santeap1-santeap3;
array R{1} ecoucor1;
array S{1} ecoucor2;

```

```
array T{1} mode1;
array U{1} mode2;

do I=1 to 1;
  if ecoutcor=1 then R{I}=1;
  else R{I}=0;

end;

do I=1 to 1;
  if ecoutcor=2 then S{I}=1;
  else S{I}=0;

end;

do I=1 to 1;
  if mode=1 then T{I}=1;
  else T{I}=0;

end;

do I=1 to 1;
  if mode=2 then U{I}=1;
  else U{I}=0;

end;

do I=1 to 4;
  if prataps=I then A{I}=1;
  else A{I}=0;

end;

do I=1 to 7;
  if apspref=I then B{I}=1;
  else B{I}=0;

end;

do I=1 to 5;
  if apsfun=I then C{I}=1;
```

```
        else C{I}=0;

end;

do I=1 to 5;
    if apscomp=I then D{I}=1;
    else D{I}=0;

end;

do I=1 to 5;
    if apsform=I then E{I}=1;
    else E{I}=0;

end;

do I=1 to 5;
    if apscorps=I then F{I}=1;
    else F{I}=0;

end;

do I=1 to 5;
    if apsrelax=I then G{I}=1;
    else G{I}=0;

end;

do I=1 to 2;
    if risq=I then H{I}=1;
    else H{I}=0;

end;

do I=1 to 5;
    if mmalaise=I then J{I}=1;
    else J{I}=0;

end;

do I=1 to 5;
    if mcompare=I then K{I}=1;
```

```
    else K{I}=0;

end;

do I=1 to 5;
    if mmoqueri=I then L{I}=1;
    else L{I}=0;

end;

do I=1 to 5;
    if pdepasse=I then M{I}=1;
    else M{I}=0;

end;

do I=1 to 3;
    if douleur=I then NN{I}=1;
    else NN{I}=0;

end;

do I=1 to 4;
    if genedire=I then O{I}=1;
    else O{I}=0;

end;

do I=1 to 3;
    if nouritur=I then P{I}=1;
    else P{I}=0;

end;

do I=1 to 3;
    if santeapp=I then Q{I}=1;
    else Q{I}=0;

end;

drop I;
```

```

RUN;

PROC PRINT data=r.dicho;

var prataps prataps1-prataps4 apspref apspref1-apspref7 apsfun apsfun1-apsfun5
    apscomp apscomp1-apscomp5 apsform apsform1-apsform5 apscorps apscor1-apskor5
    apsrelax apsrlx1-apsrlx5 risq risq1-risq2 mmalaise mmalais1-mmalais5
    mcompare mcompar1-mcompar5 mmoqueri mmoquer1-mmoquer5 pdepasse pdepase1-pdepase5
    douleur douleur1-douleur3 genedire genedir1-genedir4 nouritur nouritur1-nouritur3
    santeapp santeap1-santeap3 ecoutcor ecoucor1 ecoucor2 mode mode1 mode2 fclus;

RUN;

PROC FREQ;

RUN;

```

B.3 Programme 3 : Élimination des données manquantes

```

libname r 'C:\saslibrary'; libname s 'C:\saslibrary';

DATA s.nonmissf;

    set r.dicho;

    array Z{65} prataps prataps1-prataps4 apspref apspref1-apspref6 apsfun apsfun1 apsfun5
        apscomp apscomp1 apscomp5 apsform apsform1 apsform5 apscorps apscor1 apscor5
        apsrelax apsrlx1 apsrlx5 risq risq1 risq2 mmalaise mmalais1 mmalais4
        mcompare mcompar1 mcompar4 mmoqueri mmoquer1 mmoquer4 pdepasse pdepase1 pdepase4
        douleur douleur1-douleur3 genedire genedir1-genedir4 nouritur nouritur1-nouritur3
        santeapp santeap1-santeap3 ecoutcor ecoucor1 ecoucor2 mode mode1 mode2;

    if sexe=1;

    do I=1 to 65;
        if Z{I}='.' then delete;
    end;

```

```

end;

keep  prataps1-prataps4  apspref1-apspref6  apsfun1  apsfun5
      apscomp1  apscomp5  apsform1  apsform5  apscor1  apscor5
      apsrlx1  apsrlx5  risq1  risq2  mmalais1  mmalais4
      mcompar1  mcompar4  mmoquer1  mmoquer4  pdepase1  pdepase4
      douleur1-douleur3  genedir1-genedir4  nourtur1-nourtur3
      santeap1-santeap3  ecoucor1  ecoucor2  mode1  mode2  nivsec  fclus;

```

```
RUN;
```

```
PROC PRINT data=s.nonmissf;
```

```

var  prataps1-prataps4  apspref1-apspref6  apsfun1  apsfun5
     apscomp1  apscomp5  apsform1  apsform5  apscor1  apscor5
     apsrlx1  apsrlx5  risq1  risq2  mmalais1  mmalais4
     mcompar1  mcompar4  mmoquer1  mmoquer4  pdepase1  pdepase4
     douleur1-douleur3  genedir1-genedir4  nourtur1-nourtur3
     santeap1-santeap3  ecoucor1  ecoucor2  mode1  mode2  fclus;

```

```
RUN;
```

B.4 Programme 4 : Réduction de niveaux

```
libname s 'C:\saslibrary'; libname h 'C:\saslibrary';
```

```
DATA h.quarantecinqf;
```

```
set s.nonmissf;
```

```

if  apsfun>=4 then apsf45=1;
else apsf45=0;

```

```

if  apsform>=4 then apsfo45=1;
else apsfo45=0;

```

```
if  apscomp<=2 then apsco12=1;
```

```

else apsco12=0;

keep prataps1-prataps4 apspref1-apspref6 apsfun1 apsfun5
  apscomp1 apscomp5 apsform1 apsform5 apscor1 apscor5
  apsrlx1 apsrlx5 risq1 risq2 mmalais1 mmalais4
  mcompar1 mcompar4 mmoquer1 mmoquer4 pdepase1 pdepase4
  douleur1-douleur3 genedir1-genedir4 nourtur1-nourtur3
  santeap1-santeap3 ecoucor1 ecoucor2 mode1 mode2 apsf45 apsf45 apsco12 nivsec fclus;

RUN;

PROC PRINT data=h.quarantecinqf;
  var apsf45 apsf45 apsco12;

RUN;

```

B.5 Programme 5 : Création de la variable *summix*

```

libname w 'C:\saslibrary'; libname u 'C:\saslibrary';

DATA w.summixf;

  set h.quarantecinqf;

  summix=mmalais1+mmalais4+mcompar1+mcompar4+mmoquer1+mmoquer4+pdepase1+pdepase4;

RUN;

PROC PRINT data=w.summixf;
  var summix;

RUN;

```

B.6 Programme 6 : Transformation linéaire

```

libname h 'C:\saslibrary'; libname u 'C:\saslibrary';

DATA u.nouvelledonneef;

set h.quarantecinqf;

array x{47} prataps1-prataps4 apspref1-apspref6 apsfun1 apsf45
apsco12 apscomp5 apsform1 apsf045 apscor1 apscor5 apsr1x1 apsr1x5
risq1 risq2 mmalais1 mmalais4 mcompar1 mcompar4 mmoquer1 mmoquer4
pdepase1 pdepase4 douleur1-douleur3 ecoucor1 ecoucor2
genedir1-genedir4 nourtur1-nourtur3 mode1 mode2 santeap1-santeap3;

array y{47} y1-y47 (62 168 251 94 206 84 124 50 17 5 248 67 66 359
229 33 56 86 25 113 359 216 51 122 111 103 39 202 53 112 162 331
82 429 146 36 13 468 58 66 270 239 415 160 74 225 276);

do i=1 to 47;

x{i}=x{i}*(575-y{i})/575; end;

keep prataps1-prataps4
    apspref1-apspref6
    apsfun1 apsfun5
    apscomp1 apscomp5
    apsform1 apsform5
    apscor1 apscor5
    apsr1x1 apsr1x5
    risq1 risq2
    mmalais1 mmalais4
    mcompar1 mcompar4
    mmoquer1 mmoquer4
    pdepase1 pdepase4
    douleur1-douleur3
    genedir1-genedir4
    nourtur1-nourtur3
    santeap1-santeap3

```

```

ecoucor1 ecoucor2
mode1 mode2 apsf45 apscor12 apsf45 nivsec fclus;

RUN;

PROC PRINT data=u.nouvelledonneef;

var prataps1-prataps4 apspref1-apspref6 apsfun1 apsf45 apscor12
apscor5 apsf45 apscor1 apscor5 apsr1x1 apsr1x5 risq1
risq2 mmalais1 mmalais4 mcompar1 mcompar4 mmoquer1 mmoquer4
pdepase1 pdepase4 douleur1-douleur3 ecoucor1 ecoucor2
genedir1-genedir4 nourtur1-nourtur3 mode1 mode2 santeap1-santeap3;

RUN;

```

B.7 Programme 7 : Création des fclus et mclus

```

libname w 'C:\saslibrary'; libname v 'C:\saslibrary';

DATA v.clusterf;

set w.summixf;

fclus1=(prataps1+apspref2+apspref3+apspref4+apspref5+apspref6+apsf45+
apsfo45+apscor1+apsr1x1+mmalais1+mcompar1+mmoquer1+pdepase4+genedir1+
genedir2+genedir4+nourtur1+santeap1)/19;

fclus2=(prataps2+apspref1+apsfun1+risq2+douleur1+nourtur2+mode2+santeap2)/8;

fclus3=(prataps3+apscor5+apsform1+risq1+douleur2+ecoucor1+genedir3+
nourtur3+mode1+santeap3)/10;

fclus4=(prataps4+apscor12+apscor5+apsr1x5+mmalais4+mcompar4+mmoquer4+
pdepase1+douleur3+ecoucor2)/10;

```

```

RUN;

PROC PRINT data=v.clusterf;
var fclus1 fclus2 fclus3 fclus4 summix;

RUN;

```

B.8 Programme 8 : Ajustement des variables fclus et mclus

```

libname v 'C:\saslibrary'; libname x 'C:\saslibrary';

DATA x.clusterdeuxf;

set v.clusterf;

if summix=0 then

fclus1=(prataps1+apspref2+apspref3+apspref4+apspref5+apspref6+apsf45+
apsfo45+apscor1+apsrlx1+genedir1+genedir2+genedir4+nourtur1+santeap1)/15;

else

fclus1=(prataps1+apspref2+apspref3+apspref4+apspref5+apspref6+apsf45+
apsfo45+apscor1+apsrlx1+mmalais1+mcompar1+mmoquer1+pdepase4+genedir1+
genedir2+genedir4+nourtur1+santeap1)/19;

if summix=0 then

fclus4=(prataps4+apsco12+apscor5+apsrlx5+douleur3+ecoucor2)/6;

else

fclus4=(prataps4+apsco12+apscor5+apsrlx5+mmalais4+mcompar4+mmoquer4+
pdepase1+douleur3+ecoucor2)/10;

RUN;

PROC PRINT data=x.clusterdeuxf;
var fclus1 fclus2 fclus3 fclus4;

RUN;

```

B.9 Programme 9 : Assignment d'un individu à un cluster

```
libname x 'C:\Documents and Settings\alemieux\Bureau\saslibrary';

libname y 'C:\Documents and Settings\alemieux\Bureau\saslibrary';

DATA y.vmaxf;

set x.clusterdeuxf;

if fclus1=max(fclus1,fclus2,fclus3,fclus4)

then fclus='1';

if fclus2=max(fclus1,fclus2,fclus3,fclus4)

then fclus='2';

if fclus3=max(fclus1,fclus2,fclus3,fclus4)

then fclus='3';

if fclus4=max(fclus1,fclus2,fclus3,fclus4)

then fclus='4';

vmax=max(fclus1,fclus2,fclus3,fclus4);

if nivsec>=4 then rnivsec=2;
  else rnivsec=1;

RUN;

PROC PRINT;

var fclus1 fclus2 fclus3 fclus4 vmax fclus;

RUN;
```

```
PROC SORT;

by fclus;

RUN;

PROC MEANS data=y.vmaxf;

var fclus1 fclus2 fclus3 fclus4; by fclus;

RUN;

PROC FREQ data=y.vmaxf;

tables fclus*rnivsec/chisq;

RUN;
```

B.10 Programme 11 : ANOVA

```
libname y 'C:\saslibrary'; libname yy 'C:\saslibrary';

DATA yy.anovaf;

set y.vmaxf;

if rnivsec=1;

PROC ANOVA ;
  class fclus ;
  model fclus1-fclus4 = fclus ;
  means fclus / snk ;

RUN;

DATA anovaf;
```

```
set y.vmaxf;  
  
if rnivsec=2;  
  
PROC ANOVA ;  
  class fclus ;  
  model fclus1-fclus4 = fclus ;  
  means fclus / snk ;  
  
RUN;
```

APPENDICE C

RÉSULTATS DES PROGRAMMES SAS

C.1 Résultats programme 1

C.1.a Lecture des données originales avec label

ind	region	sexe	pratique activités physiques ou		importance	
			sportives(APS)	APS de préférence	plaisir APS	compétition APS
0001	Bas St-Laurent	fille	qq fois semaine	act indiv	peu important	très peu important
0002	Bas St-Laurent	fille	qq fois semaine	act indiv	peu important	très peu important
0003	Bas St-Laurent	fille	parfois	act à deux	important	très peu important
0004	Bas St-Laurent	fille	qq fois semaine	act coll	très important	très peu important
0005	Bas St-Laurent	fille	parfois	act indiv	très important	très peu important
0006	Bas St-Laurent	fille	parfois	act coll	important	très peu important
0007	Bas St-Laurent	fille	qq fois semaine	pas de préférence	très important	très peu important
0008	Bas St-Laurent	fille	qq fois semaine	marche	peu important	très peu important
0009	Bas St-Laurent	fille	qq fois semaine	act coll	important	peu important
0010	Bas St-Laurent	fille	qq fois semaine	pas de préférence	très important	très peu important
0011	Bas St-Laurent	fille	parfois	act indiv	moyen	peu important
0012	Bas St-Laurent	fille	qq fois semaine	act coll	peu important	très peu important
0013	Bas St-Laurent	fille	qq fois semaine	pas de préférence	moyen	très peu important
0014	Bas St-Laurent	fille	qq fois semaine	act indiv	important	très peu important
0015	Bas St-Laurent	fille	qq fois semaine	act à deux	très important	très peu important
0016	Bas St-Laurent	fille	qq fois semaine	act coll	très important	moyen
0017	Bas St-Laurent	fille	qq fois semaine	act indiv	important	très peu important
0018	Bas St-Laurent	fille	peu souvent	aime pas act	important	peu important
0019	Bas St-Laurent	garçon	souvent	act coll	très important	important

0020 Bas St-Laurent garçon qq fois semaine act à deux moyen très important
 0021 Bas St-Laurent garçon souvent plein air très important peu important

importance forme APS	importance corps APS	importance relaxer APS	préférence sexe partenaire APS
très important	moyen	important	indifférent sexe même niveau
très important	important	moyen	indifférent sexe même niveau
très important	peu important	moyen	indifférent sexe même niveau
important	peu important	moyen	autre sexe même si moins bon
important	peu important	moyen	indifférent sexe même niveau
très important	moyen	peu important	indifférent sexe même niveau
important	moyen	peu important	indifférent sexe même niveau
très important	moyen	important	indifférent sexe même niveau
très important	très peu important	moyen	indifférent sexe même niveau
important	peu important	moyen	indifférent sexe même niveau
important	très peu important	très important	indifférent sexe même niveau
très important	important	moyen	autre sexe même si moins bon
très important	important	peu important	indifférent sexe même niveau
très important	moyen	peu important	indifférent sexe même niveau
moyen	important	peu important	indifférent sexe même niveau
important	très peu important	peu important	indifférent sexe même niveau
très important	moyen	peu important	indifférent sexe même niveau
très peu important	moyen	très important	indifférent sexe même niveau
moyen	peu important	très peu important	indifférent sexe même niveau
important	peu important	très peu important	indifférent sexe même niveau
important	moyen	très peu important	indifférent sexe même niveau

attirance pour compétition APS	attirance pour risque APS	note moyenne des étapes en Education physique(EPH)	moyenne de classe EPH	perception niveau habileté APS	réaction face à un nouveau sport	a eu cours EPH mixte au secondaire
oui dépasse	non	79	80	moyen	apprendre	oui
oui dépasse	oui	80	80	faible	joue beaucoup	oui
non décourage	oui	80	80	moyen	apprendre	oui
oui jeu Pint	oui	89	80	moyen	joue beaucoup	oui
non décourage	non	67	80	faible	joue peu	oui
non conflit	oui	69	80	faible	apprendre	oui
non conflit	oui	82	80	moyen	joue beaucoup	oui

non décourage	oui	78	80	moyen	apprendre	oui
non décourage	non	67	80	moyen	apprendre	oui
oui dépasse	non	76	80	moyen	apprendre	oui
oui jeu Pint	oui	79	80	moyen	apprendre	oui
oui dépasse	non	88	80	très bon	apprendre	oui
oui jeu Pint	non	70	80	moyen	apprendre	oui
non conflit	non	60	80	faible	apprendre	oui
oui dépasse	oui	74	80	moyen	apprendre	oui
oui dépasse	oui	86	80	très bon	apprendre	oui
non décourage	non	.	80	moyen	apprendre	oui
non décourage	non	.	80	faible	joue peu	oui
oui dépasse	oui	86	75	très bon	apprendre	oui
oui dépasse	.	90	75	très bon	joue beaucoup	oui
non conflit	oui	82	75	très bon	apprendre	oui

aime pas mixité
pcq plate

aime mixité pcq
connaître autre sexe

aime pas mixité pcq
se sent dévisagé

plutôt en désaccord	plutôt en accord	plutôt en désaccord
plutôt en désaccord	plutôt en désaccord	plutôt en accord
plutôt en accord	plutôt en désaccord	plutôt en désaccord
plutôt en désaccord	tout à fait en accord	plutôt en désaccord
plutôt en désaccord	plutôt en accord	plutôt en accord
plutôt en désaccord	tout à fait en accord	tout à fait en désaccord
plutôt en désaccord	tout à fait en accord	tout à fait en désaccord
plutôt en désaccord	plutôt en accord	plutôt en désaccord
plutôt en désaccord	plutôt en accord	plutôt en accord
plutôt en accord	plutôt en désaccord	plutôt en désaccord
tout à fait en désaccord	tout à fait en accord	plutôt en accord
tout à fait en désaccord	tout à fait en accord	plutôt en désaccord
tout à fait en désaccord	plutôt en accord	plutôt en désaccord
tout à fait en désaccord	tout à fait en accord	tout à fait en désaccord
plutôt en désaccord	plutôt en accord	plutôt en désaccord
tout à fait en désaccord	tout à fait en accord	tout à fait en désaccord
plutôt en désaccord	plutôt en accord	plutôt en désaccord
tout à fait en désaccord	plutôt en accord	tout à fait en désaccord
tout à fait en accord	tout à fait en désaccord	tout à fait en accord
tout à fait en désaccord	tout à fait en accord	tout à fait en désaccord
tout à fait en désaccord	tout à fait en accord	tout à fait en désaccord

aime mixité pcq meilleur que autre sexe	aime pas mixité pcq compare autre sexe	aime mixité pcq travaille plus
plutôt en accord	plutôt en accord	plutôt en accord
plutôt en désaccord	plutôt en accord	tout à fait en accord
plutôt en accord	tout à fait en désaccord	plutôt en désaccord
tout à fait en accord	plutôt en accord	plutôt en accord
plutôt en accord	plutôt en désaccord	plutôt en accord
plutôt en accord	plutôt en désaccord	tout à fait en accord
tout à fait en désaccord	plutôt en accord	tout à fait en désaccord
plutôt en accord	tout à fait en accord	plutôt en accord
plutôt en désaccord	tout à fait en accord	plutôt en accord
plutôt en désaccord	plutôt en accord	tout à fait en accord
tout à fait en accord	plutôt en accord	tout à fait en accord
tout à fait en accord	plutôt en désaccord	tout à fait en accord
plutôt en accord	plutôt en désaccord	plutôt en accord
plutôt en désaccord	tout à fait en désaccord	plutôt en accord
tout à fait en accord	plutôt en désaccord	plutôt en accord
tout à fait en accord	tout à fait en désaccord	tout à fait en accord
plutôt en accord	tout à fait en accord	plutôt en accord
tout à fait en désaccord	tout à fait en accord	tout à fait en désaccord
tout à fait en désaccord	tout à fait en désaccord	tout à fait en désaccord
tout à fait en désaccord	tout à fait en désaccord	tout à fait en accord
plutôt en désaccord	tout à fait en désaccord	plutôt en accord
	aime mixité pcq se dépasser pour impressionner	aime pas mixité pcq se sent moins bon que autre sexe
tout à fait en désaccord	tout à fait en désaccord	tout à fait en désaccord
tout à fait en désaccord	plutôt en désaccord	plutôt en désaccord
tout à fait en désaccord	tout à fait en désaccord	tout à fait en désaccord
tout à fait en désaccord	plutôt en accord	tout à fait en désaccord
plutôt en désaccord	plutôt en accord	plutôt en désaccord
plutôt en désaccord	plutôt en accord	tout à fait en désaccord
tout à fait en désaccord	plutôt en désaccord	plutôt en désaccord
plutôt en désaccord	plutôt en désaccord	plutôt en accord
plutôt en désaccord	plutôt en désaccord	plutôt en désaccord
tout à fait en désaccord	plutôt en accord	plutôt en désaccord

tout à fait en désaccord	plutôt en accord	tout à fait en désaccord
tout à fait en désaccord	tout à fait en accord	tout à fait en désaccord
plutôt en désaccord	plutôt en désaccord	plutôt en désaccord
tout à fait en désaccord	plutôt en désaccord	tout à fait en désaccord
plutôt en désaccord	plutôt en accord	plutôt en désaccord
tout à fait en désaccord	tout à fait en accord	tout à fait en désaccord
tout à fait en accord	plutôt en désaccord	tout à fait en désaccord
tout à fait en désaccord	tout à fait en désaccord	tout à fait en désaccord
tout à fait en désaccord	plutôt en accord	tout à fait en désaccord
tout à fait en désaccord	tout à fait en accord	tout à fait en désaccord
tout à fait en désaccord	tout à fait en désaccord	tout à fait en désaccord

		pour ou	pour ou	pour ou
		contre la	contre	contre
aime mixité pqc	pour ou contre	mixité act	mixité act	mixité act
travail tet	la mixité	collective	à deux	aquatique

plutôt en accord	tout à fait en faveur	séparé	mixte	mixte
plutôt en accord	pas opinion	séparé	mixte	mixte
plutôt en désaccord	pas opinion	séparé	mixte	mixte
plutôt en accord	tout à fait en faveur	mixte	mixte	mixte
plutôt en désaccord	plutôt en faveur	mixte	séparé	séparé
plutôt en accord	tout à fait en faveur	séparé	pas opinion	mixte
plutôt en accord	tout à fait en faveur	mixte	mixte	mixte
plutôt en accord	tout à fait en faveur	séparé	mixte	séparé
tout à fait en désaccord	plutôt en faveur	mixte	pas opinion	pas opinion
tout à fait en désaccord	plutôt en faveur	pas opinion	mixte	séparé
plutôt en désaccord	tout à fait en faveur	mixte	séparé	séparé
tout à fait en accord	tout à fait en faveur	mixte	mixte	mixte
plutôt en accord	pas opinion	mixte	mixte	séparé
plutôt en désaccord	tout à fait en faveur	mixte	mixte	mixte
tout à fait en désaccord	tout à fait en faveur	pas opinion	mixte	mixte
tout à fait en accord	tout à fait en faveur	mixte	mixte	mixte
plutôt en accord	tout à fait en faveur	mixte	mixte	mixte
tout à fait en désaccord	ça dépend du prof	pas opinion	pas opinion	pas opinion
tout à fait en désaccord	tout à fait en faveur	mixte	mixte	séparé
tout à fait en désaccord	tout à fait en faveur	mixte	mixte	mixte
tout à fait en désaccord	tout à fait en faveur	séparé	séparé	séparé

pour ou	pour ou	pour ou		
contre	contre	contre	pour ou	préférence
mixité act	mixité act	mixité act	contre act	professeur de
				raison de préférence

combat	expression	plein air	individuelle	même sexe	de prof	même sexe
séparé	mixte	mixte	mixte	indifférent	.	.
séparé	mixte	mixte	mixte	indifférent	.	.
séparé	mixte	mixte	mixte	indifférent	.	.
pas opinion	mixte	mixte	mixte	indifférent	.	.
séparé	séparé	mixte	mixte	indifférent	.	.
mixte	mixte	mixte	pas opinion	indifférent	.	.
mixte	mixte	mixte	mixte	indifférent	.	.
séparé	mixte	mixte	mixte	indifférent	.	.
séparé	mixte	mixte	mixte	indifférent	.	.
mixte	mixte	mixte	mixte	indifférent	.	.
séparé	mixte	mixte	mixte	indifférent	.	.
pas opinion	mixte	mixte	mixte	indifférent	.	.
mixte	mixte	mixte	mixte	indifférent	.	.
mixte	mixte	mixte	mixte	indifférent	.	.
séparé	mixte	mixte	mixte	indifférent	.	.
mixte	mixte	mixte	mixte	indifférent	.	.
séparé	mixte	mixte	mixte	indifférent	.	.
pas opinion	pas opinion	pas opinion	pas opinion	indifférent	.	.
séparé	séparé	mixte	séparé	indifférent	.	.
pas opinion	mixte	mixte	mixte	indifférent	.	.
séparé	pas opinion	mixte	mixte	indifférent	.	.

EPH sert pour	EPH sert pour				
équipe	déjouer	technique	compétition	découvrir limites	se détendre
m important	neutre	pp important	mmm important	p important	mm important
pp important	mm important	ppp important	mmm important	m important	p important
mm important	p important	neutre	mmm important	m important	pp important
neutre	pp important	p important	mmm important	mm important	ppp important
ppp important	mm important	p important	mmm important	neutre	m important
neutre	p important	mm important	mmm important	m important	pp important
neutre	pp important	p important	mmm important	m important	ppp important
neutre	mmm important	p important	m important	pp important	mm important
m important	mmm important	pp important	neutre	p important	mm important
m important	mm important	pp important	mmm important	p important	neutre
mmm important	ppp important	mm important	neutre	m important	pp important
neutre	p important	ppp important	mmm important	m important	pp important
m important	pp important	mm important	mmm important	neutre	p important

neutre	m important	pp important	mmm important	mm important	p important
pp important	m important	ppp important	mmm important	neutre	mm important
p important	neutre	m important	mm important	mmm important	pp important
m important	p important	mm important	mmm important	neutre	ppp important
neutre	mm important	pp important	mmm important	p important	ppp important
mm important	p important	mmm important	ppp important	pp important	neutre
mm important	neutre	p important	mmm important	pp important	m important
p important	pp important	neutre	m important	mm important	ppp important

	intensité max	attention		
EPH sert pour	réaction face	malaises		choix aliments
la forme	à la douleur	corporels	géné de dire malaises	cafétéria

ppp important	arrête et prudent	oui	non pcq normal de dire	bon pour la santé
neutre	arrête et content	oui	non pcq normal de dire	bon pour la santé
ppp important	arrête et content	non	non pour attention des autres	santé pas important
m important	pousse plus	oui	non pcq normal de dire	bourratif
pp important	arrête et content	oui	non pcq normal de dire	bourratif
ppp important	arrête et content	oui	non pcq normal de dire	bon pour la santé
mm important	arrête et content	oui	non pcq normal de dire	santé pas important
ppp important	arrête et prudent	oui	non pcq normal de dire	bon pour la santé
ppp important	pousse plus	oui	non pcq normal de dire	santé pas important
ppp important	arrête et prudent	oui	non pcq normal de dire	bourratif
p important	arrête et content	oui	non pcq normal de dire	bon pour la santé
mm important	pousse plus	oui	non pcq normal de dire	bon pour la santé
ppp important	arrête et prudent	non	non pcq normal de dire	santé pas important
ppp important	arrête et content	oui	non pcq normal de dire	bon pour la santé
p important	arrête et content	oui	non pcq normal de dire	bon pour la santé
ppp important	pousse plus	non	non pcq normal de dire	bon pour la santé
pp important	arrête et content	oui	non pcq normal de dire	bon pour la santé
m important	arrête et content	oui	non pcq normal de dire	bourratif
m important	arrête et content	non	non pcq normal de dire	santé pas important
ppp important	pousse plus	oui	non pour attention des autres	santé pas important
mmm important	arrête et prudent	oui	non pcq normal de dire	bourratif

importance		importance	
dans	santé vs apparence	différences	comportement différent
habillement	physique	entre sexe	selon sexe

confort dabord	plus santé	non	.
confort dabord	plus apparence physique	non	.
confort dabord	plus apparence physique	non	.
confort dabord	corps pas important	non	.
mode dabord	plus apparence physique	non	.
confort dabord	plus santé	non	.
confort dabord	plus santé	oui	plutôt en désaccord
confort dabord	plus santé	oui	plutôt en désaccord
confort dabord	plus apparence physique	non	.
confort dabord	plus apparence physique	non	.
confort dabord	plus apparence physique	oui	tout à fait en accord
confort dabord	plus santé	non	.
confort dabord	plus apparence physique	non	.
mode dabord	plus apparence physique	non	.
confort dabord	plus santé	non	.
confort dabord	plus santé	non	.
confort dabord	plus apparence physique	non	.
confort dabord	corps pas important	non	.
mode dabord	plus apparence physique	oui	plutôt en accord
mode dabord	corps pas important	non	.
confort dabord	plus santé	non	.

psychologie différente selon sexe	habillement différent selon sexe	apparence différente selon sexe
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
plutôt en désaccord	tout à fait en accord	plutôt en accord
plutôt en désaccord	tout à fait en désaccord	tout à fait en désaccord
.	.	.
.	.	.
tout à fait en accord	tout à fait en accord	tout à fait en accord
.	.	.
.	.	.
.	.	.
.	.	.

.	.	.
.	.	.
.	.	.
tout à fait en désaccord	tout à fait en accord	plutôt en accord
.	.	.
.	.	.

filles bonne sport effort féminine	gars pas fait pour danser	gars bon sport plus populaire avec filles
plutôt en désaccord	tout à fait en désaccord	plutôt en désaccord
tout à fait en désaccord	plutôt en désaccord	plutôt en accord
plutôt en désaccord	plutôt en désaccord	plutôt en désaccord
plutôt en désaccord	plutôt en désaccord	plutôt en accord
plutôt en désaccord	plutôt en désaccord	plutôt en désaccord
tout à fait en désaccord	tout à fait en désaccord	plutôt en accord
tout à fait en désaccord	plutôt en désaccord	plutôt en désaccord
tout à fait en désaccord	tout à fait en désaccord	tout à fait en désaccord
plutôt en désaccord	tout à fait en désaccord	plutôt en accord
plutôt en désaccord	plutôt en accord	tout à fait en accord
plutôt en désaccord	plutôt en désaccord	plutôt en accord
plutôt en désaccord	tout à fait en désaccord	tout à fait en accord
plutôt en désaccord	tout à fait en désaccord	plutôt en désaccord
tout à fait en désaccord	tout à fait en désaccord	tout à fait en désaccord
tout à fait en désaccord	plutôt en désaccord	plutôt en accord
tout à fait en désaccord	tout à fait en désaccord	tout à fait en accord
plutôt en accord	tout à fait en désaccord	tout à fait en désaccord
tout à fait en désaccord	tout à fait en désaccord	plutôt en accord
tout à fait en désaccord	tout à fait en désaccord	plutôt en accord
plutôt en accord	plutôt en désaccord	plutôt en désaccord
tout à fait en désaccord	tout à fait en accord	tout à fait en désaccord

filles bonne sport plus populaire avec gars	niveau du		situation	
	secondaire	age	familiale	occupation du père
plutôt en désaccord	5	17	père et mère	technicien agent
plutôt en accord	5	16	père et mère	ouvrier spécialisé
plutôt en désaccord	5	17	père et mère	ouvrier spécialisé

plutôt en désaccord	5	17	père et mère	ouvrier non spécialisé
plutôt en désaccord	5	17	père et mère	ouvrier spécialisé
plutôt en désaccord	5	16	père et mère	technicien agent
plutôt en désaccord	5	16	père et mère	technicien agent
tout à fait en désaccord	5	17	père et mère	ouvrier non spécialisé
plutôt en désaccord	5	18	père et mère	ouvrier spécialisé
tout à fait en accord	5	16	père et mère	technicien agent
tout à fait en désaccord	5	17	père et mère	professionnel
plutôt en accord	5	17	père et mère	cadre supérieur
plutôt en désaccord	5	16	père et mère	ouvrier spécialisé
tout à fait en désaccord	5	16	autre	col blanc
plutôt en accord	5	17	père et mère	ouvrier spécialisé
plutôt en accord	5	17	père et mère	ouvrier spécialisé
tout à fait en désaccord	5	17	père et mère	professionnel
tout à fait en accord	5	17	père et mère	col blanc
plutôt en accord	5	18	père et mère	cadre moyen
plutôt en accord	5	18	père et mère	ouvrier non spécialisé
tout à fait en désaccord	5	18	mère	cadre moyen

	perception		
	scolarité	scolarité	niveau
occupation de la mère	du père	de la mère	économique
ouvrier spécialisé	technique	secondaire	moyen
ouvrier non spécialisé	primaire	primaire	pauvre
maison	ne sais pas	ne sais pas	moyen
maison	secondaire	secondaire	moyen
maison	secondaire	secondaire	moyen
maison	technique	technique	moyen
ouvrier non spécialisé	secondaire	secondaire	moyen
maison	secondaire	collégial	moyen
maison	primaire	secondaire	moyen
maison	technique	ne sais pas	moyen
ouvrier spécialisé	université	technique	moyen
professionnel	secondaire	université	moyen
maison	technique	ne sais pas	moyen
ouvrier spécialisé	technique	secondaire	moyen
technicien agent	secondaire	secondaire	moyen
col blanc	secondaire	secondaire	moyen
petit proprio	université	collégial	riche
ouvrier non spécialisé	technique	secondaire	moyen

ouvrier spécialisé	université	collégial	moyen
technicien agent	université	collégial	moyen
cadre moyen	université	université	moyen

C.1.1.b Tableaux des fréquences des variables

region	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
10	109	9.71	109	9.71
20	107	9.54	216	19.25
30	89	7.93	305	27.18
40	102	9.09	407	36.27
50	99	8.82	506	45.10
61	104	9.27	610	54.37
62	119	10.61	729	64.97
63	107	9.54	836	74.51
70	110	9.80	946	84.31
80	76	6.77	1022	91.09
90	100	8.91	1122	100.00

sexe	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	588	52.41	588	52.41
2	534	47.59	1122	100.00

pratique activités physiques ou sportives(APS)

prataps	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	84	7.50	84	7.50
2	252	22.50	336	30.00
3	477	42.59	813	72.59

4 307 27.41 1120 100.00

Frequency Missing = 2

APS de préférence

apspref	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	520	46.59	520	46.59
2	139	12.46	659	59.05
3	173	15.50	832	74.55
4	94	8.42	926	82.97
5	25	2.24	951	85.22
6	11	0.99	962	86.20
7	154	13.80	1116	100.00

Frequency Missing = 6

importance plaisir APS

apsfun	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	502	44.90	502	44.90
2	292	26.12	794	71.02
3	186	16.64	980	87.66
4	109	9.75	1089	97.41
5	29	2.59	1118	100.00

Frequency Missing = 4

importance compétition APS

apscomp	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	97	8.68	97	8.68
2	112	10.02	209	18.69

3	159	14.22	368	32.92
4	226	20.21	594	53.13
5	524	46.87	1118	100.00

Frequency Missing = 4

importance forme APS

apsform	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	410	36.61	410	36.61
2	418	37.32	828	73.93
3	216	19.29	1044	93.21
4	68	6.07	1112	99.29
5	8	0.71	1120	100.00

Frequency Missing = 2

importance corps APS

apscorps	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	84	7.51	84	7.51
2	182	16.26	266	23.77
3	328	29.31	594	53.08
4	319	28.51	913	81.59
5	206	18.41	1119	100.00

Frequency Missing = 3

importance relaxer APS

apsrelax	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	39	3.49	39	3.49
2	117	10.47	156	13.95

3	227	20.30	383	34.26
4	388	34.70	771	68.96
5	347	31.04	1118	100.00

Frequency Missing = 4

préférence sexe partenaire APS

prefsexe	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	43	3.86	43	3.86
2	120	10.76	163	14.62
3	33	2.96	196	17.58
4	861	77.22	1057	94.80
5	58	5.20	1115	100.00

Frequency Missing = 7

attirance pour compétition APS

compet	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	64	5.72	64	5.72
2	309	27.61	373	33.33
3	395	35.30	768	68.63
4	128	11.44	896	80.07
5	20	1.79	916	81.86
6	203	18.14	1119	100.00

Frequency Missing = 3

attirance pour risque APS

risq	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	754	68.48	754	68.48

2	347	31.52	1101	100.00
---	-----	-------	------	--------

Frequency Missing = 21

perception niveau habileté APS

habil	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	349	31.13	349	31.13
2	669	59.68	1018	90.81
3	103	9.19	1121	100.00

Frequency Missing = 1

réaction face à un nouveau sport

newspe	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	327	29.20	327	29.20
2	672	60.00	999	89.20
3	121	10.80	1120	100.00

Frequency Missing = 2

a eu cours EPH mixte au secondaire

epmix	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	891	79.62	891	79.62
2	228	20.38	1119	100.00

Frequency Missing = 3

aime pas mixité pcq plate

mplate	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	51	5.60	51	5.60
2	207	22.72	258	28.32
3	365	40.07	623	68.39
4	288	31.61	911	100.00

Frequency Missing = 211

aime mixité pcq connaître autre sexe

pconnait	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	412	45.23	412	45.23
2	400	43.91	812	89.13
3	71	7.79	883	96.93
4	28	3.07	911	100.00

Frequency Missing = 211

aime pas mixité pcq se sent dévisagé

mmalaise	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	65	7.15	65	7.15
2	149	16.39	214	23.54
3	357	39.27	571	62.82
4	338	37.18	909	100.00

Frequency Missing = 213

aime mixité pcq meilleur que autre sexe

pmeieur	Frequency	Percent	Cumulative Frequency	Cumulative Percent
---------	-----------	---------	-------------------------	-----------------------

1	144	15.84	144	15.84
2	265	29.15	409	44.99
3	281	30.91	690	75.91
4	219	24.09	909	100.00

Frequency Missing = 213

aime pas mixité pcq compare autre sexe

mcompare	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	168	18.48	168	18.48
2	242	26.62	410	45.10
3	267	29.37	677	74.48
4	232	25.52	909	100.00

Frequency Missing = 213

aime mixité pcq travaille plus

pequil	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	296	32.49	296	32.49
2	322	35.35	618	67.84
3	182	19.98	800	87.82
4	111	12.18	911	100.00

Frequency Missing = 211

aime pas mixité pcq autre sexe se moque

mmoqueri	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	47	5.16	47	5.16
2	80	8.79	127	13.96
3	282	30.99	409	44.95
4	501	55.05	910	100.00

Frequency Missing = 212

aime mixité pcq se dépasser pour impressionner

pdepasse	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	115	12.64	115	12.64
2	264	29.01	379	41.65
3	307	33.74	686	75.38
4	224	24.62	910	100.00

Frequency Missing = 212

aime pas mixité pcq se sent moin bon que autre sexe

mcomplex	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	25	2.74	25	2.74
2	85	9.33	110	12.07
3	345	37.87	455	49.95
4	456	50.05	911	100.00

Frequency Missing = 211

aime mixité pcq travail tet

ptetgars	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	169	18.59	169	18.59
2	295	32.45	464	51.05
3	246	27.06	710	78.11
4	199	21.89	909	100.00

Frequency Missing = 213

pour ou contre la mixité

prefmix	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	508	55.70	508	55.70
2	215	23.57	723	79.28
3	48	5.26	771	84.54
4	21	2.30	792	86.84
5	33	3.62	825	90.46
6	87	9.54	912	100.00

Frequency Missing = 210

pour ou contre la mixité act collective

collmix	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	558	61.05	558	61.05
2	293	32.06	851	93.11
3	63	6.89	914	100.00

Frequency Missing = 208

pour ou contre mixité act à deux

deuxmix	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	758	83.02	758	83.02
2	111	12.16	869	95.18
3	44	4.82	913	100.00

Frequency Missing = 209

pour ou contre mixité act aquatique

aquamix	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	560	61.34	560	61.34
2	253	27.71	813	89.05
3	100	10.95	913	100.00

Frequency Missing = 209

pour ou contre mixité act combat

combmix	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	344	37.68	344	37.68
2	475	52.03	819	89.70
3	94	10.30	913	100.00

Frequency Missing = 209

pour ou contre mixité act expression

expremix	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	640	70.18	640	70.18
2	170	18.64	810	88.82
3	102	11.18	912	100.00

Frequency Missing = 210

pour ou contre mixité act plein air

plenamix	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	875	95.84	875	95.84
2	14	1.53	889	97.37
3	24	2.63	913	100.00

Frequency Missing = 209

pour ou contre act individuelle

indimix	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	830	90.91	830	90.91
2	24	2.63	854	93.54
3	59	6.46	913	100.00

Frequency Missing = 209

préférence professeur de même sexe

sexeprof	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	91	8.14	91	8.14
2	105	9.39	196	17.53
3	105	9.39	301	26.92
4	95	8.50	396	35.42
5	722	64.58	1118	100.00

Frequency Missing = 4

raison de préférence de prof même sexe

raison	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	36	21.82	36	21.82
1	22	13.33	58	35.15
2	12	7.27	70	42.42
3	43	26.06	113	68.48
4	16	9.70	129	78.18
5	6	3.64	135	81.82
6	18	10.91	153	92.73

7	7	4.24	160	96.97
8	1	0.61	161	97.58
9	4	2.42	165	100.00

Frequency Missing = 957

EPH sert pour équipe

ephequip	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	114	10.22	114	10.22
2	154	13.80	268	24.01
3	201	18.01	469	42.03
4	226	20.25	695	62.28
5	185	16.58	880	78.85
6	164	14.70	1044	93.55
7	72	6.45	1116	100.00

Frequency Missing = 6

EPH sert pour défouler

ephdefou	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	184	16.50	184	16.50
2	189	16.95	373	33.45
3	158	14.17	531	47.62
4	156	13.99	687	61.61
5	125	11.21	812	72.83
6	176	15.78	988	88.61
7	127	11.39	1115	100.00

Frequency Missing = 7

EPH sert pour technique

Cumulative Cumulative

ephtech	Frequency	Percent	Frequency	Percent
1	153	13.72	153	13.72
2	187	16.77	340	30.49
3	208	18.65	548	49.15
4	189	16.95	737	66.10
5	203	18.21	940	84.30
6	115	10.31	1055	94.62
7	60	5.38	1115	100.00

Frequency Missing = 7

EPH sert pour compétition

ephcomp	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	42	3.77	42	3.77
2	52	4.66	94	8.43
3	77	6.91	171	15.34
4	116	10.40	287	25.74
5	168	15.07	455	40.81
6	186	16.68	641	57.49
7	474	42.51	1115	100.00

Frequency Missing = 7

EPH sert pour découvrir limites

ephlimit	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	55	4.94	55	4.94
2	133	11.94	188	16.88
3	154	13.82	342	30.70
4	203	18.22	545	48.92
5	204	18.31	749	67.24
6	224	20.11	973	87.34
7	141	12.66	1114	100.00

Frequency Missing = 8

EPH sert pour se détendre

ephrelax	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	195	17.49	195	17.49
2	191	17.13	386	34.62
3	140	12.56	526	47.17
4	117	10.49	643	57.67
5	134	12.02	777	69.69
6	158	14.17	935	83.86
7	180	16.14	1115	100.00

Frequency Missing = 7

EPH sert pour la forme

ephform	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	394	35.30	394	35.30
2	210	18.82	604	54.12
3	176	15.77	780	69.89
4	108	9.68	888	79.57
5	94	8.42	982	87.99
6	86	7.71	1068	95.70
7	48	4.30	1116	100.00

Frequency Missing = 6

intensité max réaction face à la douleur

douleur	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	265	23.68	265	23.68
2	567	50.67	832	74.35

3 287 25.65 1119 100.00

Frequency Missing = 3

attention malaises corporels

ecoutcor	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	785	70.28	785	70.28
2	332	29.72	1117	100.00

Frequency Missing = 5

gène de dire malaises

genedire	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	59	5.28	59	5.28
2	46	4.11	105	9.39
3	893	79.87	998	89.27
4	120	10.73	1118	100.00

Frequency Missing = 4

choix aliments cafétéria

nouritur	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	167	14.94	167	14.94
2	549	49.11	716	64.04
3	402	35.96	1118	100.00

Frequency Missing = 4

importance dans habillement

mode	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	837	75.00	837	75.00
2	279	25.00	1116	100.00

Frequency Missing = 6

sante vs apparence physique

santeapp	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	222	19.87	222	19.87
2	364	32.59	586	52.46
3	531	47.54	1117	100.00

Frequency Missing = 5

importance différences entre sexe

sexediff	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	758	67.74	758	67.74
2	361	32.26	1119	100.00

Frequency Missing = 3

comportement différent selon sexe

diffcmt	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	72	15.22	72	15.22
2	125	26.43	197	41.65
3	166	35.10	363	76.74
4	110	23.26	473	100.00

Frequency Missing = 649

psychologie différente selon sexe

diffpsy	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	53	11.21	53	11.21
2	157	33.19	210	44.40
3	155	32.77	365	77.17
4	108	22.83	473	100.00

Frequency Missing = 649

habillement différent selon sexe

diffmode	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	135	28.60	135	28.60
2	170	36.02	305	64.62
3	108	22.88	413	87.50
4	59	12.50	472	100.00

Frequency Missing = 650

apparence différente selon sexe

diffapp	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	123	26.06	123	26.06
2	166	35.17	289	61.23
3	121	25.64	410	86.86
4	62	13.14	472	100.00

Frequency Missing = 650

fille bonne sport effort féminine

bonpfemi	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	55	4.95	55	4.95
2	194	17.45	249	22.39
3	496	44.60	745	67.00
4	367	33.00	1112	100.00

Frequency Missing = 10

gars pas fait pour danser

exprenga	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	104	9.34	104	9.34
2	197	17.68	301	27.02
3	420	37.70	721	64.72
4	393	35.28	1114	100.00

Frequency Missing = 8

gars bon sport plus populaire avec filles

garsppop	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
1	119	10.69	119	10.69
2	361	32.43	480	43.13
3	367	32.97	847	76.10
4	266	23.90	1113	100.00

Frequency Missing = 9

fille bonne sport plus populaire avec gars

fillpop	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	65	5.83	65	5.83
2	256	22.98	321	28.82
3	504	45.24	825	74.06
4	289	25.94	1114	100.00

Frequency Missing = 8

niveau du secondaire

nivsec	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1	0.09	1	0.09
2	568	50.76	569	50.85
3	3	0.27	572	51.12
4	2	0.18	574	51.30
5	545	48.70	1119	100.00

Frequency Missing = 3

age	Frequency	Percent	Cumulative Frequency	Cumulative Percent
12	1	0.09	1	0.09
13	139	12.43	140	12.52
14	323	28.89	463	41.41
15	71	6.35	534	47.76
16	200	17.89	734	65.65
17	319	28.53	1053	94.19
18	53	4.74	1106	98.93
19	12	1.07	1118	100.00

Frequency Missing = 4

situation familiale

famille	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	818	73.10	818	73.10
2	211	18.86	1029	91.96
3	74	6.61	1103	98.57
4	16	1.43	1119	100.00

Frequency Missing = 3

occupation du père

occupere	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	28	2.72	28	2.72
2	277	26.87	305	29.58
3	126	12.22	431	41.80
4	224	21.73	655	63.53
5	31	3.01	686	66.54
6	81	7.86	767	74.39
7	140	13.58	907	87.97
8	69	6.69	976	94.67
9	55	5.33	1031	100.00

Frequency Missing = 91

occupation de la mère

occumere	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	41	3.83	41	3.83
1	211	19.72	252	23.55
2	151	14.11	403	37.66
3	216	20.19	619	57.85
4	216	20.19	835	78.04
5	6	0.56	841	78.60
6	22	2.06	863	80.65
7	104	9.72	967	90.37

8	24	2.24	991	92.62
9	79	7.38	1070	100.00

Frequency Missing = 52

scolarité du père

scolpere	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	106	9.65	106	9.65
2	279	25.41	385	35.06
3	134	12.20	519	47.27
4	134	12.20	653	59.47
5	266	24.23	919	83.70
6	179	16.30	1098	100.00

Frequency Missing = 24

scolarité de la mère

scolmere	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	88	7.95	88	7.95
2	349	31.53	437	39.48
3	114	10.30	551	49.77
4	200	18.07	751	67.84
5	176	15.90	927	83.74
6	180	16.26	1107	100.00

Frequency Missing = 15

perception niveau économique

nivecono	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	163	14.58	163	14.58


```

5 4 0 0 0 1 0 3 0 0 1 0 0 2 0 1 2 0 1 0 0 0 3 0 0 1 0 0 3 0 0 1
6 3 0 0 1 0 0 4 0 0 0 1 0 1 1 0 4 0 0 0 1 0 3 0 0 1 0 0 3 0 0 1
7 3 0 0 1 0 0 4 0 0 0 1 0 1 1 0 4 0 0 0 1 0 2 0 1 0 0 0 4 0 0 0
8 3 0 0 1 0 0 2 0 1 0 0 0 1 1 0 3 0 0 1 0 0 1 1 0 0 0 0 3 0 0 1

```

```

m m p p p p p p p d d d g g g g g n n n n n s s s s e e e
m m d d d d d d d o o o e e e e e e o o o o a a a a c c c
o o e e e e e e o u u u n n n n n n u u u u n n n n o o o
q q p p p p p p u l l l e e e e e r r r r r t t t t u u u m m
u u a a a a a a l e e e d d d d d i t t t e e e e t c c m o o
0 e e s s s s s s e u u u i i i i i t u u u a a a a c o o o d d
b r r s e e e e e u r r r r r r r r u r r r p p p p o r r d e e
s 4 5 e 1 2 3 4 5 r 1 2 3 e 1 2 3 4 r 1 2 3 p 1 2 3 r 1 2 e 1 2

```

```

1 1 0 4 0 0 0 1 0 1 1 0 0 3 0 0 1 0 3 0 0 1 3 0 0 1 1 1 0 1 1 0
2 1 0 3 0 0 1 0 0 2 0 1 0 3 0 0 1 0 3 0 0 1 2 0 1 0 1 1 0 1 1 0
3 1 0 4 0 0 0 1 0 2 0 1 0 4 0 0 0 1 2 0 1 0 2 0 1 0 2 0 1 1 1 0
4 1 0 2 0 1 0 0 0 3 0 0 1 3 0 0 1 0 1 1 0 0 1 1 0 0 1 1 0 1 1 0
5 0 0 2 0 1 0 0 0 2 0 1 0 3 0 0 1 0 1 1 0 0 2 0 1 0 1 1 0 2 0 1
6 0 0 2 0 1 0 0 0 2 0 1 0 3 0 0 1 0 3 0 0 1 3 0 0 1 1 1 0 1 1 0
7 1 0 3 0 0 1 0 0 2 0 1 0 3 0 0 1 0 2 0 1 0 3 0 0 1 1 1 0 1 1 0
8 0 0 3 0 0 1 0 0 1 1 0 0 3 0 0 1 0 3 0 0 1 3 0 0 1 1 1 0 1 1 0

```

C.3 Résultats programme 3

C.3.a Variables utilisées pour l'analyse

```

keep   prataps1-prataps4  apspref1-apspref6  apsfun1  apsfun5
        apscomp1  apscomp5  apsform1  apsform5  apscor1  apscor5
        apsr1x1  apsr1x5  risq1  risq2  mmalais1  mmalais4  mcompar1          mcompar4  mmoquer1  mmoquer4  pdepase1  pdepase4  dc

```

C.3.b Sortie des 47 variables dichotomiques

```

p p p p a a a a a a      a a a a      m m
r r r r p p p p p p a a p p p p a a a a      m m
a a a a s s s s s s p p s s s s p p p p      a a

```

t t t t p p p p p p s s c c f f s s s s r r l l
 a a a a r r r r r r f f o o o o c c r r i i a a
 0 p p p p e e e e e e u u m m r r o o l l s s i i
 b s s s s f f f f f f n n p p m m r r x x q q s s
 s 1 2 3 4 1 2 3 4 5 6 1 5 1 5 1 5 1 5 1 2 1 4

1 0 0 1 0 0 0 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0 0
 2 0 0 1 0 0 0 1 0 0 0 0 0 0 1 1 0 0 0 0 0 1 0 0 0
 3 0 1 0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1 0 0 0
 4 0 0 1 0 1 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 0 0 0
 5 0 1 0 0 0 0 1 0 0 0 1 0 0 1 0 0 0 0 0 0 0 1 0 0
 6 0 1 0 0 1 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1 0 0 1
 7 0 0 1 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 0 0 1
 8 0 0 1 0 0 0 0 0 1 0 0 0 0 1 1 0 0 0 0 0 1 0 0 0
 9 0 0 1 0 1 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 1 0 0
 10 0 0 1 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 1 0 0
 11 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0
 12 0 0 1 0 1 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0 0
 13 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0 0
 14 0 0 1 0 0 0 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0 1
 15 0 0 1 0 0 1 0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 0 0 0
 16 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 1 0 0 1
 17 0 0 1 0 0 0 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0 0

m m m m p p d d d g g g g n n n s s s e e
 c c m m d d o o o e e e e o o o a a a c c
 o o o o e e u u u n n n n u u u n n n o o
 m m q q p p l l l e e e e r r r t t t u u m m
 p p u u a a e e e d d d d t t t e e e c c o o
 0 a a e e s s u u u i i i i u u u a a a o o d d
 b r r r r e e r r r r r r r r r r r p p p r r e e
 s 1 4 1 4 1 4 1 2 3 1 2 3 4 1 2 3 1 2 3 1 2 1 2

1 0 0 0 1 0 1 1 0 0 0 0 1 0 0 0 1 0 0 1 1 0 1 0
 2 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 1 0 1 0 1 0
 3 0 1 0 1 0 1 0 1 0 0 0 0 1 0 1 0 0 1 0 0 1 1 0
 4 0 0 0 1 0 0 0 0 1 0 0 1 0 1 0 0 1 0 0 1 0 1 0
 5 0 0 0 0 0 0 0 1 0 0 0 1 0 1 0 0 0 1 0 1 0 0 1
 6 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 1 1 0 1 0
 7 0 0 0 1 0 0 0 1 0 0 0 1 0 0 1 0 0 0 1 1 0 1 0
 8 1 0 0 0 0 0 1 0 0 0 0 1 0 0 0 1 0 0 1 1 0 1 0
 9 1 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 1 0 1 0 1 0

```

10 0 0 0 1 0 0 1 0 0 0 0 1 0 1 0 0 0 1 0 1 0 1 0
11 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 1 0 1 0 1 0
12 0 0 0 1 1 0 0 0 1 0 0 1 0 0 0 1 0 0 1 1 0 1 0
13 0 0 0 0 0 0 1 0 0 0 0 1 0 0 1 0 0 1 0 0 1 1 0
14 0 1 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 1 0 1 0 0 1
15 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 1 1 0 1 0
16 0 1 0 1 1 0 0 0 1 0 0 1 0 0 0 1 0 0 1 0 1 1 0
17 1 0 1 0 0 0 0 1 0 0 0 1 0 0 0 1 0 1 0 1 0 1 0

```

C.4 Résultats programme 4

C.4.a Réduction de niveaux chez les filles

Variable	niveau supérieur (=1)	niveau inférieur (=0)
apsfun	4 ou 5	1, 2 ou 3
apsform	4 ou 5	1, 2 ou 3
apscomp	1 ou 2	3, 4 ou 5

C.4.b Réduction de niveaux chez les gars

Variable	niveau supérieur (=1)	niveau inférieur (=0)
apsfun	4 ou 5	1, 2 ou 3
apsform	4 ou 5	1, 2 ou 3
apsrelax	1 ou 2	3, 4 ou 5
mmalaise	1 ou 2	3 ou 4
mmoqueri	1 ou 2	3 ou 4

C.5 Résultats programme 5 : Calcul de la variable summix

```
Obs summix fclus mmalais1 mmalais4 mcompar1 mcompar4 mmoquer1 mmoquer4 pdepase1 pdepase4
```

```

1 2 0001 0 0 0 0 0 1 0 1
2 1 0002 0 0 0 0 0 1 0 0
3 3 0003 0 0 0 1 0 1 0 1
4 1 0004 0 0 0 0 0 1 0 0
5 0 0005 0 0 0 0 0 0 0 0
6 1 0006 0 1 0 0 0 0 0 0

```

7	2	0007	0	1	0	0	0	1	0	0
8	1	0008	0	0	1	0	0	0	0	0
9	1	0009	0	0	1	0	0	0	0	0
10	1	0010	0	0	0	0	0	1	0	0
11	1	0011	0	0	0	0	0	1	0	0
12	2	0012	0	0	0	0	0	1	1	0
13	0	0013	0	0	0	0	0	0	0	0
14	3	0014	0	1	0	1	0	1	0	0
15	0	0015	0	0	0	0	0	0	0	0
16	4	0016	0	1	0	1	0	1	1	0
17	2	0017	0	0	1	0	1	0	0	0
18	4	0018	0	1	1	0	0	1	0	1

C.6 Résultats programme 6 : Sortie des transformations linéaires

Obs prataps1 prataps2 prataps3 prataps4 apspref1 apspref2 apspref3
 apspref4 apspref5 apspref6

1	0.00000	0.00000	0.56348	0.00000	0.00000	0.00000	0.78435	0.00000	0.00000	0.00000
2	0.00000	0.00000	0.56348	0.00000	0.00000	0.00000	0.78435	0.00000	0.00000	0.00000
3	0.00000	0.70783	0.00000	0.00000	0.00000	0.85391	0.00000	0.00000	0.00000	0.00000
4	0.00000	0.00000	0.56348	0.00000	0.64174	0.00000	0.00000	0.00000	0.00000	0.00000
5	0.00000	0.70783	0.00000	0.00000	0.00000	0.00000	0.78435	0.00000	0.00000	0.00000
6	0.00000	0.70783	0.00000	0.00000	0.64174	0.00000	0.00000	0.00000	0.00000	0.00000
7	0.00000	0.00000	0.56348	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Obs apsfun1 apsf45 apsc012 apsc0p5 apsf0m1 apsf045 apscor1 apscor5 apsr1x1 apsr1x5 risq1

1	0.00000	0	0.88522	0.37565	0.60174	0	0.00000	0.00000	0.00000	0.00000	0.00000
2	0.00000	0	0.88522	0.37565	0.60174	0	0.00000	0.00000	0.00000	0.00000	0.37565
3	0.00000	0	0.88522	0.37565	0.60174	0	0.00000	0.00000	0.00000	0.00000	0.37565
4	0.56870	0	0.88522	0.37565	0.00000	0	0.00000	0.00000	0.00000	0.00000	0.37565
5	0.56870	0	0.88522	0.37565	0.00000	0	0.00000	0.00000	0.00000	0.00000	0.00000
6	0.00000	0	0.88522	0.37565	0.60174	0	0.00000	0.00000	0.00000	0.00000	0.37565
7	0.56870	0	0.88522	0.37565	0.00000	0	0.00000	0.00000	0.00000	0.00000	0.37565

Obs risq2 mmalais1 mmalais4 mcompar1 mcompar4 mmoquer1 mmoquer4 pdepase1 pdepase4 douleur1

1	0.62435	0.00000	0.00000	0.00000	0.00000	0.00000	0.64870	0.00000	0.80522	0.71826
---	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------

2	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.64870	0.00000	0.00000	0.00000
3	0.00000	0.00000	0.00000	0.00000	0.82087	0.00000	0.64870	0.00000	0.80522	0.00000
4	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.64870	0.00000	0.00000	0.00000
5	0.62435	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
6	0.00000	0.00000	0.78783	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
7	0.00000	0.00000	0.78783	0.00000	0.00000	0.00000	0.64870	0.00000	0.00000	0.00000

Obs	douleur2	douleur3	ecoucor1	ecoucor2	genedir1	genedir2	genedir3	genedir4	nourtur1	nourtur2
1	0.00000	0.00000	0.25391	0.00000	0	0	0.18609	0.00000	0.00000	0.00000
2	0.42435	0.00000	0.25391	0.00000	0	0	0.18609	0.00000	0.00000	0.00000
3	0.42435	0.00000	0.00000	0.74609	0	0	0.00000	0.89913	0.00000	0.53043
4	0.00000	0.85739	0.25391	0.00000	0	0	0.18609	0.00000	0.88522	0.00000
5	0.42435	0.00000	0.25391	0.00000	0	0	0.18609	0.00000	0.88522	0.00000
6	0.42435	0.00000	0.25391	0.00000	0	0	0.18609	0.00000	0.00000	0.00000
7	0.42435	0.00000	0.25391	0.00000	0	0	0.18609	0.00000	0.00000	0.53043

Obs	nourtur3	mode1	mode2	santeap1	santeap2	santeap3
1	0.58435	0.27826	0.00000	0.00000	0.00000	0.52
2	0.58435	0.27826	0.00000	0.00000	0.60870	0.00
3	0.00000	0.27826	0.00000	0.00000	0.60870	0.00
4	0.00000	0.27826	0.00000	0.87130	0.00000	0.00
5	0.00000	0.00000	0.72174	0.00000	0.60870	0.00
6	0.58435	0.27826	0.00000	0.00000	0.00000	0.52
7	0.00000	0.27826	0.00000	0.00000	0.00000	0.52

C.7 Résultats programme 7 : Calcul des fclus pour les filles

Obs	fclus1	fclus2	fclus3	fclus4	summix
1	0.10526	0.250	0.8	0.2	2
2	0.05263	0.125	0.9	0.2	1
3	0.15789	0.375	0.5	0.4	3
4	0.10526	0.250	0.6	0.3	1
5	0.10526	0.625	0.4	0.1	0
6	0.00000	0.250	0.9	0.2	1
7	0.00000	0.250	0.8	0.3	2
8	0.10526	0.125	0.9	0.1	1
9	0.05263	0.500	0.5	0.3	1

10	0.05263	0.500	0.5	0.2	1
11	0.10526	0.250	0.6	0.3	1
12	0.00000	0.250	0.8	0.4	2
13	0.00000	0.500	0.5	0.2	0
14	0.05263	0.375	0.7	0.4	3
15	0.05263	0.125	0.9	0.1	0
16	0.00000	0.250	0.6	0.8	4
17	0.15789	0.250	0.8	0.1	2
18	0.36842	0.125	0.4	0.3	4
19	0.10526	0.125	0.8	0.1	0
20	0.15789	0.000	0.6	0.5	2

C.8 Résultats programme 8 : Recalcul des fclus pour les filles selon summix

Obs	fclus1	fclus2	fclus3	fclus4	summix
1	0.10526	0.250	0.8	0.20000	2
2	0.05263	0.125	0.9	0.20000	1
3	0.15789	0.375	0.5	0.40000	3
4	0.10526	0.250	0.6	0.30000	1
5	0.13333	0.625	0.4	0.16667	0
6	0.00000	0.250	0.9	0.20000	1
7	0.00000	0.250	0.8	0.30000	2
8	0.10526	0.125	0.9	0.10000	1
9	0.05263	0.500	0.5	0.30000	1
10	0.05263	0.500	0.5	0.20000	1
11	0.10526	0.250	0.6	0.30000	1
12	0.00000	0.250	0.8	0.40000	2
13	0.00000	0.500	0.5	0.33333	0
14	0.05263	0.375	0.7	0.40000	3
15	0.06667	0.125	0.9	0.16667	0
16	0.00000	0.250	0.6	0.80000	4
17	0.15789	0.250	0.8	0.10000	2
18	0.36842	0.125	0.4	0.30000	4
19	0.13333	0.125	0.8	0.16667	0
20	0.15789	0.000	0.6	0.50000	2

C.9 Résultats programme 9

C.9.a L'assignation des 20 premières filles

Obs	fclus1	fclus2	fclus3	fclus4	vmax	fclus
1	0.15789	0.250	0.8	0.10000	0.80000	3
2	0.10526	0.125	0.9	0.10000	0.90000	3
3	0.15789	0.375	0.5	0.30000	0.50000	3
4	0.10526	0.250	0.6	0.20000	0.60000	3
5	0.13333	0.625	0.4	0.00000	0.62500	2
6	0.00000	0.250	0.9	0.10000	0.90000	3
7	0.00000	0.250	0.8	0.20000	0.80000	3
8	0.15789	0.125	0.9	0.00000	0.90000	3
9	0.05263	0.500	0.5	0.20000	0.50000	3
10	0.05263	0.500	0.5	0.10000	0.50000	3
11	0.10526	0.250	0.6	0.20000	0.60000	3
12	0.05263	0.250	0.8	0.30000	0.80000	3
13	0.00000	0.500	0.5	0.16667	0.50000	3
14	0.05263	0.375	0.7	0.30000	0.70000	3
15	0.06667	0.125	0.9	0.00000	0.90000	3
16	0.00000	0.250	0.6	0.70000	0.70000	4
17	0.15789	0.250	0.8	0.00000	0.80000	3
18	0.42105	0.125	0.4	0.20000	0.42105	1
19	0.13333	0.125	0.8	0.00000	0.80000	3
20	0.15789	0.000	0.6	0.40000	0.60000	3

C.9.b Descriptions statistiques des clusters filles

----- fclus=1 -----

Variable	Nb	Moyenne	Écart-type	Minimum	Maximum
----------	----	---------	------------	---------	---------

fclus1	1	0.4210526	.	0.4210526	0.4210526
fclus2	1	0.1250000	.	0.1250000	0.1250000
fclus3	1	0.4000000	.	0.4000000	0.4000000
fclus4	1	0.2000000	.	0.2000000	0.2000000

----- fclus=2 -----

Variable	Nb	Moyenne	Écart-type	Minimum	Maximum
fclus1	94	0.1262038	0.0878580	0	0.3333333
fclus2	94	0.6223404	0.0986786	0.3750000	0.8750000
fclus3	94	0.3936170	0.1024522	0.2000000	0.6000000
fclus4	94	0.1556738	0.1330991	0	0.5000000

----- fclus=3 -----

Variable	Nb	Moyenne	Écart-type	Minimum	Maximum
fclus1	341	0.1068066	0.0812914	0	0.3684211
fclus2	341	0.2866569	0.1451854	0	0.5000000
fclus3	341	0.6595308	0.1382828	0.4000000	1.0000000
fclus4	341	0.1905181	0.1492400	0	0.6000000

----- fclus=4 -----

Variable	Nb	Moyenne	Écart-type	Minimum	Maximum
fclus1	34	0.0968008	0.0856562	0	0.3684211
fclus2	34	0.2720588	0.1392301	0	0.5000000
fclus3	34	0.4500000	0.1022475	0.2000000	0.6000000
fclus4	34	0.5647059	0.0773906	0.5000000	0.7000000

C.9.c Descriptions statistiques des clusters garçons

----- mclus=2 -----

Variable	Nb	Moyenne	Écart-type	Minimum	Maximum
mclus1	231	0.1022311	0.0933877	0	0.4615385
mclus2	231	0.7099567	0.1223342	0.3750000	1.0000000
mclus3	231	0.3743146	0.1602359	0	0.7000000
mclus4	231	0.1225108	0.1236453	0	0.6000000
mclus5	231	0.2474747	0.1770999	0	0.6666667

----- mclus=3 -----

Variable	Nb	Moyenne	Écart-type	Minimum	Maximum
mclus1	99	0.0738151	0.0829064	0	0.3846154
mclus2	99	0.4646465	0.1072030	0	0.6250000
mclus3	99	0.6710438	0.1177690	0.4000000	1.0000000
mclus4	99	0.1552189	0.1279670	0	0.5000000
mclus5	99	0.1936027	0.1627204	0	0.5000000

----- mclus=4 -----

Variable	Nb	Moyenne	Écart-type	Minimum	Maximum
mclus1	9	0.0854701	0.0810840	0	0.2307692
mclus2	9	0.3611111	0.1159951	0.1250000	0.5000000
mclus3	9	0.4222222	0.0971825	0.2000000	0.5000000
mclus4	9	0.5000000	0.0500000	0.4000000	0.6000000
mclus5	9	0.1481481	0.1001542	0	0.3333333

----- mclus=5 -----

Variable	Nb	Moyenne	Écart-type	Minimum	Maximum
mclus1	64	0.0937500	0.0816252	0	0.3076923
mclus2	64	0.4648438	0.1123980	0.1250000	0.6250000
mclus3	64	0.3338542	0.1636760	0	0.6666667
mclus4	64	0.0916667	0.1107932	0	0.5000000
mclus5	64	0.6588542	0.1374679	0.5000000	1.0000000

C.9.d Tableau croisé files

Table of fclus by rnivsec

fclus	rnivsec			
Frequency,				
Percent ,				
Row Pct ,				
Col Pct ,	1,	2,	Total	
1	, 0 ,	1 ,	1	
	, 0.00 ,	0.21 ,	0.21	
	, 0.00 ,	100.00 ,		
	, 0.00 ,	0.41 ,		

2	,	47	,	47	,	94
	,	10.00	,	10.00	,	20.00
	,	50.00	,	50.00	,	
	,	20.98	,	19.11	,	
3	,	157	,	184	,	341
	,	33.40	,	39.15	,	72.55
	,	46.04	,	53.96	,	
	,	70.09	,	74.80	,	
4	,	20	,	14	,	34
	,	4.26	,	2.98	,	7.23
	,	58.82	,	41.18	,	
	,	8.93	,	5.69	,	
Total		224		246		470
		47.66		52.34		100.00

Statistics for Table of fclus by rnivsec

Statistic	DF	Value	Prob
Chi-Square	3	3.1738	0.3656
Likelihood Ratio Chi-Square	3	3.5606	0.3130
Mantel-Haenszel Chi-Square	1	0.2099	0.6468
Phi Coefficient		0.0822	
Contingency Coefficient		0.0819	
Cramer's V		0.0822	

C.9.e Tableau croisé garçons

Table of mclus by rnivsec

mclus	rnivsec		
Frequency,			
Percent ,			
Row Pct ,			
Col Pct ,	1,	2,	Total
2	, 110 ,	121 ,	231

		, 27.30 ,	30.02 ,	57.32
		, 47.62 ,	52.38 ,	
		, 57.89 ,	56.81 ,	
3		, 47 ,	52 ,	99
		, 11.66 ,	12.90 ,	24.57
		, 47.47 ,	52.53 ,	
		, 24.74 ,	24.41 ,	
4		, 6 ,	3 ,	9
		, 1.49 ,	0.74 ,	2.23
		, 66.67 ,	33.33 ,	
		, 3.16 ,	1.41 ,	
5		, 27 ,	37 ,	64
		, 6.70 ,	9.18 ,	15.88
		, 42.19 ,	57.81 ,	
		, 14.21 ,	17.37 ,	
Total		190	213	403
		47.15	52.85	100.00

Statistic	DF	Value	Prob
Chi-Square	3	2.0328	0.5656
Likelihood Ratio Chi-Square	3	2.0516	0.5618
Mantel-Haenszel Chi-Square	1	0.2727	0.6016
Phi Coefficient		0.0710	
Contingency Coefficient		0.0708	
Cramer's V		0.0710	

C.10 Résultats programme 11

C.10.a Tableau anova filles

Dependent Variable: fclus1

Source	DF	Anova SS	Mean Square	F Value	Pr > F
fclus	3	0.49963416	0.16654472	37.23	.0001

rnivsec	1	0.01208027	0.01208027	2.70	0.1010
---------	---	------------	------------	------	--------

Dependent Variable: fclus2

Source	DF	Anova SS	Mean Square	F Value	Pr F
fclus	3	4.34818591	1.44939530	246.61	.0001
rnivsec	1	0.00863691	0.00863691	1.47	0.2260

Dependent Variable: fclus3

Source	DF	Anova SS	Mean Square	F Value	Pr F
fclus	3	1.82488342	0.60829447	247.17	.0001
rnivsec	1	0.03456491	0.03456491	14.04	0.0002

Dependent Variable: fclus4

Source	DF	Anova SS	Mean Square	F Value	Pr F
fclus	3	5.58178837	1.86059612	245.81	.0001
rnivsec	1	0.03224589	0.03224589	4.26	0.0396

C.10.b Tableau anova garçons

Dependent Variable: mclus1

Source	DF	Anova SS	Mean Square	F Value	Pr F
mclus	4	1.02139148	0.25534787	40.32	.0001
rnivsec	1	0.01105601	0.01105601	1.75	0.1872

Dependent Variable: mclus2

Source	DF	Anova SS	Mean Square	F Value	Pr F
mclus	4	1.52567614	0.38141904	91.76	.0001
rnivsec	1	0.00016661	0.00016661	0.04	0.8414

Dependent Variable: mclus3

Source	DF	Anova SS	Mean Square	F Value	Pr F
mclus	4	2.85515830	0.71378957	126.24	.0001
rnivsec	1	0.05800847	0.05800847	10.26	0.0015

Dependent Variable: mclus4

Source	DF	Anova SS	Mean Square	F Value	Pr F
mclus	4	2.57795650	0.64448912	85.34	.0001
rnivsec	1	0.08995359	0.08995359	11.91	0.0006

Dependent Variable: mclus5

Source	DF	Anova SS	Mean Square	F Value	Pr F
mclus	4	2.80414478	0.70103620	106.66	.0001
rnivsec	1	0.00449122	0.00449122	0.68	0.4089

APPENDICE D

PROGRAMME BIPLLOT

```
/******  
* Titre : Biplot linéaire *  
* Définition: Ce programme permet de calculer le biplot linéaire *  
* Création : Session automne 2003 *  
* Créateur : P.Rousseau, S.Émond, M.Côté *  
*****/  
PROC IML;  
%MACRO traitement(nom_donne, list_var, individu);  
  
    USE &nom_donne;  
    SETIN &nom_donne;  
    READ ALL VAR{&list_var} INTO donne [ROWNAME = &individu COLNAME = Variable];  
    /*PRINT donne;  
  
    /* Nombre de lignes de la matrice X correspond au nombres d'observations */  
    /* Nombre de colonnes de la matrice X correspond au nombre de Variables */  
  
    n_obs = NROW(donne);  
    n_var = NCOL(donne);  
  
    /* Calcul de Diff, la matrice de données centrées  
    MOY EST UN VECTEUR LIGNE  
    V_ONE EST UN VECTEUR COLONNE DONT LES ELEMENTS SONT 1 */  
    moy = donne [+,]/n_obs;  
    v_one = J(n_obs,1,1);  
    diff = (donne - v_one * moy);
```

```

/*calcul de la matrice par la methode de Gower*/

    one = J(1,n_obs,1);
    diag_L = 1/sqrt(one*donne);
    L = diag(diag_L);
    print L;
    print one;
    print diag_L;

/* Calcul de Z, la matrice de données centrées et réduites */
    ss = diff[##,];
    std = sqrt(ss/(n_obs-1));
    /*methode de Gower*/
    z=donne*L;
    /*z=diff;*/
    /*z = diff*diag(1/std);*/
    PRINT Z;

/* Décomposition en valeurs singulières de Z */
/* ordonne U Q ET V selon l'ordre des valeurs singulières de Q */
CALL SVD (U, Q, V, Z);

/* IND_COOR: COORDONNEES DES INDIVIDUS
AXE_BI:COORDONNEES DES AXES-BIPLLOT
ACP_VAR: COORDONNEES DES VECTEURS-VARIABLES POUR ACP ESPACE DES VARIABLES*/
IND_COOR = U*DIAG(Q);
AXE_BI = &mult*V;
ACP_VAR = (V*DIAG(Q))/SQRT(n_obs);

/* Goodness of fit */

CUM = J(n_var,1,1);
CUM_PROP = CUM;
/* division par n_obs pour faire la correspondance avec l'acp de Proc Princomp */
Q_2 = Q##2/n_obs;
Sum_tot = Q_2[+,];

DO i = 1 to n_var;
    Q_TEMP = Q_2[1:i,1];
    CUM[I,1] = Q_TEMP[+,1];
    CUM_PROP[I,1] = (CUM[I,1]/SUM_TOT)*100;

```

```

END;

/*print U;*/
title 'valeurs propres et criteres d ajustement';
  PRINT Q Q_2 CUM_PROP;
  PRINT V Q_2 CUM_PROP;
title;

/* Identification des coordonnes
(individus=0, AXES-BIPLLOT=1, ACP DES variables=2 */
IND = J(n_obs,1,0);
Vari = J(n_var,1,1);
Vari2 = J(n_var,1,2);

/* Concaténation horizontale */
  Ind2_coo = ind_coor||ind;
  axe_bi2 = axe_bi||vari;
title 'direction axes biplot';
print axe_bi2;
  acp_var2 = acp_var||vari2;
print acp_var2;

/* Concaténation verticale */
  coor_acp = ind2_coo // axe_bi2 // acp_var2;
print coor_acp;

/* Varname = 'PRIN1 PRIN2 PRIN3' etc.; */
create coordin FROM coor_acp [COLNAME={&l_nompr 'ident'}];
  append from coor_acp;

/* Vecteur des identificateurs en alpha numérique */
noms = &individu//Variable'//variable';
create identch from noms [COLNAME={'Nom'}];
  append from noms;

RUN;

%MEND traitement;

/*****
* Macro imprimer : Permet d'effectuer une impression *
*****/

```

```

QUIT;

%MACRO imprimer(nom_donne, titre, titre2, titre3);

PROC PRINT DATA = &NOM_DONNE;

TITLE &titre; TITLE2 &titre2; TITLE3 &titre3;

RUN;

%MEND IMPRIMER;

/*****
 * Création de l'ensemble de données BiplotB *
 *****/

%MACRO BIPLotr (PR1, PR2);

DATA biplotb;

/* CONCATENATION HORIZONTALE DES COORDONNES ET
   DES LABELS INDICATEURS DE
   POINTS-INDIVIDUS: IDENT=0
   POINTS SUR AXES BIPLot: IDENT=1
   POINTS POUR VECTEURS DES VARAIBLES: IDENT=2
 */

   set identch; set coordin;

RUN;

/*****
 * Impression de l'ensemble des données du BiplotB *
 *****/

%IF &iprint=1 %THEN
%imprimer(BiplotB, "BIPLot SUR ESPACE DES INDIVIDUS ACM", "MATRICES CONCATENEES",
          "IDENT=0:INDIVIDUS;
          IDENT=1:AXES BIPLot;
          IDENT=2:VECTEUR DES VARIABLES DANS ESPACE DUAL");

```

```

/*****
*DATA BIPLLOT:
  balise les axes biplot;
  inscrit texte pour identifier les individus*
*****/

DATA BIPLLOT;

/* IDENT=1: POINTS SUR AXES BIPLLOT alors BALISE DES AXES BIPLLOT
  IDENT=0: POINTS-INDIVIDUS alors INSCRIPTION DU TEXTE
  IDENT=2: POINTS POUR VECTEURS DES VARIABLES alors delete
*/

  set BiplotB;

  /* NECESSAIRE AU SYSTEME SASGRAPH*/
  XSYS = "2";
  YSYS = "2";
  SIZE = 1;

IF IDENT = 1 THEN                                /* points
pour tracer axes biplot */ DO;
      /*&max, &min: nombres maximaux de ecart-types pour axes biplot*/

      X = &PR1;
      Y = &PR2;

      TEXT = SUBSTR(NOM,1,&len_var);

      IF &MINPR1 <= X <= &MAXPR1
          AND
          &MINPR2 <= Y <= &MAXPR2 THEN OUTPUT;

END;

/*points-individus */ IF IDENT=0 AND NOM IN (&l_IND) THEN

DO;

      TEXT = SUBSTR(NOM,1,&LEN_IND);

```

```

/* nombre de caractères (&len_ind) a partir de la premiere colonne*/
Y = &PR2;
X = &PR1;

IF &MINPR1 <= X <= &MAXPR1
    AND
    &MINPR2 <= Y <= &MAXPR2 THEN OUTPUT;

END;

IF IDENT = 2 THEN DELETE;

KEEP X Y TEXT XSYS YSYS SIZE NOM ;

RUN;
%if &iprint=1 %then
%imprimer(Biplot, "BIPLOT: ESPACE DES INDIVIDUS","MATRICE POUR BILOT",' ');
/*****
* Impression du graphique Biplot *
*****/

/* Option d'impression */ GOPTIONS RESET = ALL HSIZE = 8 IN VSIZE
= 8 IN TARGET = PS COLORS =(black);

PROC GPLOT DATA = BILOT;

TITLE1 "BILOT SUR ESPACE DES INDIVIDUS ACM";
title2 &titr_pr; /*titre du graph */

SYMBOL1 V = none;
SYMBOL2 V = none;
SYMBOL3 V = none;
SYMBOL4 V = none;
SYMBOL5 V = none;
SYMBOL6 V = none;
SYMBOL7 V = none;
SYMBOL8 V = none;
SYMBOL9 V = none;
SYMBOL10 V = none;
SYMBOL11 V = none;
SYMBOL12 V = none;

```

```

SYMBOL13 V = none;

axis1 length = 6.5 in order =(&minind TO &maxind BY &incr_ind);
axis2 length = 6.5 in order =(&minind TO &maxind BY &incr_ind);

PLOT Y*X / ANNOTATE = BIPLLOT FRAME VAXIS = AXIS1 HAXIS = AXIS2 HREF = 0 VREF = 0;

RUN;

%MEND BIPLOTR(PR1,PR2);

/*****
*****
FIN DE LA PROCEDURE BIPLLOT
*****
*****
DEBUT DES AJUSTEMENTS POUR LES PARAMETRS DE LA PROCEDURE BIPLLOT
*****/

/* POINTS TERMINAUX DES AXES BIPLLOT
PAR DEFAUT, on fixe AXE=30: les coordonnees des points terminaux
des axes biplots ne peuvent pas dépasser la valeur AXE ou -AXE.
On peut modifier l'emplacement de ces ponts terminaux en modifiant
inc_axe.
{max, min} = points terminaux des axes biplot
abs(max)+abs(min) < abs(minpr1)+abs(maxpr1)*/

%LET AXE=30;%LET MAX=&AXE;%LET MIN=-&AXE;%LET inc_axe=0.001;
/* ne pas modifier la commande precedente*/

/* LABEL DE L'AXE BIPLLOT*/ /* len_var=longueur du nom du label de
la variable (texte sur graph)*/
%LET len_var=8;

/* POINT_INDIVIDUS *****ZOOM*****: IMPRIMER SEULEMENT LES
INDIVIDUS AVEC
MINPR1<= PRIN1<=MAXPR1
MINPR2<= PRIN2<=MAXPR12
extr, extr2= valeurs du min et max*/
%LET extr=0.2;%LET EXTR2=0.2;

```

```

%LET MINPR1=-&extr;%LET MAXPR1=&extr2;
%LET MINPR2=-&extr;%LET MAXPR2=&extr2;

/* len_ind:longueur du texte pour identifier un individu*/
%LET len_ind=1;

/* CARACTERISTIQUES PHYSIQUES DU GRAPHIQUE ESPACE-INDIVIDU
   minind, maxind: etendue des valeurs pour axes cartesiens
   des graphes espace-individu*/
%LET minind=-0.2;%LET maxind=0.2;%LET incr_ind=0.1;

/*CARACTERISTIQUES PHYSIQUES DU GRAPHIQUE ESPACE-DES-VARIABLES
   minvar, maxvar: etendue des valeurs pour axes cartesiens
   des graphes espace-des-variables*/
%LET minvar=-0.2;%LET maxvar=0.2;%LET incr_var=0.1;

/* mult multiplie la longueur des points-catégories
   mult=1 pour variables quantitatives*/
%LET mult=1;

TITLE 'VECTEURS PROPRES DE XX (pxp): U'; title2 'VECTEURS PROPRES
DE XX (nxn): V'; title3 'valeurs propres et valeurs propres
cumulees: Q, cum, cum_prop';

RUN;

/*
*****
*****
                               FIN DE LA GETION DS PARAMETRES
*****
*****
*/ /*GARCONS*/ libname g 'C:\Documents and
Settings\alemieux\Bureau\saslibrary';

%let choix=('1','2','3','4','5');

DATA BPLGTG;

set g.vmaxg; if mclus in &choix;

keep mclus prataps1-prataps4

```

```
apspref1-apspref6
apsfun1 apsfun5
apscomp1 apscomp5
apsform1 apsform5
apscor1 apscor5
apsrlx1 apsrlx5
risq1 risq2
mmalais1 mmalais4
mcompar1 mcompar4
mmoquer1 mmoquer4
pdepase1 pdepase4
douleur1-douleur3
genedir1-genedir4
nourtur1-nourtur3
santeap1-santeap3
ecoucor1 ecoucor2
model mode2;

rename prataps1=Apratap1
      prataps2=Apratap2
      apspref2=Aapspre2
      apspref3=Aapspre3
      apspref5=Aapspre5
      apspref6=Aapspre6
      apsform5=Aapsfor5
      apscor1=Aapscor1
      apscor5=Aapscor5
      apsrlx1=aApsrlx1
      genedir1=Agenedi1
      genedir4=Agenedi4
      nourtur1=Anourtu1

      prataps3=Bpratap3
      apsfun1=Bapsfu1
      risq1=Brisq1
      douleur2=Bdouleu2
      ecoucor1=Becouco1
      genedir3=Bgenedi3
      nourtur2=Bnourtu2
      mode1=Bmode1
```

```

prataps4=Cpratap4
apspref1=Capspre1
apsrlx5=Capsrlx5
mmalais4=Cmmalai4
mcompar4=Cmcompa4
mmoquer4=Cmmoque4
pdepase4=Cpdepas4
douleur3=Cdouleu3
ecoucor2=Cecouco2
santeap1=Csantea1

```

```

apspref4=Dapspre4
apsfun5=Dapsfu5
apscomp1=Dapscom1
mmalais1=Dmmalai1
mcompar1=Dmcompa1
mmoquer1=Dmmoque1
pdepase1=Dpdepas1
genedir2=Dgenedi2
mode2=Dmode2
santeap2=Dsantea2

```

```

apscomp5=Eapscom5
apsform1=Eapsfor1
risq2=Erisq2
douleur1=Edouleu1
nourtur3=Enourtu3
santeap3=Esantea3;

```

```
RUN;
```

```

%LET l_nompr='prin1' 'prin2' 'prin3' 'prin4' 'prin5' 'prin6' 'prin7' 'prin8' 'prin9'
          'prin10' 'prin11' 'prin12' 'prin13' 'prin14' 'prin15' 'prin16' 'prin17'
          'prin18' 'prin19' 'prin20' 'prin21' 'prin22' 'prin23' 'prin24' 'prin25'
          'prin26' 'prin27' 'prin28' 'prin29' 'prin30' 'prin31' 'prin32' 'prin33'
          'prin34' 'prin35' 'prin36' 'prin37' 'prin38' 'prin39' 'prin40' 'prin41'
          'prin42' 'prin43' 'prin44' 'prin45' 'prin46' 'prin47';

```

```

%let iprint=1;

PROC IML;

%traitement(BPLOTG, Apratap1 Apratap2 Aapspre2 Aapspre3
           Aapspre5 Aapspre6 Aapsfor5 Aapscom1 Aapscom5 Aapsrlx1 Agenedi1 Agenedi4 Anourtu1

           Bpratap3 Bapsfu1 Brisq1 Bdouleu2 Becouco1 Bgenedi3 Bnourtu2 Bmode1

           Cpratap4 Capspre1 Capsrlx5 Cmmalai4 Cmcompa4 Cmmoque4 Cpdepas4
           Cdouleu3 Cecouco2 Csantea1

           Dapspre4 Dapsfu5 Dapscom1 Dmmalai1 Dmcompa1 Dmmoque1
           Dpdepas1 Dgenedi2 Dmode2 Dsantea2

           Eapscom5 Eapsfor1 Erisq2 Edouleu1 Enourtu3
           Esantea3, mclus);

QUIT;

title;title2;title3;

%macro partiel(liste,m1,m2,k,n1,n2);
/** liste = numeros i des l_clusi; exemple de numeros des liste:
1 2
    m1, m2    = numero des liste de m1 a m2 pour l_clusi;
    pr1      = composante principale en abcisse
    n1, n2=   composantes principales (de n1 à n2) en ordonnes (de prin&n1 a prin&n2)
*****/

/** execute pour &pr1 (en abcisse) de prin*n1 jusqu'a prin&n2
***/
%do j=&n1 %to &n2;
%put prin&j;

/** imprime chaque biplot avec les individus dans la liste
&l_clus&i
    un graphique par liste

```

```

    pour les liste de &l_clus&m1 jusqu'a l_clus&m2   ***/
%do i=&m1 %to &m2;
/** choix de la liste ***/
    %if %index(&liste,&i) > 0 %then %do;

/** n'imprimera que les individus dans la liste l_ind ***/
    %let l_ind=&l_clus&i;
    %LET titr_pr= 'prin&k prin&j ' ;
    %put &titr_pr;
    %put &l_ind;
/** biplot pour &pr1 en abcisse et prin&j en ordonnee ***/
    %BIPLOTR(prin&k,prin&j);
    %end;
%end;
%end;

%mend partiel;
%let l_clus1='1';
%let l_clus2='2';
%let l_clus3='3';
%let l_clus4='4';
%let l_clus5='5';
%let l_clus6='1' '2' '3' '4' '5';
%let l_clus7=' ' ;

/* Si iprint=0 pas d impression ; si iprint=1 alors impression*/
%let iprint=1;

/* %partiel( 13,13,13,prin1,2,12); */

/*%macro partiel(liste,m1,m2,pr1,n1,n2);*/
/* première entrée: liste i des l_clus&i(dernier chiffre i de
l_clusi:
        exemple : 1 2 5 ou 2
chiffre no 2 et 3 = chiffre début du premier numéro de la liste
        et dernier numéro de la liste
chiffre 4: prin&k en abcisse
deux dernieres entres: chiffres avant-dernier et dernier:
        printj à prinnk en ordonnée*/

%partiel(6,6,6,2,3,3);

```

```
run;

/*FILLES*/

libname y 'C:\Documents and Settings\alemieux\Bureau\saslibrary';

%let choix=('1','2','3','4');

DATA BPLOTF;

set y.vmaxf; if fclus in &choix;

keep fclus prataps1-prataps4
    apspref1-apspref6
    apsfun1 apsfun5
    apscomp1 apscomp5
    apsform1 apsform5
    apscor1 apscor5
    apsrlx1 apsrlx5
    risq1 risq2
    mmalais1 mmalais4
    mcompar1 mcompar4
    mmoquer1 mmoquer4
    pdepase1 pdepase4
    douleur1-douleur3
    genedir1-genedir4
    nourtur1-nourtur3
    santeap1-santeap3
    ecoucor1 ecoucor2
    mode1 mode2;

rename prataps1=Apratap1
    apspref2=Aapspre2
    apspref3=Aapspre3
    apspref4=Aapspre4
    apspref5=Aapspre5
    apspref6=Aapspre6
    apsfun5=Aapsfu5
```

apsform5=Aapsfor5
apscor1=Aapscor1
apsrlx1=Aapsrlx1
mmalais1=Ammalai1
mcompar1=Amcompar1
mmoquer1=Ammoque1
pdepase4=Apdepas4
genedir1=Agenedir1
genedir2=Agenedi2
genedir4=Agenedi4
nourtur1=Anourtu1
santeap1=Asantea1

prataps2=Bpratap2
apspref1=Bapspre1
apsfun1=Bapsfu1
risq2=Brisq2
douleur1=Bdouleu1
nourtur2=Bnourtu2
mode2=Bmode2
santeap2=Bsantea2

prataps3=Cpratap3
apscomp5=Capscom5
apsform1=Capsfor1
risq1=Crisq1
douleur2=Cdouleu2
ecoucor1=Cecouco1
genedir3=Cgenedi3
nourtur3=Cnourtu3
mode1=Cmode1
santeap3=Csantea3

prataps4=Dpratap4
apscomp1=Dapscom1
apscor5=Dapscor5
apsrlx5=Dapsrlx5
mmalais4=Dmmalai4
mcompar4=Dmcompa4
mmoquer4=Dmmoque4
pdepase1=Dpdepas1

```
douleur3=Ddouleu3  
ecoucor2=Decouco2;
```

```
RUN;
```

```
%LET l_nompr='prin1' 'prin2' 'prin3' 'prin4' 'prin5' 'prin6' 'prin7' 'prin8' 'prin9'  
          'prin10' 'prin11' 'prin12' 'prin13' 'prin14' 'prin15' 'prin16' 'prin17'  
          'prin18' 'prin19' 'prin20' 'prin21' 'prin22' 'prin23' 'prin24' 'prin25'  
          'prin26' 'prin27' 'prin28' 'prin29' 'prin30' 'prin31' 'prin32' 'prin33'  
          'prin34' 'prin35' 'prin36' 'prin37' 'prin38' 'prin39' 'prin40' 'prin41'  
          'prin42' 'prin43' 'prin44' 'prin45' 'prin46' 'prin47';
```

```
%let iprint=1;
```

```
PROC IML;
```

```
%traitement(BPLOTf, Apratap1  
            Aapspre2  
            Aapspre3  
            Aapspre4  
            Aapspre5 Aapspre6 Aapsfu5 Aapsfor5 Aapscor1  
            Aapsrlx1  
            Ammalai1  
            Amcompar1  
            Ammoque1  
            Apdepas4  
            Agenedir1  
            Agenedi2  
            Agenedi4  
            Anourtu1  
            Asanteal  
  
            Bpratap2  
            Bapspre1  
            Bapsfu1  
            Brisq2  
            Bdouleu1  
            Bnourtu2  
            Bmode2
```

```

Bsantea2

Cpratap3
Capscom5
Capsfor1
Crisq1
Cdouleu2
Cecouco1
Cgenedi3
Cnourt3
Cmode1
Csantea3

Dpratap4
Dapscom1
Dapscom5
Dapsrlx5
Dmmalai4
Dmcompa4
Dmmoque4
Dpdepas1
Ddouleu3
Decouco2, Fclus);

QUIT;

title;title2;title3;

%macro partiel(liste,m1,m2,k,n1,n2);
/** liste = numeros i des l_clusi; exemple de numeros des liste:
1 2
    m1, m2    = numero des liste de m1 a m2 pour l_clusi;
    pr1      = composante principale en abcisse
    n1, n2= composantes principales (de n1 à n2) en ordonnees (de prin&n1 a prin&n2)
*****/

/** execute pour &pr1 (en abcisse) de prin*n1 jusqu'a prin&n2
***/
%do j=&n1 %to &n2;
%put prin&j;

```

```

/** imprime chaque biplot avec les individus dans la liste
&l_clus&i
    un graphique par liste
    pour les liste de &l_clus&m1 jusqu'a l_clus&m2  ***/
%do i=&m1 %to &m2;
/** choix de la liste ***/
%if %index(&liste,&i) > 0 %then %do;

/** n'imprimera que les individus dans la liste l_ind ***/
%let l_ind=&l_clus&i;
%LET titr_pr= 'prin&k prin&j ' ;
%put &titr_pr;
%put &l_ind;

/** biplot pour &pr1 en abcisse et prin&j en ordonnee ***/
%BIPLOTR(prin&k,prin&j);
%end;
%end;

%mend partiel;
%let l_clus1='1';
%let l_clus2='2';
%let l_clus3='3';
%let l_clus4='4';
%let l_clus5='1' '2' '3' '4';
%let l_clus6=' ';
%let l_clus7='2' '3' '4';
%let l_clus8='3' '4';

/* Si iprint=0 pas d impression ; si iprint=1 alors impression*/
%let iprint=1;

/* %partiel( 13,13,13,prin1,2,12); */

/*%macro partiel(liste,m1,m2,pr1,n1,n2);*/
/* première entrée: liste i des l_clus&i(dernier chiffre i de
l_clusi:
    exemple : 1 2 5 ou 2
    chiffre no 2 et 3 = chiffre début du premier numéro de la liste
    et dernier numéro de la liste
    chiffre 4: prin&k en abcisse
    deux dernieres entres: chiffres avant-dernier et dernier:

```

```
printj à printk en ordonnée*/
```

```
%partiel(5,5,5,2,3,3);
```

```
run;
```

APPENDICE E

BIPLOTS DES FILLES

BIPLOT SUR ESPACE DES INDIVIDUS ACM

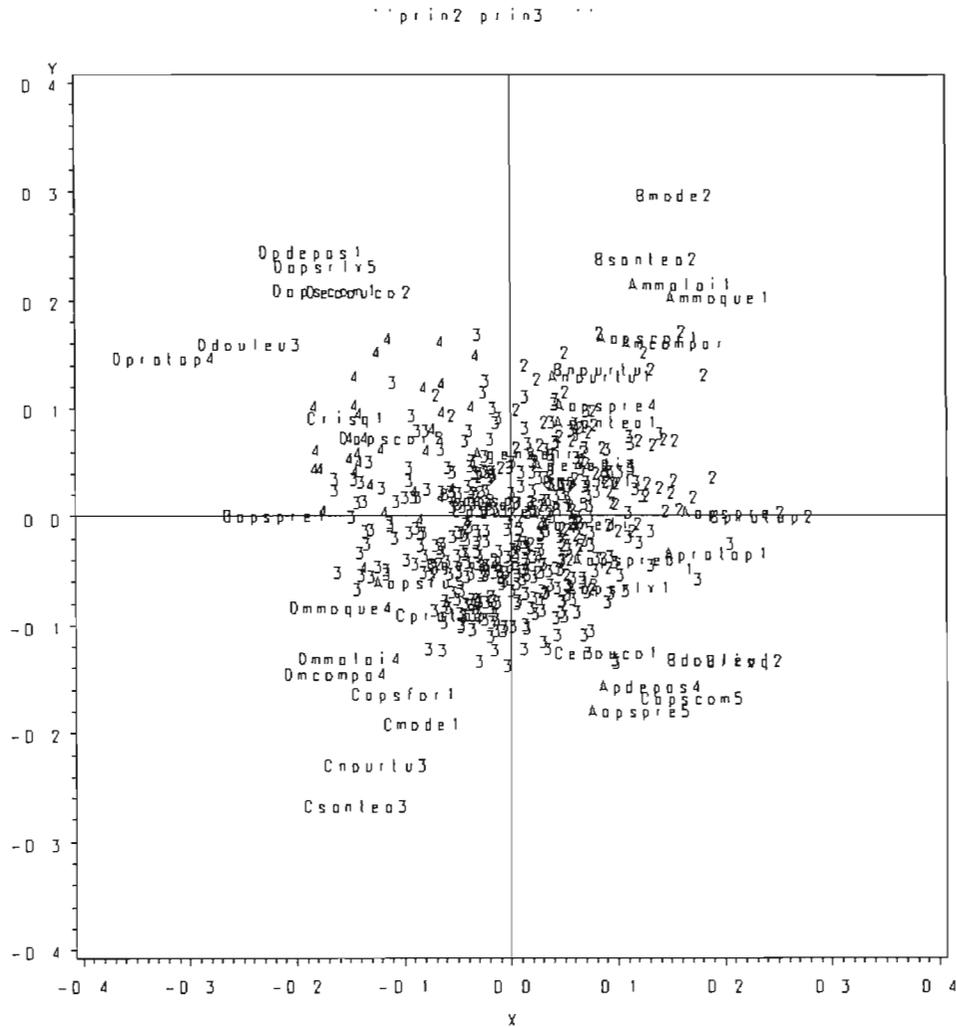


Figure E.1 Biplot des 470 filles identifiées par leur cluster et les 47 variables utilisées pour le clustering

BIPLOT SUR ESPACE DES INDIVIDUS ACM

prin2 prin3

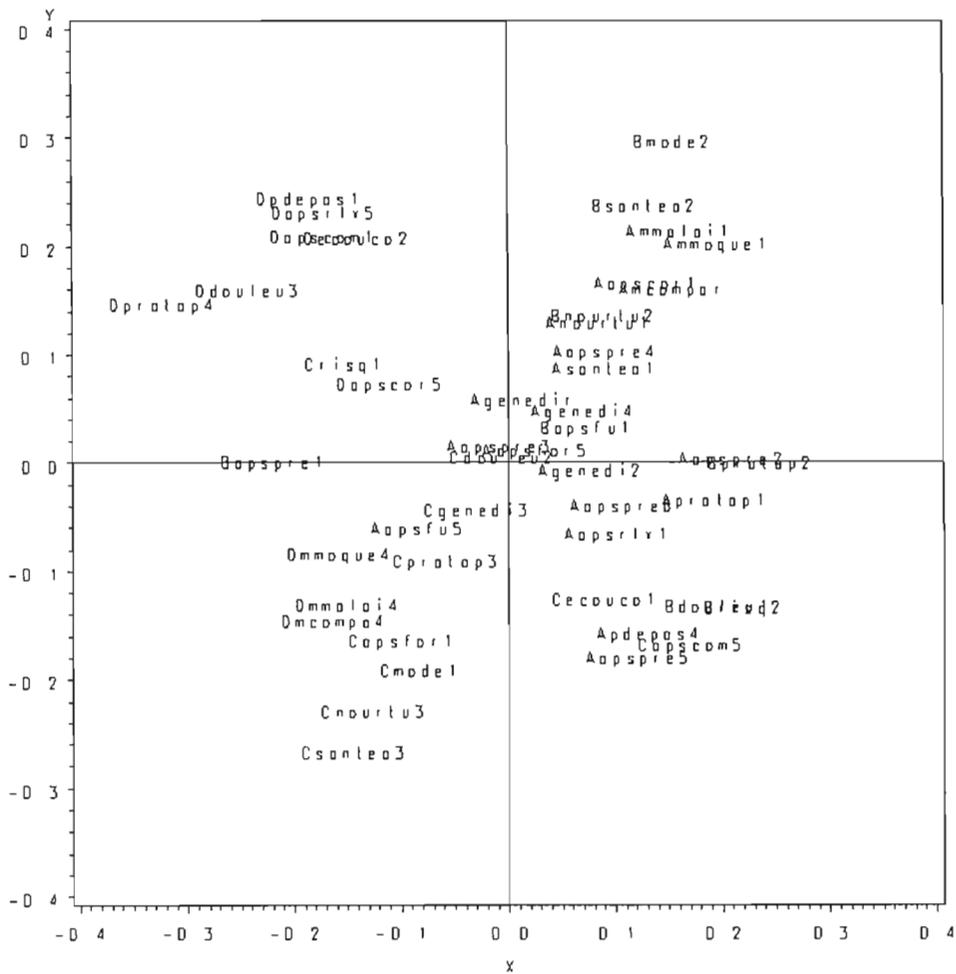


Figure E.3 Biplot construit avec 470 filles et 47 variables : 47 variables affichées

BIPLOT SUR ESPACE DES INDIVIDUS ACM

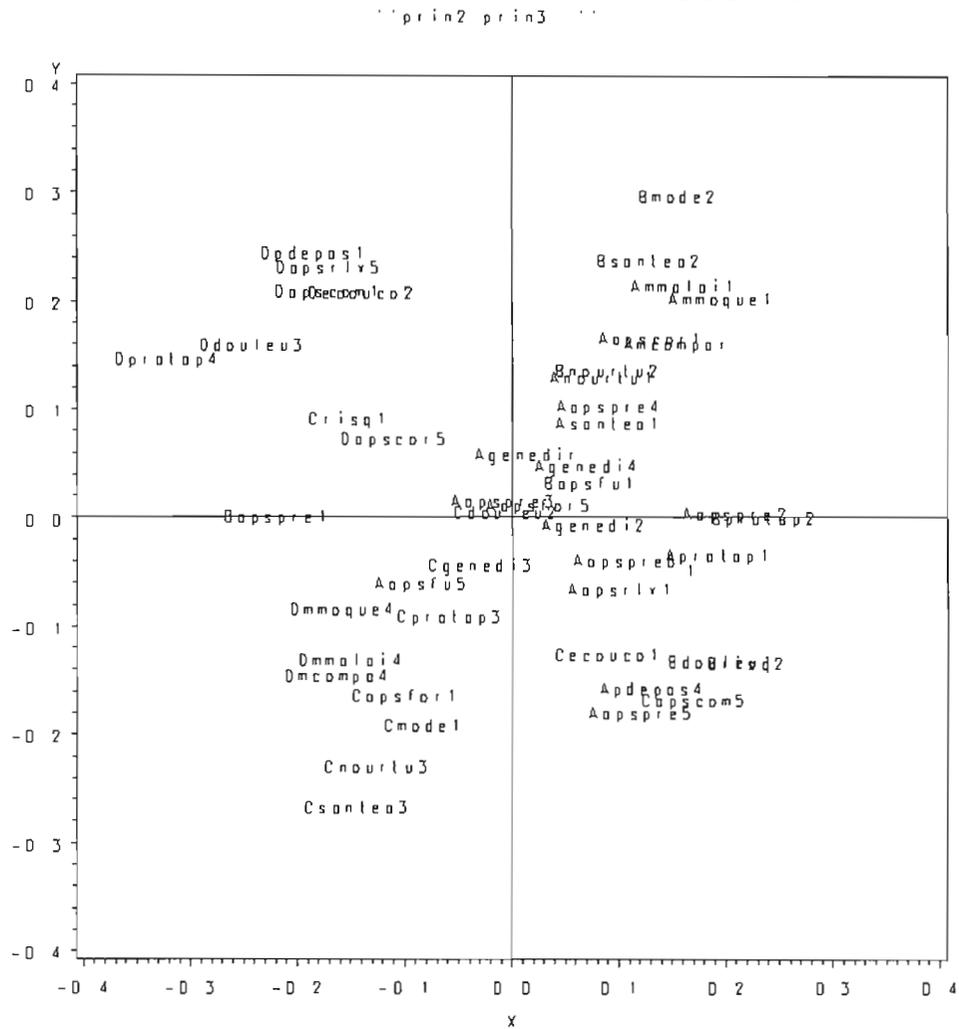


Figure E.4 Biplot construit avec 470 filles et 47 variables : 47 variables et cluster 1 affichés

BIPLOT SUR ESPACE DES INDIVIDUS ACM

prin2 prin3

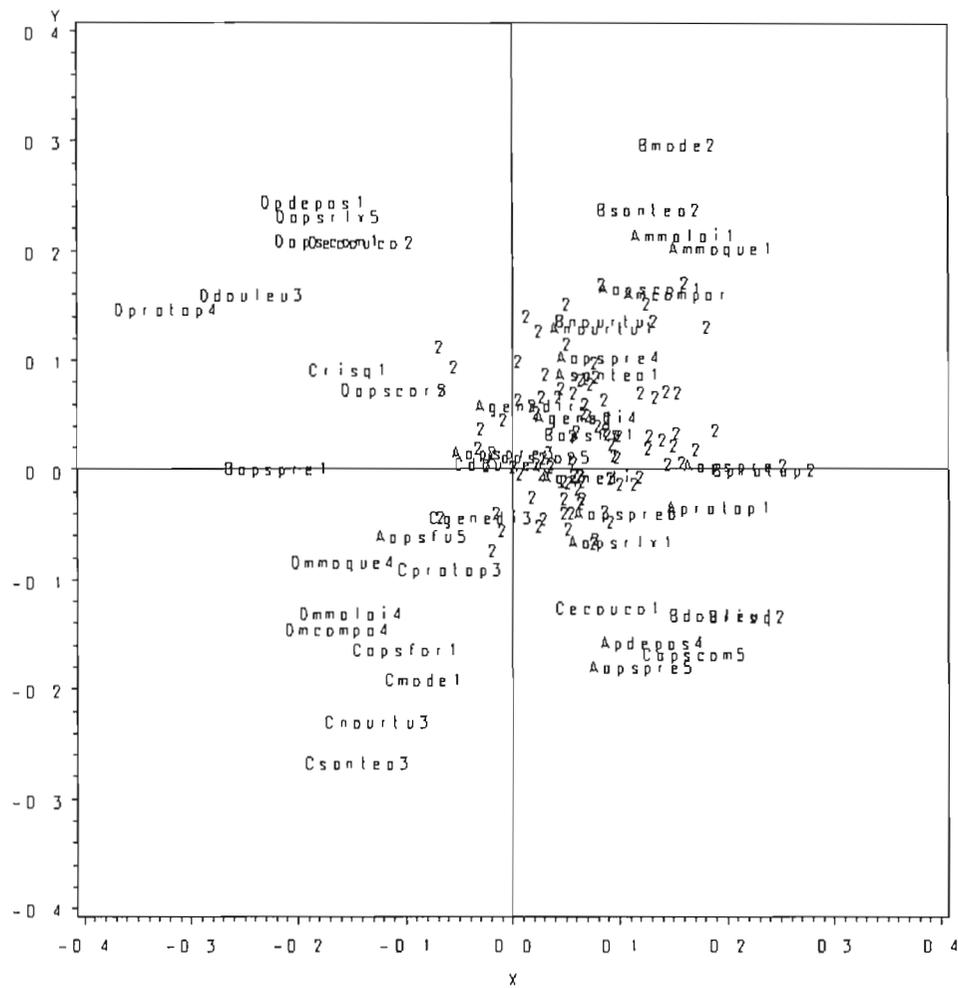


Figure E.5 Biplot construit avec 470 filles et 47 variables : 47 variables et cluster 2 affichés

BIPLOT SUR ESPACE DES INDIVIDUS ACM

prin2 prin3

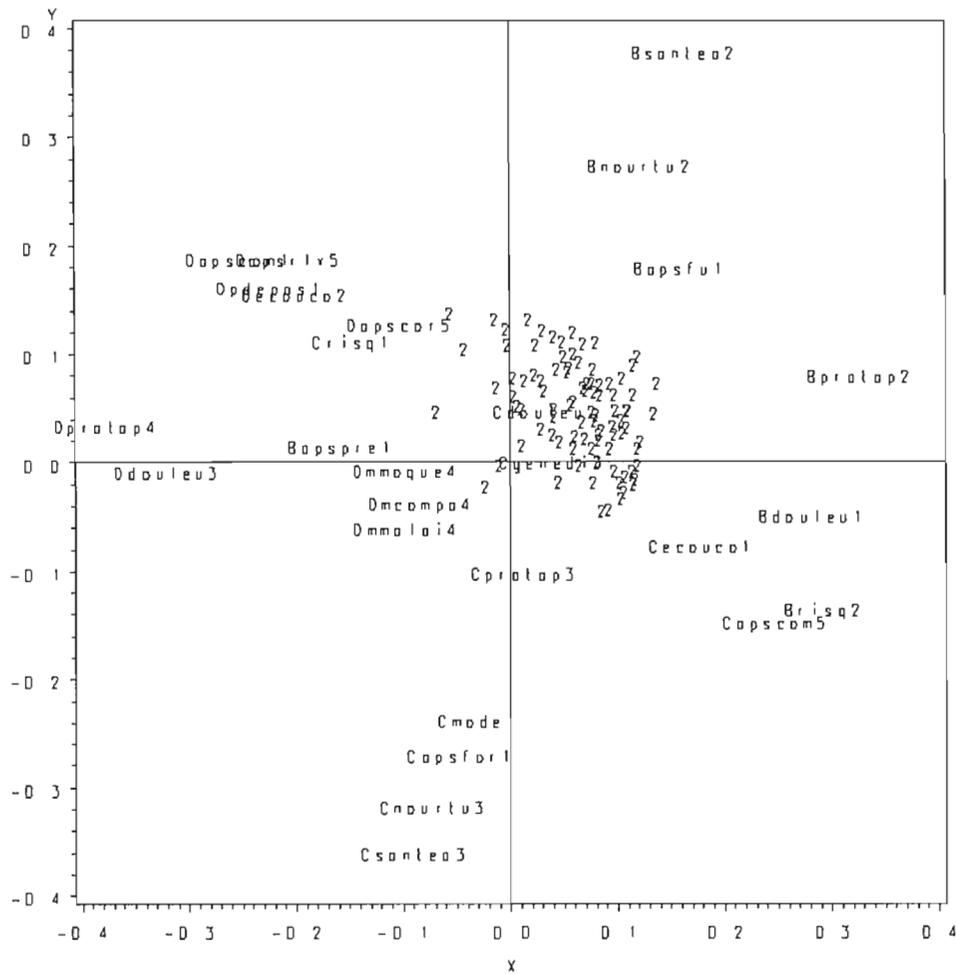


Figure E.6 Biplot construit avec 470 filles et 47 variables : 28 variables et cluster 2 affichés

BIPLOT SUR ESPACE DES INDIVIDUS ACM

prin2 prin3

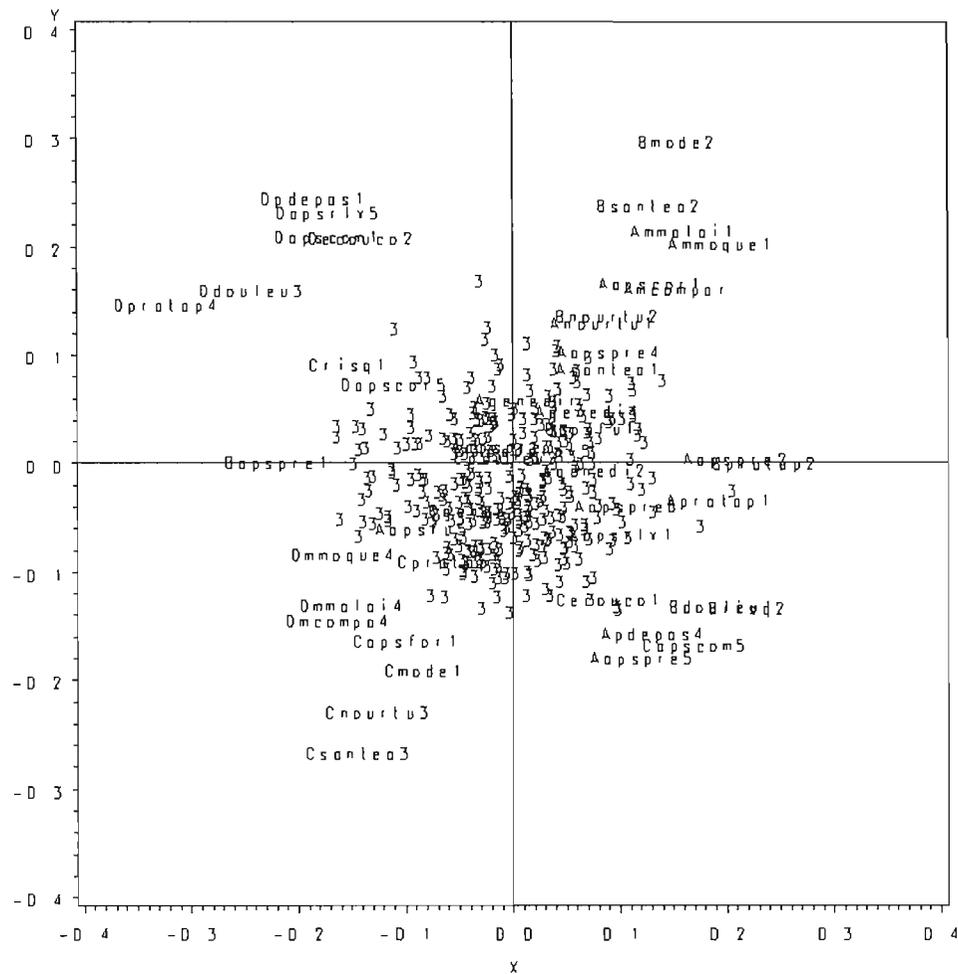


Figure E.7 Biplot construit avec 470 filles et 47 variables : 47 variables et cluster 3 affichés

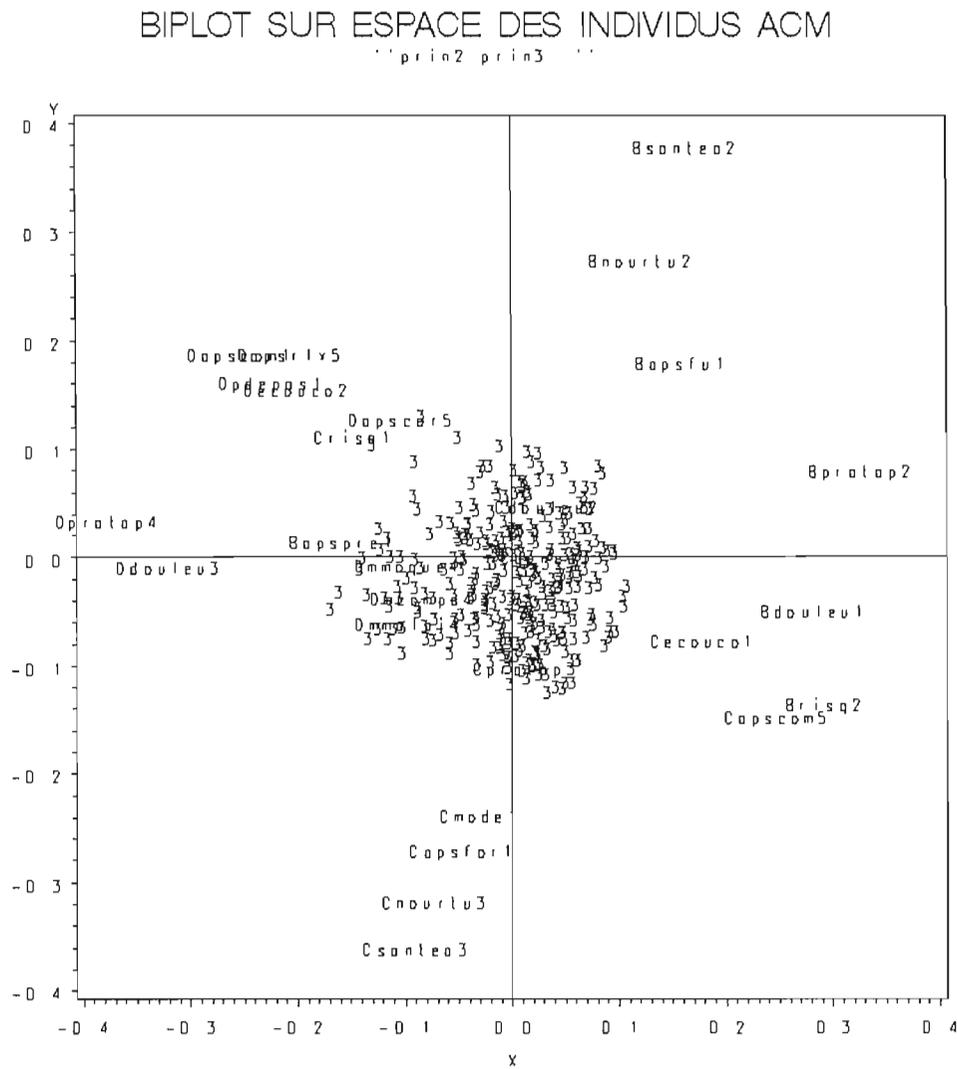


Figure E.8 Biplot construit avec 470 filles et 47 variables : 28 variables et cluster 3 affichés

BIPLOT SUR ESPACE DES INDIVIDUS ACM

'prin2 prin3'

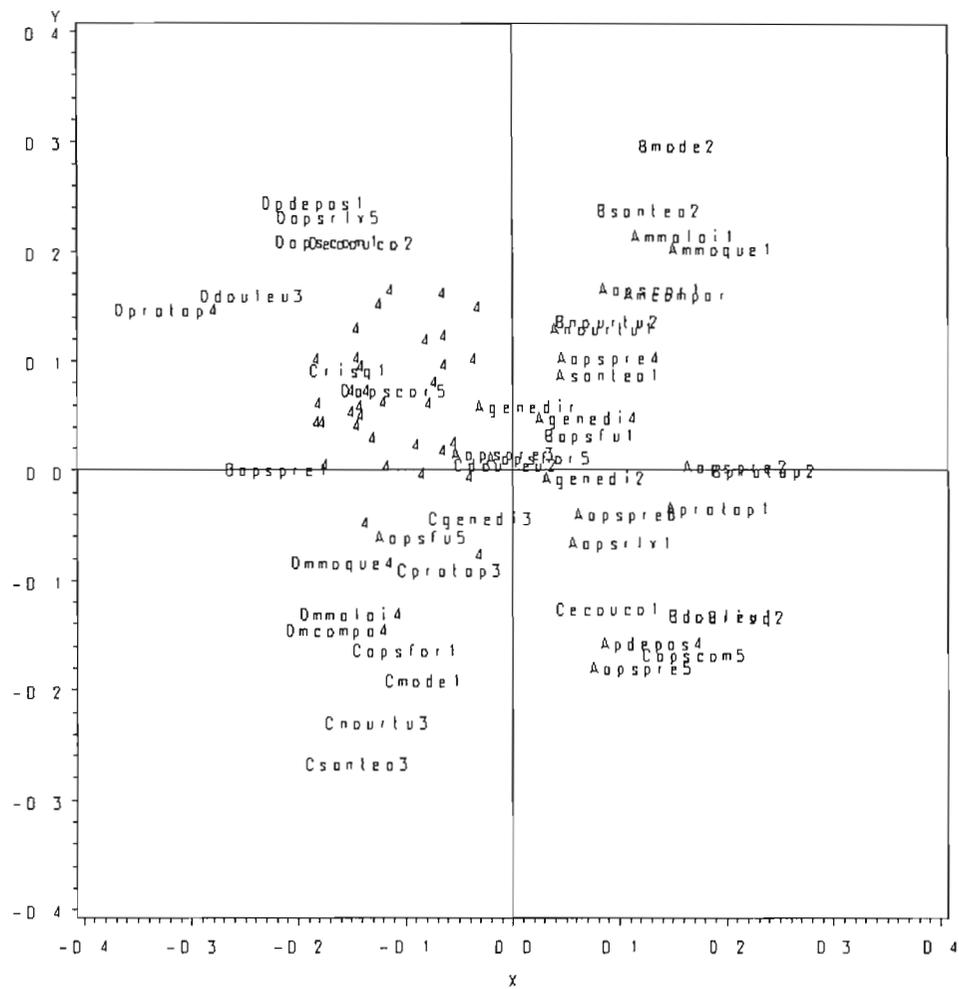


Figure E.9 Biplot construit avec 470 filles et 47 variables : 47 variables et cluster 4 affichés

BIPLOT SUR ESPACE DES INDIVIDUS ACM

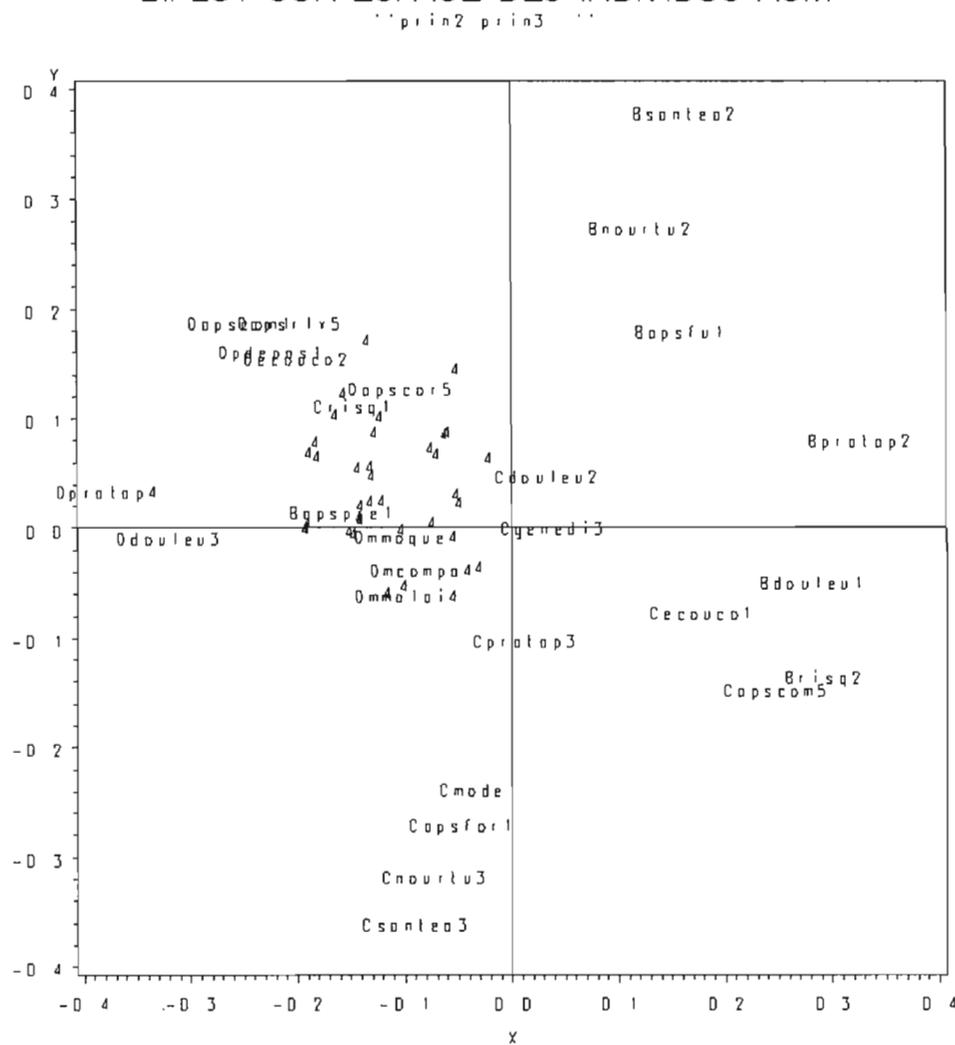


Figure E.10 Biplot construit avec 470 files et 47 variables : 28 variables et cluster 3 affichés

APPENDICE F

BILOTS DES GARÇONS

BIPLOT SUR ESPACE DES INDIVIDUS ACM
 'prin2 prin3'

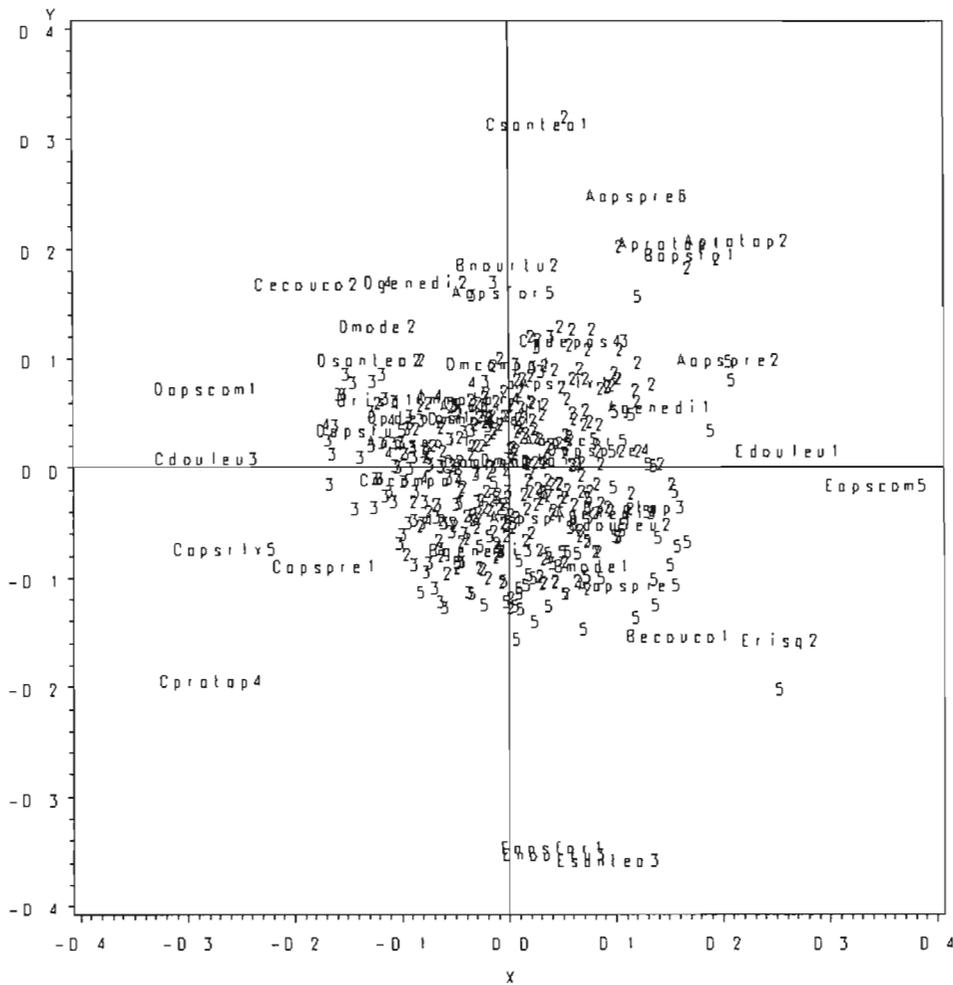


Figure F.1 Biplot des 403 garçons identifiés par leur cluster et les 47 variables utilisées pour le clustering

BIPLOT SUR ESPACE DES INDIVIDUS ACM

''prin2 prin3 ''

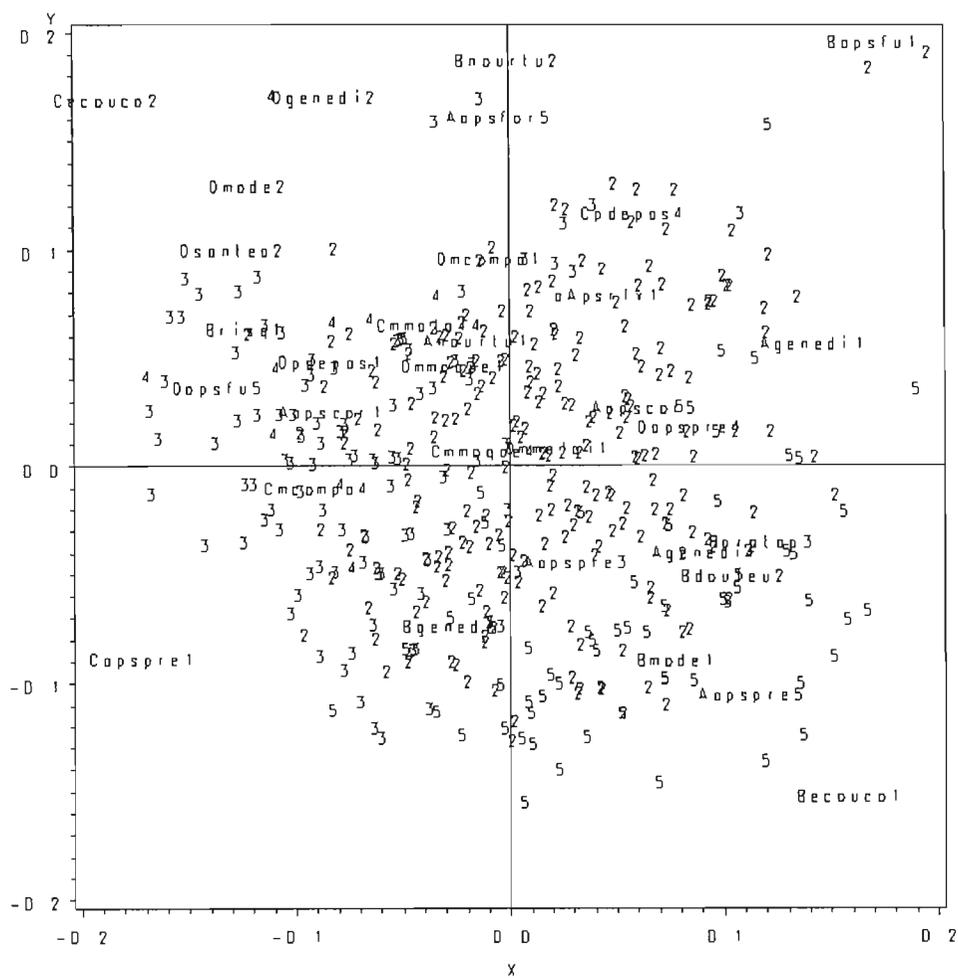


Figure F.2 Biplot des 403 garçons identifiés par leur cluster

BIPLOT SUR ESPACE DES INDIVIDUS ACM

''prin2 prin3 ''

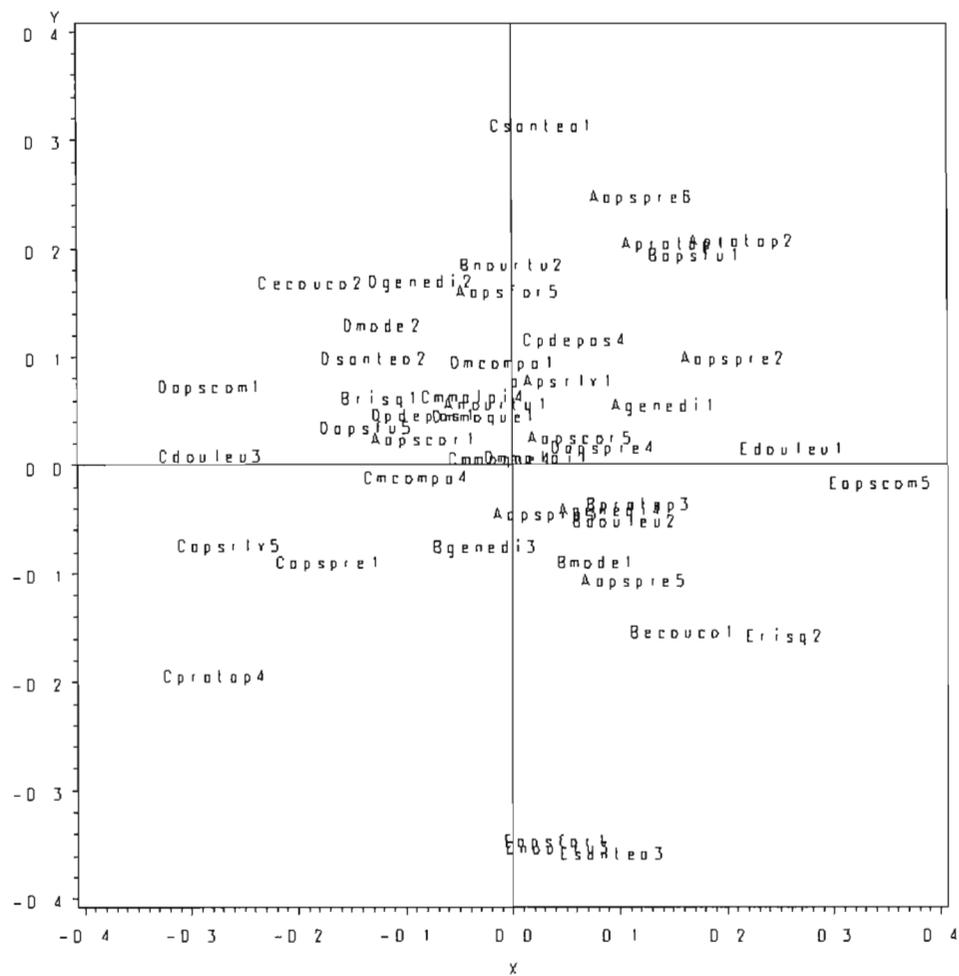


Figure F.3 Biplot construit avec 403 garçons et 47 variables : 47 variables affichées

BIPLOT SUR ESPACE DES INDIVIDUS ACM

prin2 prin3

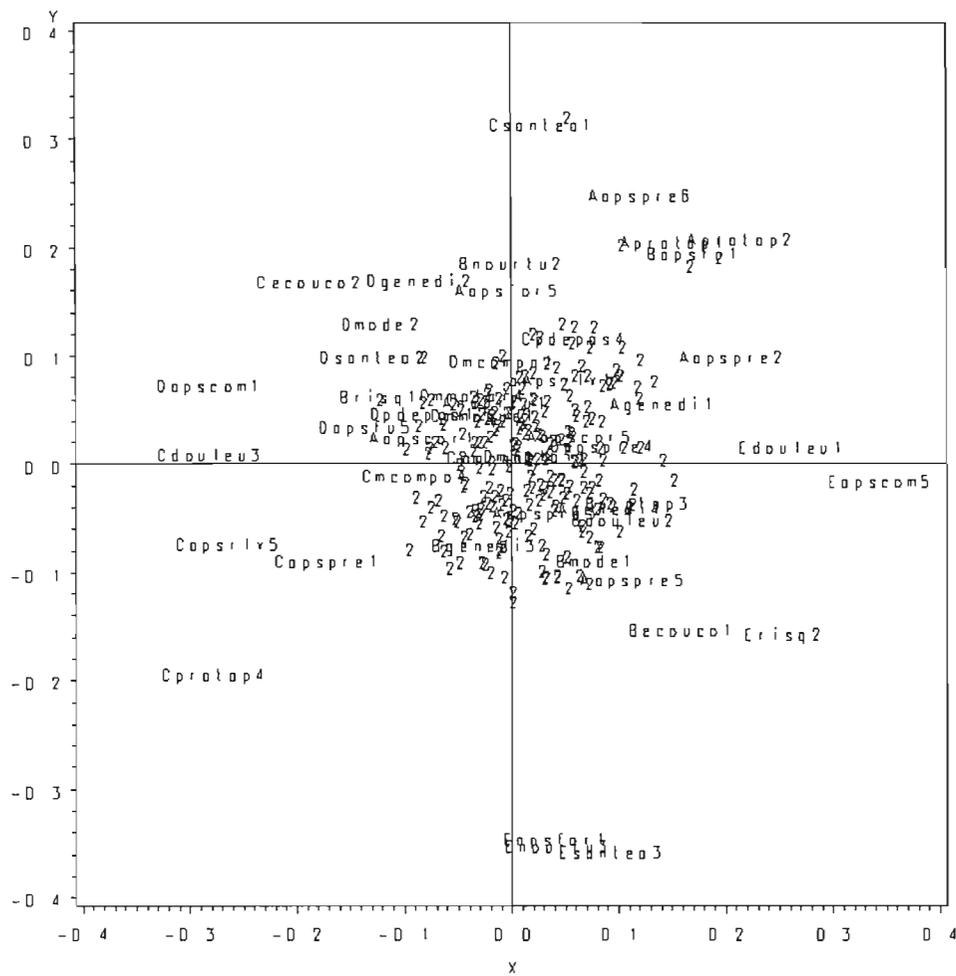


Figure F.5 Biplot construit avec 403 garçons et 47 variables : 47 variables et cluster 2 affichés

BIPLOT SUR ESPACE DES INDIVIDUS ACM

''prin2 prin3 ''

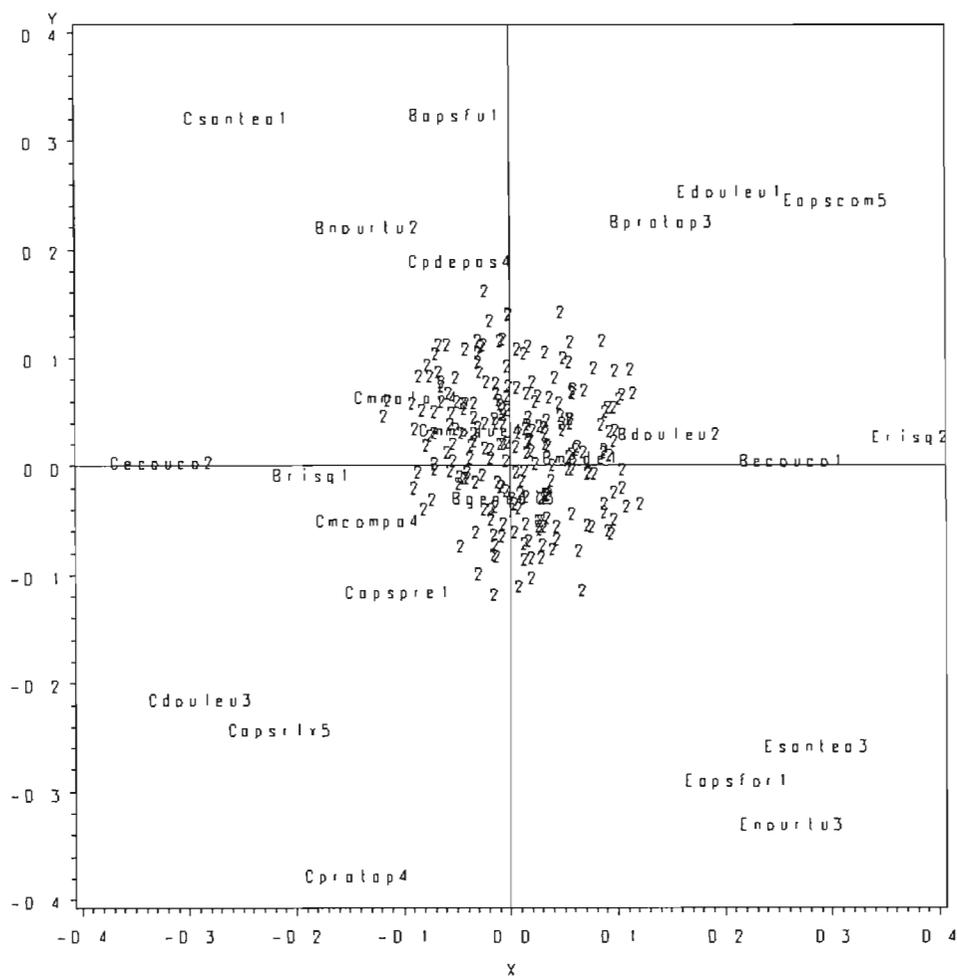


Figure F.6 Biplot construit avec 403 garçons et 47 variables : 24 variables et cluster 2 affichés

BIPLOT SUR ESPACE DES INDIVIDUS ACM

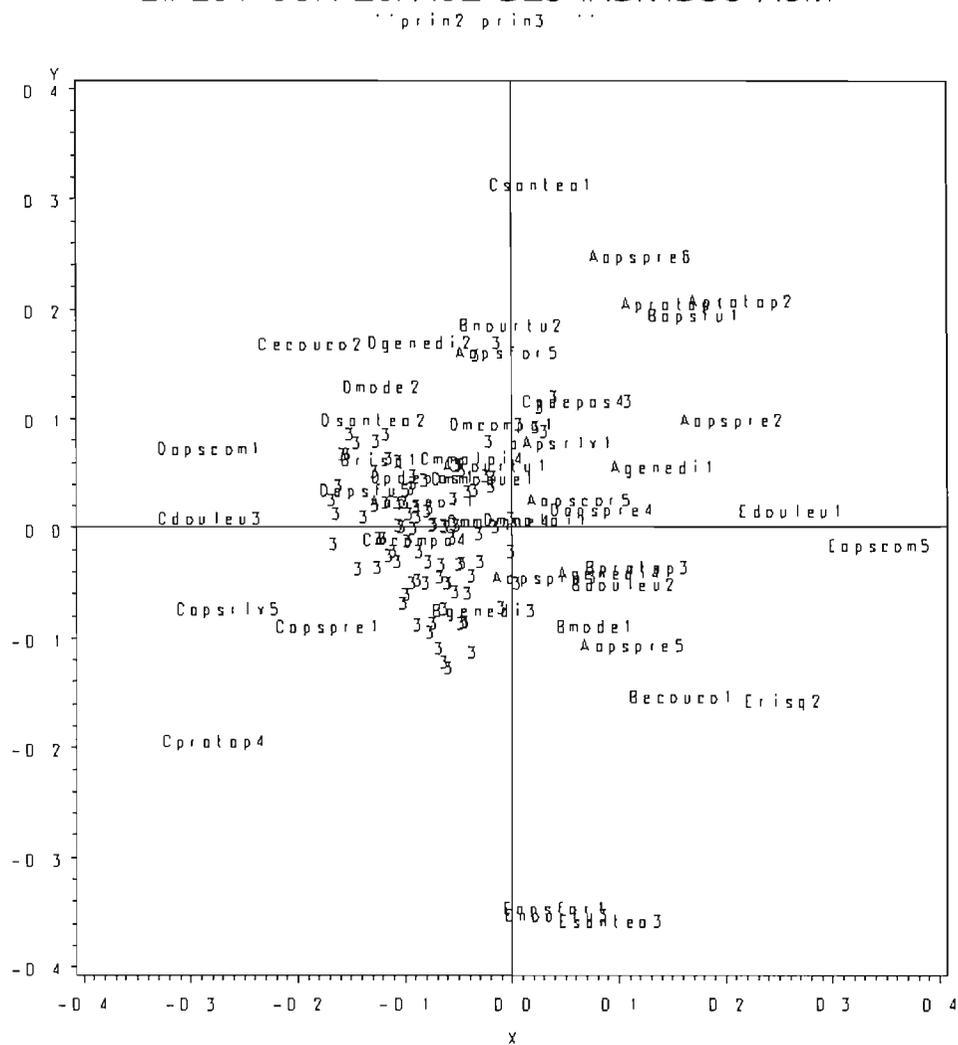


Figure F.7 Biplot construit avec 403 garçons et 47 variables : 47 variables et cluster 3 affichés

BIPLOT SUR ESPACE DES INDIVIDUS ACM

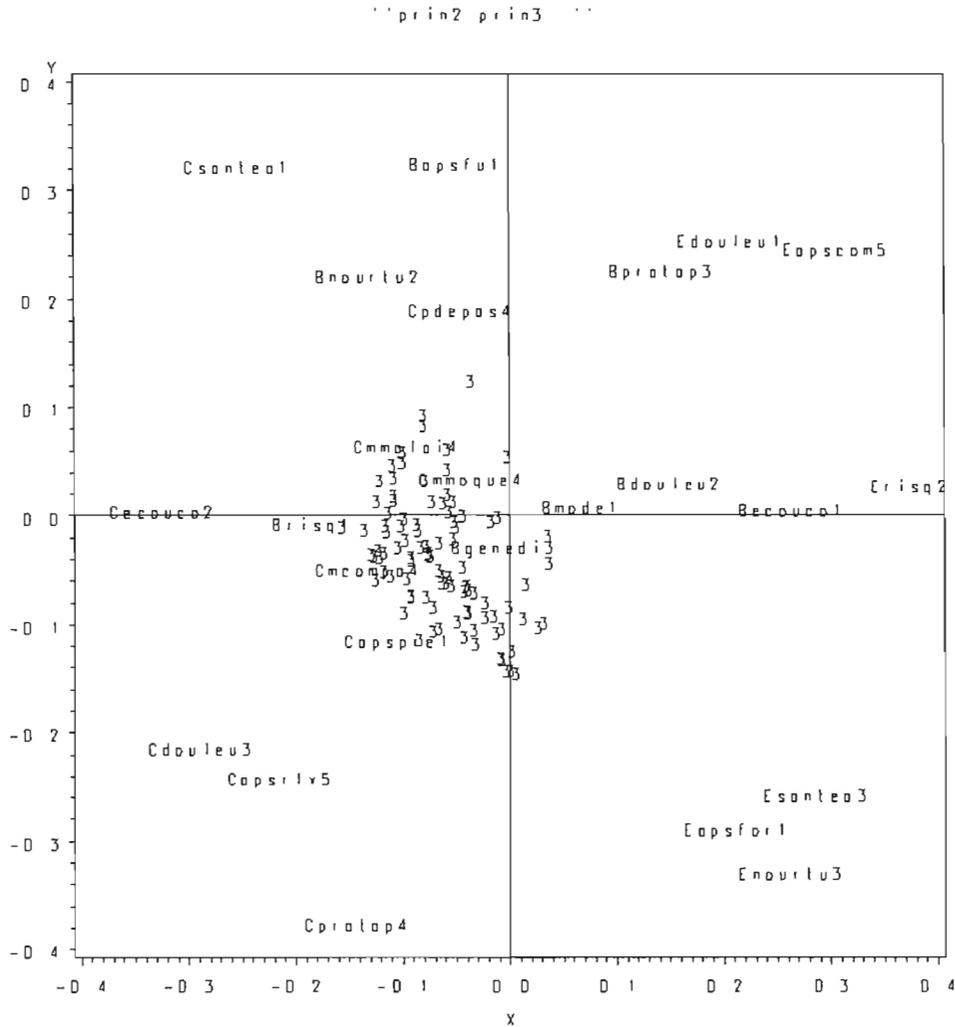


Figure F.8 Biplot construit avec 403 garçons et 47 variables : 24 variables et cluster 3 affichés

BIPLOT SUR ESPACE DES INDIVIDUS ACM

''prin2 prin3 ''

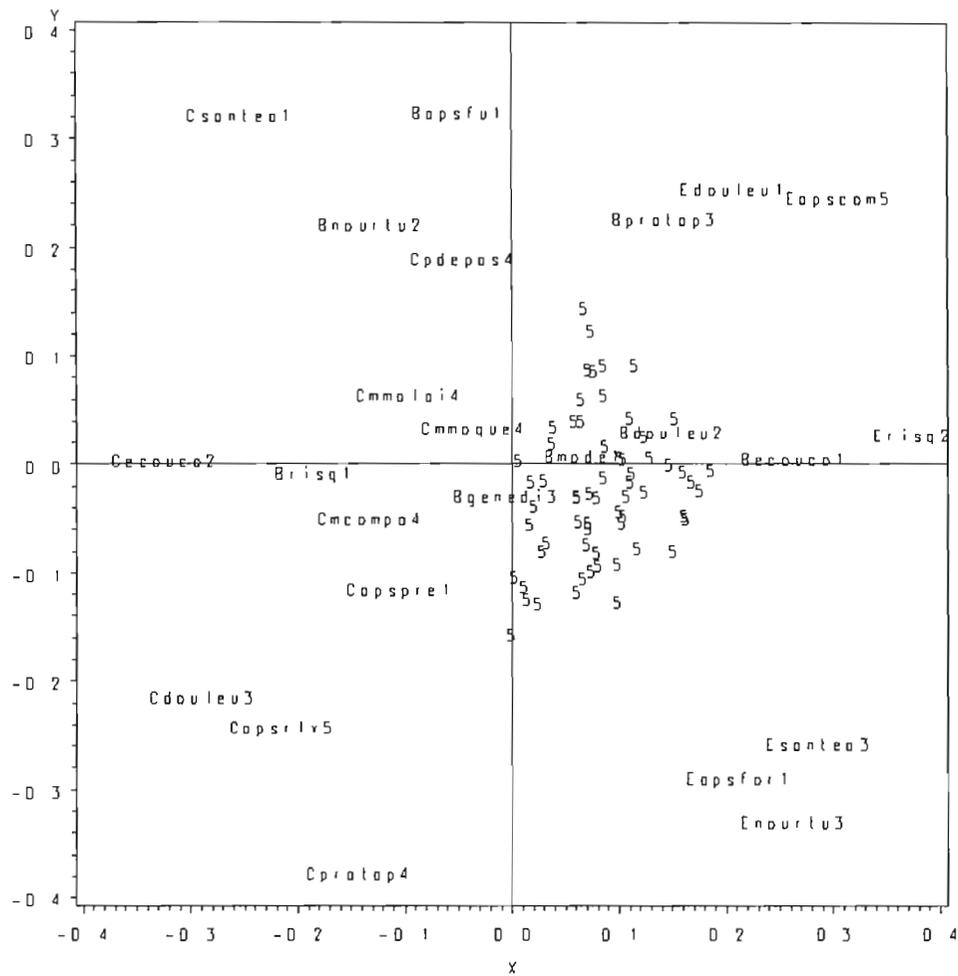


Figure F.10 Biplot construit avec 403 garçons et 47 variables : 24 variables et cluster 5 affichés

APPENDICE G

LES 47 VARIABLES DICHOTOMIQUES

G.1 Description des 47 variables

Tableau G.1 Description des variables du cas réel

Variable	Description
prataps1	pratique activités physiques ou sportives (APS) «peu souvent»
prataps2	pratique APS «parfois»
prataps3	pratique APS «qq fois par semaine»
prataps4	pratique APS «souvent»
apspref1	APS de préférences : «activités collectives»
apspref2	APS de préférences : «activités à deux»
apspref3	APS de préférences : «activités individuelles»
apspref4	APS de préférences : «plein» air»
apspref5	APS de préférences : «marche»
apspref6	APS de préférences : «aime pas faire activités»
apsfun1	importance du plaisir dans les APS «très important»
apsfun45	importance du plaisir dans les APS «peu ou très peu important»
apsform1	importance de la forme dans les APS «très important»
apsfor45	importance de la forme dans les APS «peu ou très peu important»
apscor1	importance du corps dans les APS «très important»
apscor5	importance du corps dans les APS «très peu important»
apscomp1	importance de la compétition dans APS «très important»
apscomp5	importance de la compétition dans APS «très peu important»
apsrlx12	importance de relaxer dans APS «très important ou important»
apsrlx5	importance de relaxer dans APS «très peu important»
genedir1	géné de dire les malaises dans les APS «oui à cause de la perception des autres»
genedir2	géné de dire les malaises dans les APS «oui pq pas se plaindre»
<i>suite à la prochaine page</i>	

<i>suite de la page précédente</i>	
Variable	Description
genedir3	gène de dire les malaises dans les APS «non pcq normal de dire»
genedir4	gène de dire les malaises dans les APS «non pour attirer l'attention»
nourtur1	choix des aliments à la cafétéria «bourratif»
nourtur2	choix des aliments à la cafétéria «santé pas important»
nourtur3	choix des aliments à la cafétéria «bon pour la santé»
risq1	attirance pour le risque «oui»
risq2	attirance pour le risque «non»
douleur1	réaction face à la douleur dans l'intensité «arrête et prudent»
douleur2	réaction face à la douleur dans l'intensité «arrête et content»
douleur3	réaction face à la douleur dans l'intensité «pousse plus»
ecoucor1	attention aux malaises corporels «oui»
ecoucor2	attention aux malaises corporels «non»
mmalais12	aime pas la mixité pcq se sent dévisagé «tout à fait ou plutôt en accord»
mmalais4	aime pas la mixité pcq se sent dévisagé «tout à fait en désaccord»
mcompar1	aime pas la mixité pcq compare aux autres «tout à fait en accord»
mcompar4	aime pas la mixité pcq compare aux autres «tout à fait en désaccord»
mmoquer12	aime pas la mixité pcq autre sexe se moque «tout à fait ou plutôt en accord»
mmoquer4	aime pas la mixité pcq autre sexe se moque «tout à fait en désaccord»
pdepase1	aime la mixité pcq se dépasse pour impressionner «tout à fait en accord»
pdepase4	aime la mixité pcq se dépasse pour impressionner «tout à fait en désaccord»
santeap1	santé vs apparence physique «corps pas important»
santeap2	santé vs apparence physique «plus apparence physique»
santeap3	santé vs apparence physique «plus santé»
mode1	confort d'abord
mode2	mode d'abord

G.2 Clusters des filles

- cluster1 : prataps1 apspref2 apspref3 apspref4 apspref5 apspref6 apsf45 apsf45 apscor1 apsr1x1 mmalais1 mcompar1 mmoquer1 pdepase4 genedir1 genedir2 genedir4 nourtur1 santeap1
- cluster2 : prataps2 apspref1 apsfun1 risq2 douleur1 nourtur2 mode2 santeap2
- cluster3 : prataps3 apscorp5 apsforn1 risq1 douleur2 ecoucor1 genedir3 nourtur3 mode1 santeap3
- cluster4 : prataps4 apscor12 apscor5 apsr1x5 mmalais4 mcompar4 mmoquer4 pdepase1 douleur3 ecoucor2

G.3 Clusters des garçons

- cluster1 : prataps1 prataps2 apspref2 apspref3 apspref5 apspref6 apsf45 apscor1 apscor5 apsr1x12 genedir1 genedir4 nourtur1
- cluster2 : prataps3 apsfun1 risq1 douleur2 ecoucor1 genedir3 nourtur2 mode1

- cluster3 : prataps4 apspref1 apsr1x5 mmalais4 mcompar4 mmoquer4 pdepase4 douleur3 ecoucor2 santeap1
- cluster4 : apspref4 apsfun45 apscomp1 malais12 mcompar1 moquer12 pdepase1 genedir2 mode2 santeap2
- cluster5 : apscomp5 apsform1 risq2 douleur1 nourtur3 santeap3

G.4 Description des clusters des filles

Tableau G.2 Description des clusters des filles

No cluster	Définition des variables
cluster1	pratique activités physiques ou sportives (APS) «peu souvent» APS de préférences : «activités à deux» APS de préférences : «activités individuelles» APS de préférences : «plein» air» APS de préférences : «marche» APS de préférences : «aime pas faire activités» importance du plaisir dans les APS «peu ou très peu important» importance de la forme dans les APS «peu ou très peu important» importance du corps dans les APS «très important» importance de relaxer dans APS «très important» aime pas la mixité pcq se sent dévisagé «tout à fait en accord» aime pas la mixité pcq compare aux autres «tout à fait en accord» aime pas la mixité pcq autre sexe se moque «tout à fait en accord» aime la mixité pcq se dépasse pour impressionner «tout à fait en désaccord» gêné de dire les malaises dans les APS «oui pcq pas se plaindre» gêné de dire les malaises dans les APS «oui pcq perception des autres» gêné de dire les malaises dans les APS «non pour attirer l'attention» choix des aliments à la cafétéria «bourratif» santé vs apparence physique «corps pas important»
cluster2	pratique APS «parfois» APS de préférences : «activités collectives» importance du plaisir dans les APS «très important» attirance pour le risque «non» réaction face à la douleur dans l'intensité «arrête et prudent» choix des aliments à la cafétéria «santé pas important» mode d'abord santé vs apparence physique «plus apparence physique»
<i>suite à la prochaine page</i>	

suite de la page précédente

No cluster	Définition des variables
cluster3	pratique APS «qq fois par semaine» importance de la compétition dans APS «très peu important» importance de la forme dans les APS «très important» attirance pour le risque «oui» réaction face à la douleur dans l'intensité «arrête et content» attention aux malaises corporels «oui» gêné de dire les malaises dans les APS «non pcq normal de dire» choix des aliments à la cafétéria «bon pour la santé» confort avant la mode santé vs apparence physique «plus santé»
cluster4	pratique APS «souvent» importance de la compétition dans APS «très important ou important» importance du corps dans les APS «très peu important» importance de relaxer dans APS «très peu important» aime pas la mixité pcq se sent dévisagé «tout à fait en désaccord» aime pas la mixité pcq compare aux autres «tout à fait en désaccord» aime pas la mixité pcq autre sexe se moque «tout à fait en désaccord» aime la mixité pcq se dépasse pour impressionner «tout à fait en accord» réaction face à la douleur dans l'intensité «pousse plus» attention aux malaises corporels «non»

G.5 Description des clusters des garçons

Tableau G.3 Description des clusters des garçons

No cluster	Définition des variables
cluster1	pratique activités physiques ou sportives (APS) «peu souvent» ou «parfois» APS de préférences : «activités à deux» APS de préférences : «activités individuelles» APS de préférences : «marche» APS de préférences : «aime pas faire activités» importance de la forme dans les APS «peu ou très peu important» importance du corps dans les APS «très important» importance du corps dans les APS «très peu important» importance de relaxer dans APS «très important» ou «important» gêné de dire les malaises dans les APS «oui pcq pas se plaindre» gêné de dire les malaises dans les APS «non pour attirer l'attention» choix des aliments à la cafétéria «bourratif»

suite à la prochaine page

<i>suite de la page précédente</i>	
No cluster	Définition des variables
cluster2	pratique APS «qq fois par semaine» importance du plaisir dans les APS «très important» attirance pour le risque «oui» réaction face à la douleur dans l'intensité «arrête et content» attention aux malaises corporels «oui» gêné de dire les malaises dans les APS «non pcq normal de dire» choix des aliments à la cafétéria «santé pas important» confort avant mode
cluster3	pratique APS «souvent» APS de préférences : «activités collectives» importance de relaxer dans APS «très peu important» aime pas la mixité pcq se sent dévisagé «tout à fait en désaccord» aime pas la mixité pcq compare aux autres «tout à fait en désaccord» aime pas la mixité pcq autre sexe se moque «tout à fait en désaccord» aime la mixité pcq se dépasse pour impressionner «tout à fait en désaccord» réaction face à la douleur dans l'intensité «pousse plus» attention aux malaises corporels «non» santé vs apparence physique «corps pas important»
cluster4	APS de préférences : «plein» air» importance du plaisir dans les APS «peu ou très peu important» importance de la compétition dans APS «très important» aime pas la mixité pcq se sent dévisagé «tout à fait ou plutôt en accord» aime pas la mixité pcq compare aux autres «tout à fait en accord» aime pas la mixité pcq autre sexe se moque «tout à fait ou plutôt en accord» aime la mixité pcq se dépasse pour impressionner «tout à fait en accord» gêné de dire les malaises dans les APS «oui pcq pas se plaindre» mode d'abord santé vs apparence physique «plus apparence physique»
cluster5	importance de la compétition dans APS «très peu important» importance de la forme dans les APS «très important» attirance pour le risque «non» réaction face à la douleur dans l'intensité «arrête et prudent» choix des aliments à la cafétéria «bon pour la santé» santé vs apparence physique «plus santé»

BIBLIOGRAPHIE

- Agresti, A. 1984. *Analysis of ordinal categorical data*. Toronto : J. Wiley. 287 p.
- . 1990. *Categorical data analysis*. Toronto : J. Wiley. 558 p.
- . 1996. *An introduction to categorical data analysis*. Toronto : J. Wiley. 290 p.
- . 1997. *Statistical methods for the social sciences*. Toronto : J. Wiley.
- Der, G. et B. S. Everitt. 2002. *A Handbook of Statistical Analyses using SAS*. London : Chapman and Hall, 2^e édition. 360 p.
- Everitt, B. S. 1993. *Applied multivariate data analysis*. London : Arnold, 3^e édition. 456 p.
- Everitt, B. S. et G. Dunn. 2001. *Applied multivariate data analysis*. London : Arnold; New York : Oxford University Press, 2^e édition. 342 p.
- Gower, J. C. et D. J. Hand. 1996. *Biplots*. London : Chapman and Hall. 277 p.
- Lebart, L., A. Morineau, et M. Piron. 2000. *Statistique exploratoire multidimensionnelle*. Paris : Dunod, 3^e édition. 439 p.
- Leonard, T. 2000. *A course in categorical data analysis*. London : Chapman and Hall. 183 p.