

Exemple d'analyse des données à partir de données réelles :

Analyse de la production agricole départementale

Arnaud MARTIN
Septembre 2004

Réf. : 1464 DE/SCO – S2(module)

Avertissement

Ce document a pour but de présenter les différentes analyses étudiées en cours (analyse en composantes principales, analyse factorielle des correspondances, analyse factorielle des correspondances multiples et classification hiérarchique). Dans ce document, nous nous appuyons sur des données réelles, et nous montrons comment il est possible d'interpréter les résultats des analyses. Le fait de travailler avec des données particulières ne nous a pas permis de présenter toutes les options auxquelles vous avez accès sous SPAD. A vous de vérifier si les approches proposées dans ce document peuvent être appliquées à vos données et s'il n'y a pas d'autres études envisageables pour vos données.

Table des matières

1.	Description des données.....	6
2.	Objectif de l'étude.....	8
3.	Analyse en composantes principales.....	9
4.	Classification hiérarchique.....	17
5.	Analyse factorielle des correspondances.....	22
6.	Analyse des correspondances multiples.....	28
7.	Conclusion.....	33
8.	Annexes.....	34

1. Description des données

Les données concernent des informations sur les départements français. Elles ont été recueillies à partir de plusieurs sites Internet tels que les sites de l'INSEE, du ministère de l'industrie et de l'agriculture. Les individus sont donc les départements français avec les départements d'outre-mer et les territoires d'outre-mer. Ils sont au nombre de 106. Les variables sont constituées d'informations sur la population en 1999 et 1990 et sa répartition en âge, sexe, selon le niveau d'étude, la nationalité et l'activité, sur la géographie (superficie, altitude maximale, proximité de la mer, d'une frontière), et sur la production agricole (nombre d'exploitations et la superficie, pour le blé tendre et les bovins). Un extrait des données est fourni en Annexe (Tableau A.1) ainsi que les principales statistiques des variables continues (Tableau A.2). Afin de faciliter la lecture du rapport une carte en Figure 1.1 représente les départements de métropole, on pourra s'y reporter.



Figure 1.1 : Carte des départements français issue du Quid.

2. Objectif de l'étude

L'objectif de l'étude est de faire ressortir les caractéristiques agricoles des départements français. A ce titre nous restreindrons souvent les études aux variables concernées par cette problématique.

Nous étudierons tout d'abord les corrélations entre les variables agricoles et certaines variables caractéristiques de la population et de la géographie des départements pour aboutir à l'étude des individus dans les premiers plans factoriels issus de l'analyse en composantes principales. A l'issue de cette analyse nous effectuerons une classification hiérarchique pour étudier les regroupements des départements en classe. Nous étudierons ensuite le nombre d'exploitations (avec ses distinctions pour l'élevage bovin, la culture du blé, etc.) et le nombre d'essais OGM en fonction des caractéristiques de la population à l'aide d'une analyse factorielle des correspondances à la suite de laquelle une classification hiérarchique sera réalisée. Pour finir nous étudierons la production agricole de manière plus grossière après avoir transformé les variables continues en variables nominales en deux classes, de nouveau en fonction de la géographie caractérisée par la variable *montagneux*. Une classification hiérarchique sera réalisée à la suite de cette analyse.

3. Analyse en composantes principales

Nous allons effectuer une ACP sur les variables données dans le Tableau 3.1, pour dans un premier temps dégager leur corrélation. Nous étudierons ensuite les ressemblances entre individus projetés dans les premiers plans factoriels, et nous caractériserons ces ressemblances.

Libellé de la variable	Effectif	Moyenne	Ecart-type	Minimum	Maximum
Superficie	106	10442.600	42034.100	105.000	432000.000
altitude max	102	1158.390	1113.630	55.000	4807.000
Population 1999	105	624222.000	693336.000	6600.000	2554449.000
nb d'essais OGM en 2000	106	1.594	3.123	0.000	16.000
nb d'exploitations agricoles	96	6914.660	3732.840	0.000	18242.000
exploitation avec superficie agricole	96	6803.020	3684.450	0.000	18183.000
superficie agricole utilisée	96	289534.000	146942.000	0.000	559062.000
exploitation en blé tendre	96	2865.310	2217.240	0.000	11312.000
superficie en blé tendre	96	51005.300	52412.900	0.000	195012.000
exploitation en bovin	96	2937.590	2534.310	0.000	12444.000
effectifs bovins	96	211030.000	196406.000	0.000	764522.000

Tableau 3.1 : Variables continues étudiées par l'ACP.

Notons à partir du Tableau 3.1 que nous avons des valeurs manquantes pour certains individus.

L'étude des valeurs propres (Figure 3.1) montre que les quatre premiers axes permettent d'obtenir un pourcentage cumulé de 76.92%, ce qui reste peu. Cependant les valeurs propres associées à ces axes sont supérieures ou proches de 1. Nous allons donc restreindre l'étude aux quatre premiers axes factoriels.

Une première projection sur les 2 premiers axes factoriels montre (Figure 3.2) que les variables *exploitation avec superficie agricole* et *nb d'exploitations agricoles*, *effectifs bovins* et *exploitation en bovin*, ainsi que les 2 variables *superficie en blé tendre* et *superficie agricole utilisée* sont très corrélées avec respectivement des coefficients de corrélation de 1, 0.92 et 0.70. Ces résultats se retrouvent également dans le tableau 3.2 donnant la matrice de corrélation. On peut également remarquer une corrélation assez élevée (0.68) de la variable *exploitation en bovin* avec *nb d'exploitations agricoles* et *exploitations avec superficie*, ainsi que la corrélation des variables *exploitations en blé tendre* et *superficie agricole utilisée* (0.63). Cette dernière s'explique par le fait que la culture de blé est très consommatrice de superficie agricole, et qu'elle est liée à la variable *superficie en blé tendre*. Il est étonnant de ne pas constater de corrélation entre ces 2 variables. Notons de plus que la superficie du département n'est pas corrélée avec la superficie agricole utilisée, ce qui *a priori* n'est pas sûr. Il est intéressant de remarquer qu'il n'y a pas de variables inversement corrélées ce qui traduit par exemple que les productions de blé et de bovins ne sont pas antagonistes.

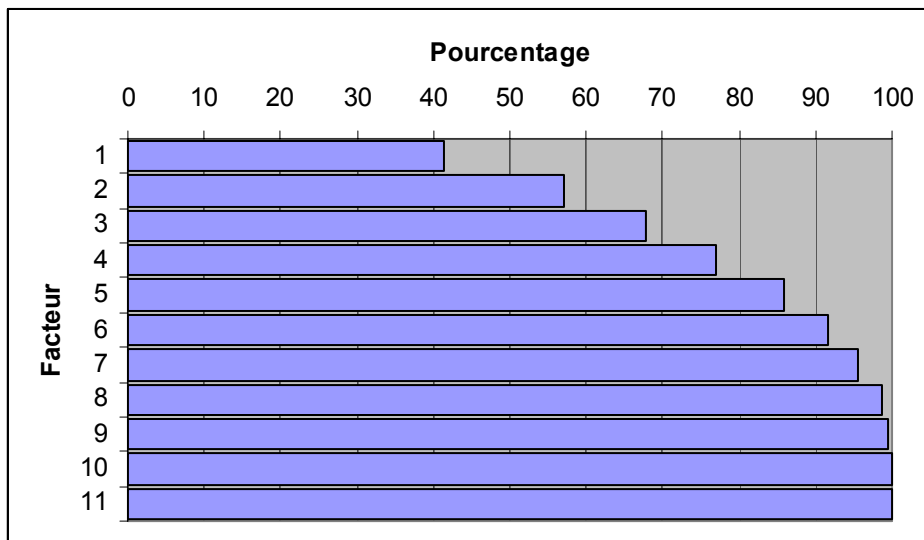


Figure 3.1 : Histogramme des valeurs propres (% cumulé) pour l'ACP.

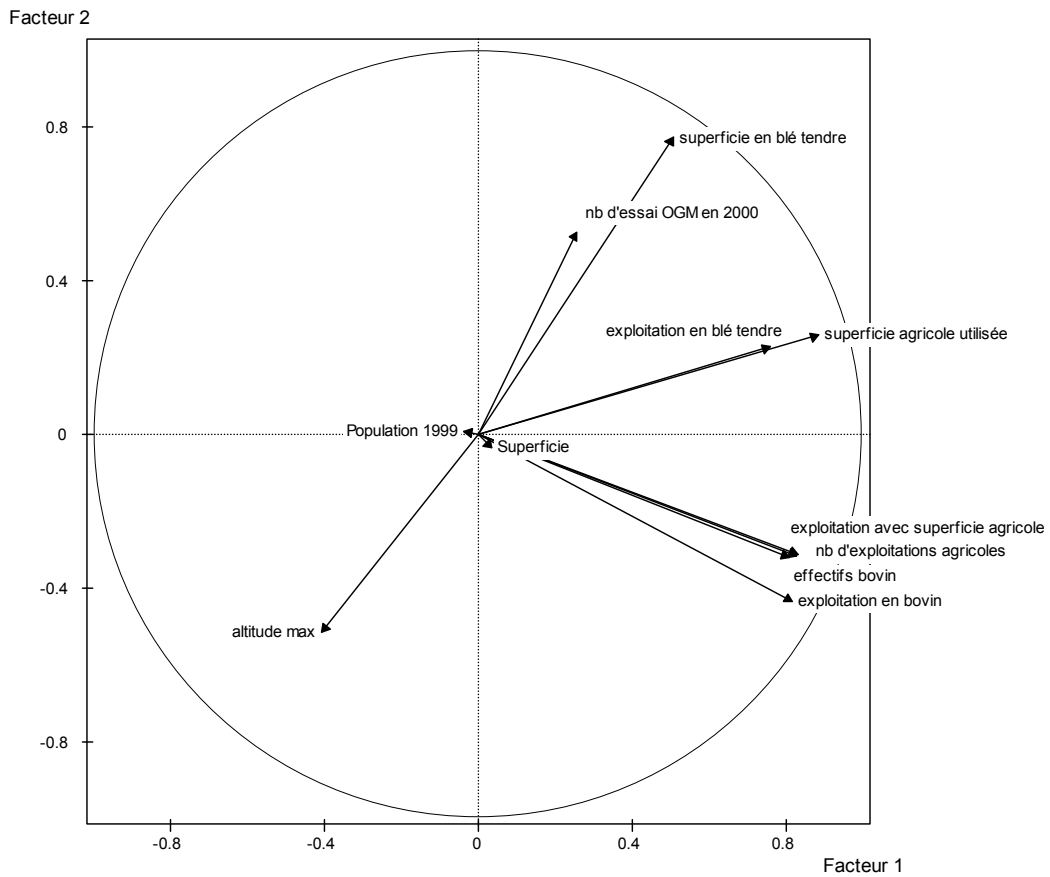


Figure 3.2 : Projection des variables sur les 2 premiers axes factoriels.

	Superficie	altitude max	Population 1999	nb d'essais OGM en 2000	nb d'exploitations agricoles	exploitation avec superficie	superficie agricole utilisée	exploitation en blé tendre	superficie en blé tendre	exploitation en bovin	effectifs bovins
Superficie	1.00										
altitude max	-0.01	1.00									
Population 1999	-0.02	0.02	1.00								
nb d'essais OGM en 2000	-0.04	-0.09	0.04	1.00							
nb d'exploitations agricoles	0.03	-0.14	0.02	0.21	1.00						
exploitation avec superficie agricole	0.03	-0.14	0.02	0.21	1.00	1.00					
superficie agricole utilisée	0.03	-0.42	-0.15	0.29	0.61	0.61	1.00				
exploitation en blé tendre	0.02	-0.29	0.08	0.27	0.50	0.49	0.63	1.00			
superficie en blé tendre	0.02	-0.52	-0.02	0.36	0.14	0.14	0.70	0.55	1.00		
exploitation en bovin	0.02	-0.15	-0.03	-0.05	0.68	0.68	0.60	0.52	0.09	1.00	
effectifs bovins	0.02	-0.27	-0.08	-0.10	0.57	0.57	0.67	0.52	0.18	0.92	1.00

Tableau 3.2 : Matrice de corrélation.

La figure 3.2 montre de plus que 7 variables contribuent aux axes 1 et 2, les variables : *exploitation avec superficie agricole*, *nb d'exploitations agricoles*, *effectifs bovins*, *exploitation en bovin*, *superficie agricole utilisée*, *exploitation en blé tendre* et *superficie en blé tendre* qui sont proches du cercle unité. Les 6 premières contribuent davantage au premier axe, alors que la dernière contribue au deuxième axe. Nous pouvons confirmer ces contributions par l'étude des corrélations des variables avec les axes (*cf.* Tableau 3.3). Ce tableau montre également qu'il n'y a pas de variables qui contribuent fortement à l'axe 3, tandis que pour l'axe 4, la superficie a une corrélation très élevée.

Libellé de la variable	Axe 1	Axe 2	Axe 3	Axe 4
Superficie	0.03	-0.03	-0.19	0.97
altitude max	-0.41	-0.51	0.37	0.03
Population 1999	-0.04	0.01	0.57	0.17
nb d'essais OGM en 2000	0.25	0.52	0.58	0.02
nb d'exploitations agricoles	0.83	-0.31	0.30	0.04
exploitation avec superficie agricole	0.83	-0.32	0.30	0.04
superficie agricole utilisée	0.88	0.26	-0.13	-0.02
exploitation en blé tendre	0.76	0.23	0.09	0.03
superficie en blé tendre	0.51	0.77	-0.11	0.02
exploitation en bovin	0.82	-0.44	-0.14	-0.05
effectifs bovins	0.81	-0.32	-0.31	-0.08

Tableau 3.3 : corrélation des variables avec les facteurs.

Une fois que les axes ont été caractérisés par les variables, nous représentons la projection des individus (Figure 3.3) dans le premier plan factoriel en fonction de leur contribution au plan. Cette figure montre que les départements de la région parisienne (75, 90, 92, 95, 91, 78)

et la Corse (2A, 2B) sont peu agricoles dans le sens de leur production en bovins et de leur nombre d'exploitations et leur superficie agricole employée, à la différence des départements de l'ouest de la France tels que 29, 22, 35, 85, 49, 50, etc... Nous remarquons de plus que les départements 45, 60, 28, 2, 51 et 62, qui se trouvent plutôt dans le nord de la France, présentent une superficie en blé tendre importante.

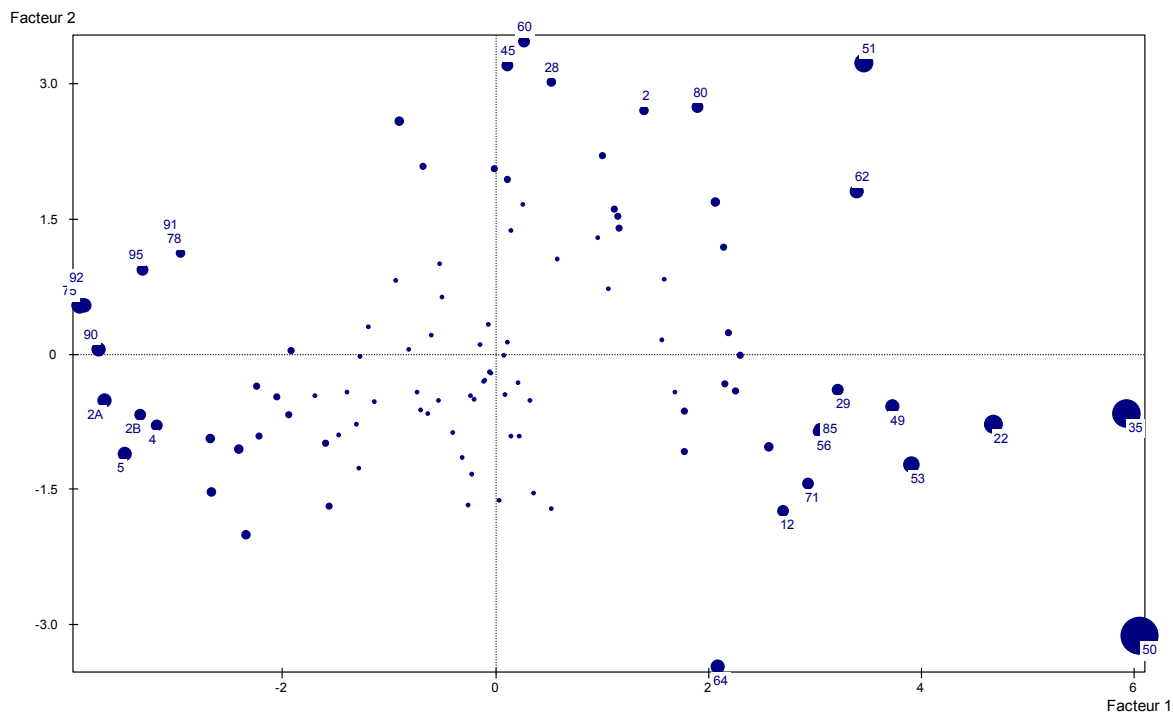


Figure 3.3 : Projection des individus dans le premier plan factoriel.

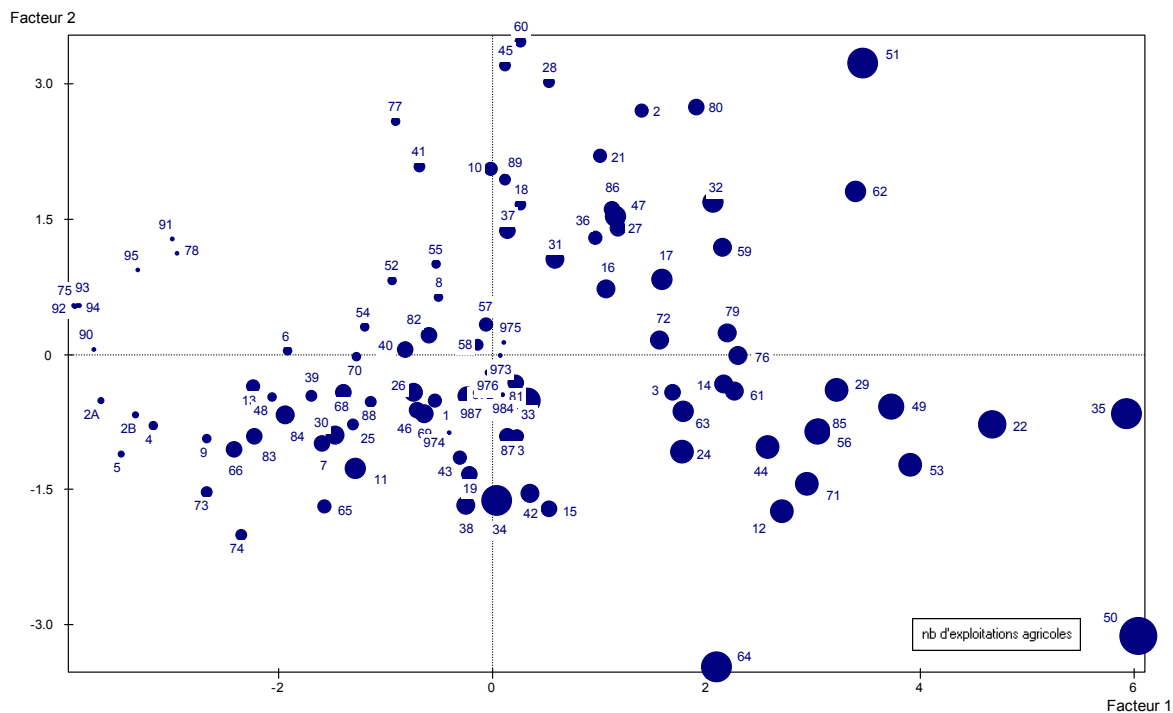


Figure 3.4 : Projection des individus dans le premier plan factoriel en fonction du nombre d'exploitations agricoles.

Nous pouvons constater à partir de la Figure 3.4 qui représente la même projection des individus en habillant les points en fonction du nombre d'exploitations, que l'axe 1 peut être assez bien résumé par la variable *nb d'exploitations agricoles*. Ainsi les individus sur la droite du graphique contiennent un grand nombre d'exploitations à la différence des départements sur la gauche du graphique, la taille du symbole variant dans ce sens.

L'étude du plan constitué de l'axe 1 et 3 n'apportant pas d'information supplémentaire par rapport au plan constitué de l'axe 2 et 3, nous nous limitons à l'étude de ce dernier.

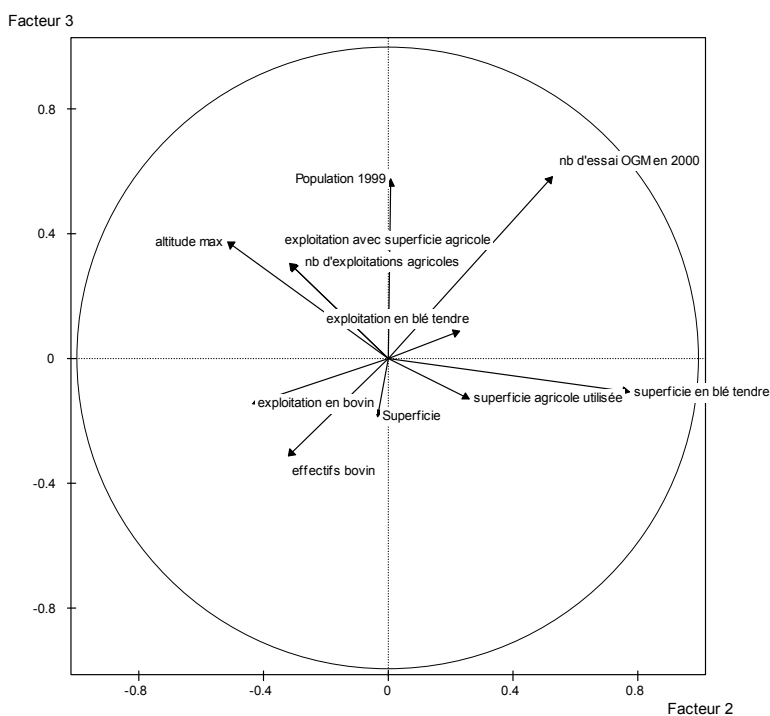


Figure 3.5 : Projection des variables sur le plan constitué de l'axe 2 et 3.

La figure 3.5 ne fait pas apparaître de variables qui contribuent fortement à l'axe 3, en effet seuls la *population* et le *nombre d'essais OGM* ont une corrélation (0.57 et 0.58 respectivement cf. Tableau 3.3) supérieure à 0,50. L'axe 2 reste caractérisé par la superficie en blé tendre.

Sur la figure 3.6 (projection des individus sur le plan formé par les axes 2 et 3 en fonction de leur contribution), l'Isère (38) apparaît avoir une population importante, et la Marne (51) et Haute Garonne (31) ont effectué un nombre important d'essais OGM. Nous retrouvons que les départements 45, 60, 28, 2, 51 et 62 présentent une superficie en blé tendre importante, à l'inverse de la Manche (50) et des Pyrénées-Atlantiques (64). Cette dernière information n'apparaissait pas aussi clairement dans le premier plan de projection.

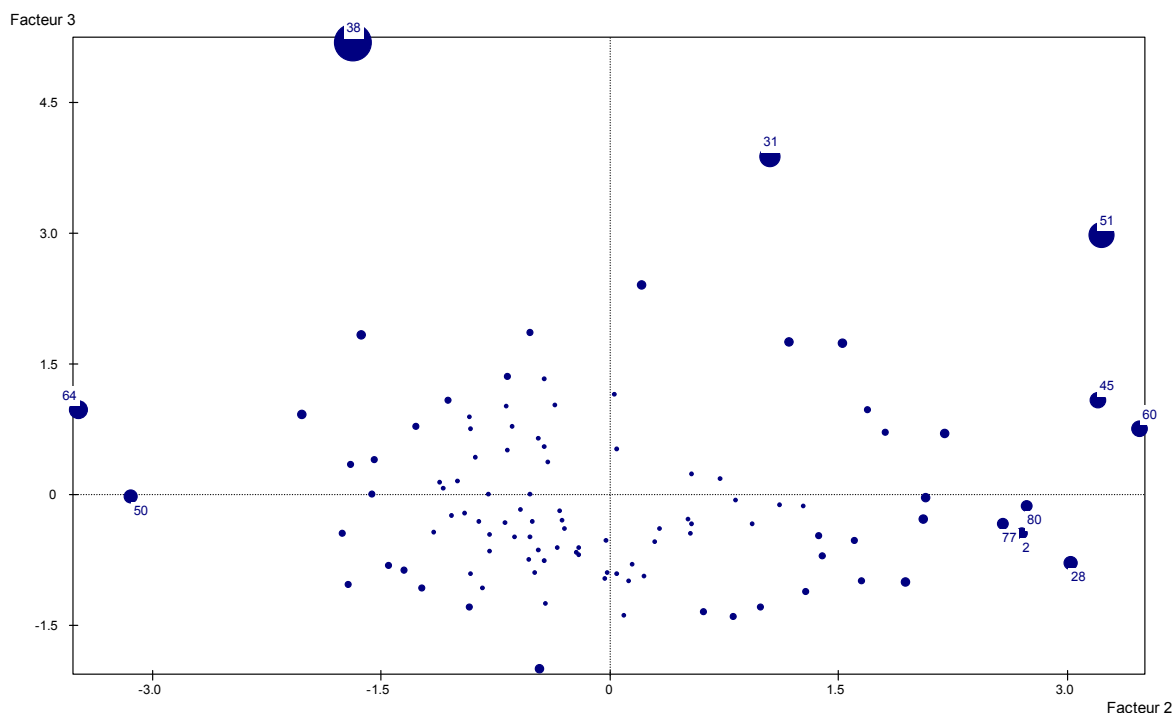


Figure 3.6 : Projection des individus dans le plan formé par les axes 2 et 3.

Les plans constitués de l'axe 4 sont intéressants pour l'interprétation de la variable superficie qui possède une forte corrélation avec cet axe (*cf.* Tableau 3.3 et Figure 3.7). Nous avons donc étudié le plan de projection constitué des facteurs 1 et 4. L'axe 4 est donc caractérisé par la variable superficie et l'axe 1 par les variables agricoles fortement corrélées entre elles (Figure 3.6). La projection des individus sur ce plan (Figure 3.8) montre que l'individu Territoires Australes et Antarctiques Françaises est caractérisé par sa superficie. Dans la suite des analyses factorielles nous ne considérerons plus cet individu. Dans une moindre mesure la Guyane (973) et l'Isère (38) sont également des départements de superficie importante. L'axe 1 fait ressortir dans la direction positive des départements agricoles de l'ouest tels que l'Ile et Vilaine, la Manche, les Côtes d'Armor et la Mayenne, tandis que la direction opposée regroupe des départements de la région parisienne à faible production agricole (Paris, Hauts de Seine, Seine St Denis et Val de Marne, ...).

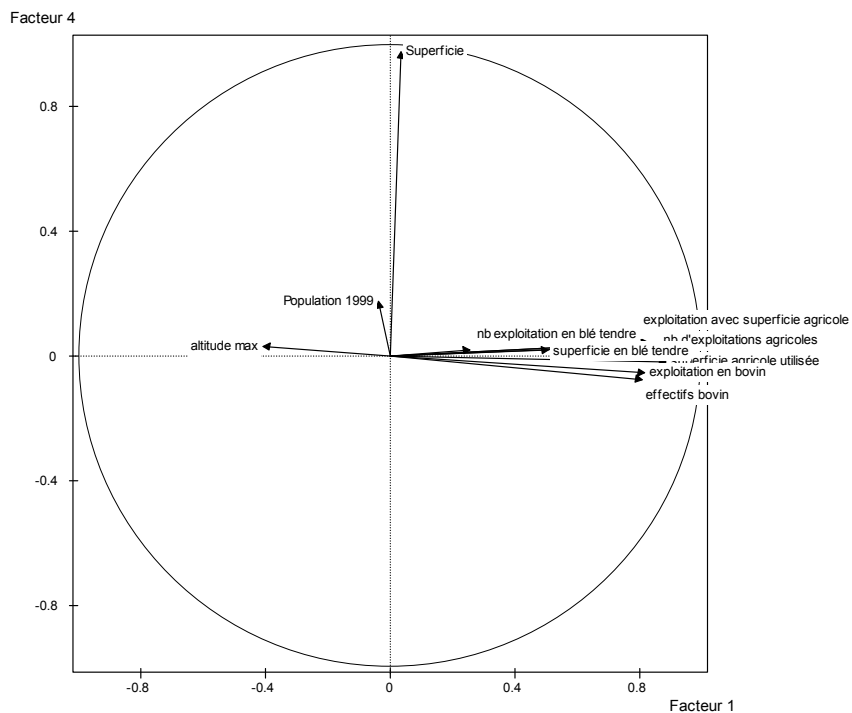


Figure 3.7 : projection des variables sur le plan formé par les axes 1 et 4.

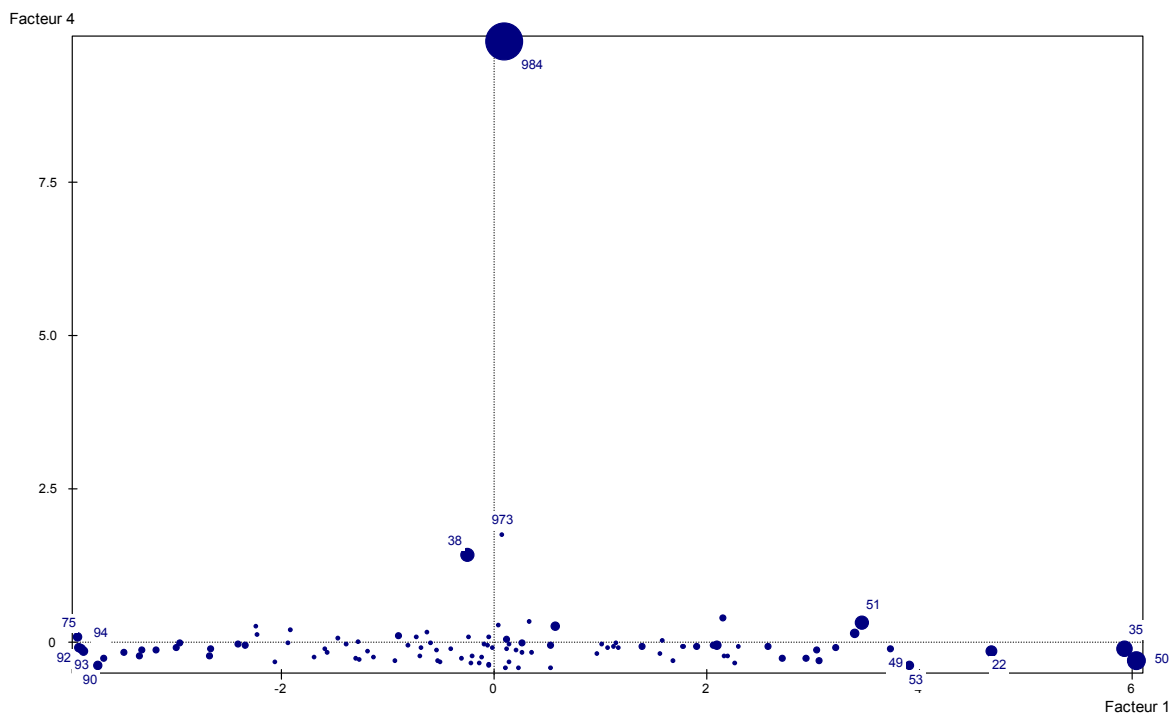


Figure 3.8 : Projection des individus sur le plan formé des axes 1 et 4.

Cette analyse a permis de faire apparaître des corrélations entre variables (*exploitation avec superficie* et *nb d'exploitations agricoles*, *effectifs bovins* et *exploitation en bovin*, *superficie en blé tendre* et *superficie agricole utilisée*) qui nous serviront dans la suite de l'étude. Cette analyse a de plus fait ressortir principalement 3 groupes d'individus : la région

parisienne à faible production agricole, des départements de l'ouest de la France à forte production agricole et des départements du nord à forte production en blé tendre. Les DOM TOM apparaissent comme très particuliers de part leurs caractéristiques géographiques ; nous n'étudierons plus ces individus dans les autres analyses factorielles (paragraphe 5 et 6). Nous pouvons noter également la particularité de l'Isère qui se dégage des autres départements. Nous allons affiner l'étude de ces regroupements par une classification hiérarchique.

4. Classification hiérarchique

Nous avons réalisé dans un premier temps la classification hiérarchique après l'analyse en composantes principales précédente. Ainsi tous les individus sont considérés. L'histogramme des indices pour les 20 premiers nœuds (Figure 4.1) et le dendrogramme (Figure 4.2) montre que nous pouvons envisager la coupure à différents niveaux. Une première coupure possible serait avec 3 classes, cependant la perte d'information à ce niveau est trop importante. En effet nous obtenons un quotient d'inertie interclasse sur l'inertie totale (intra et interclasse) de 37.4%. Deux autres coupures sont envisageables, une coupure en 5 ou 6 classes. En effet nous observons une baisse d'indice assez importante pour une coupure en 5 classes et une stabilisation de l'indice pour une coupure à 6 classes (Figure 4.1). De plus le quotient d'inertie interclasse sur l'inertie totale dépasse 50% (nous obtenons respectivement pour 5 et 6 classes, 54.08% et 60.07%). On peut noter d'après le dendrogramme (Figure 4.2) que la sixième classe est constituée d'un seul département (l'Isère 38), ce qui ne va pas compliquer l'analyse. Nous allons donc choisir une coupure en 6 classes pour la suite de ce paragraphe.

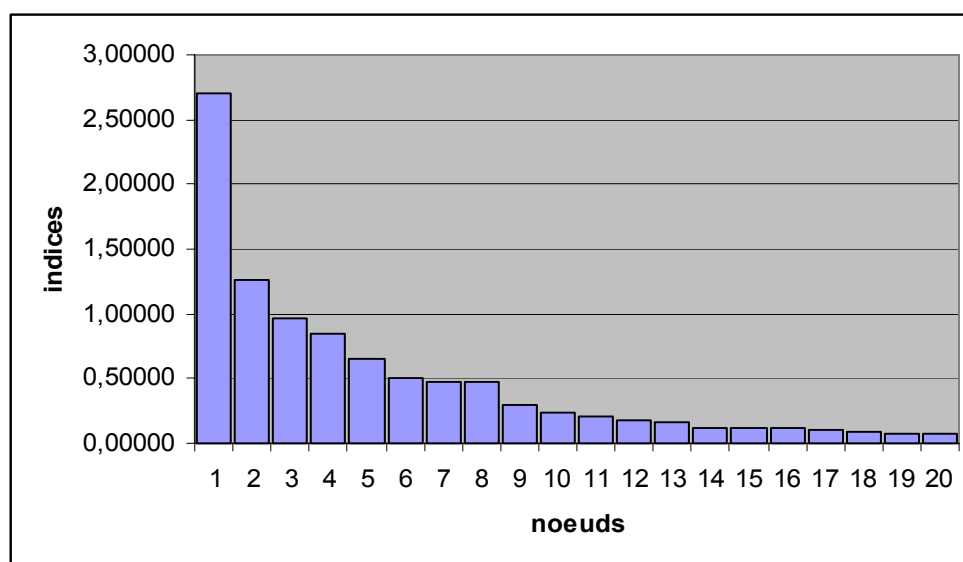


Figure 4.1 : Histogramme des indices des 20 premiers nœuds.

Le Tableau 4.1 présente d'une part les départements dans chaque classe répertoriée par ordre croissant de la distance au centre de la classe, et d'autre part les valeurs tests (supérieures à 2 en valeur absolue) des variables actives uniquement. Rappelons que les valeurs tests représentent une statistique issue d'un test, elles mesurent l'écart entre les moyennes de l'ensemble des individus et des classes. Cet écart est significatif au seuil de 5% pour les valeurs tests supérieures à 2 en valeur absolue.

Notons tout d'abord que 2 individus se retrouvent seuls dans leur classe : l'Isère et les Terres Australes et Antarctiques Françaises. La première classe est constituée de 41 départements en partie des DOM TOM et de départements plutôt au centre de la France ayant une faible culture en blé tendre (en exploitation et superficie) ainsi qu'une faible population. La troisième classe contient 20 départements essentiellement de la région parisienne avec quelques départements du sud de la France dont la Corse. Ces départements sont regroupés par un comportement similaire au niveau de la faible production agricole (superficie, nombre d'exploitations) aussi bien du blé que des bovins. La cinquième classe de 21 départements est marquée par la production de blé (caractérisée par les variables *superficie en blé tendre*,

superficie agricole utilisée, nb d'essais OGM et faiblement par exploitation en blé tendre) ainsi que par une faible altitude. Les départements de cette classe se trouvent principalement dans le nord de la France. La sixième classe représente 22 départements à forte production bovine (*effectif bovin, exploitation en bovin*) et agricole en général (*nb d'exploitations agricoles, exploitation avec superficie, superficie agricole utilisée et exploitation en blé tendre*). Notons que ces départements sont également de faible altitude. Cette classe contient essentiellement des départements de l'ouest de la France.

Afin d'étudier en profondeur ces 4 classes, nous avons regardé les valeurs tests à la suite d'une coupure en 9 classes (qui représente un saut d'indice important sur la Figure 4.1, et un quotient d'inertie interclasse sur l'inertie totale de 72.93%). Ainsi les classes 1, 3 et 5 sont divisées en 2 classes, et la classe 6 est inchangée. Sur le Tableau 4.2, nous n'avons donc représenté que ces 3 classes qui sont maintenant 6 classes pour la coupure en 9 classes (classe 1a, 1b, 2, 3a, 3b, 5a et 5b).

Nous remarquons que les caractéristiques des classes 1a et 1b (qui forment la classe 1 pour une coupure en 6 classes) sont différentes. En effet la classe 1 qui contient le plus grand nombre de départements (29) est la classe la plus disparate ; elle est caractérisée principalement par peu d'essais OGM et une faible population. Elle contient les départements des DOM TOM ainsi que quelques départements du centre de la France. La classe 2 est quant à elle un peu plus homogène, elle est constituée de 13 départements de faible production agricole qui se trouvent majoritairement en périphérie de l'hexagone. Les classes 3a et 3b (qui constituent la classe 3 pour une coupure en 6 classes) se ressemblent davantage. Leur distinction provient de 10 départements caractérisés par leur altitude (Corse et Alpes) et de 8 départements caractérisés par leur population (la région parisienne). Ces départements se ressemblent par leur faible production agricole. Les classes 5a et 5b (qui sont la classe 5 pour une coupure en 6 classes) se distinguent par leur nombre d'essais OGM. La classe 5a est formée par 14 départements à forte production en blé, alors que la classe 5b est formée de 8 départements qui ont réalisé un grand nombre d'essais OGM.

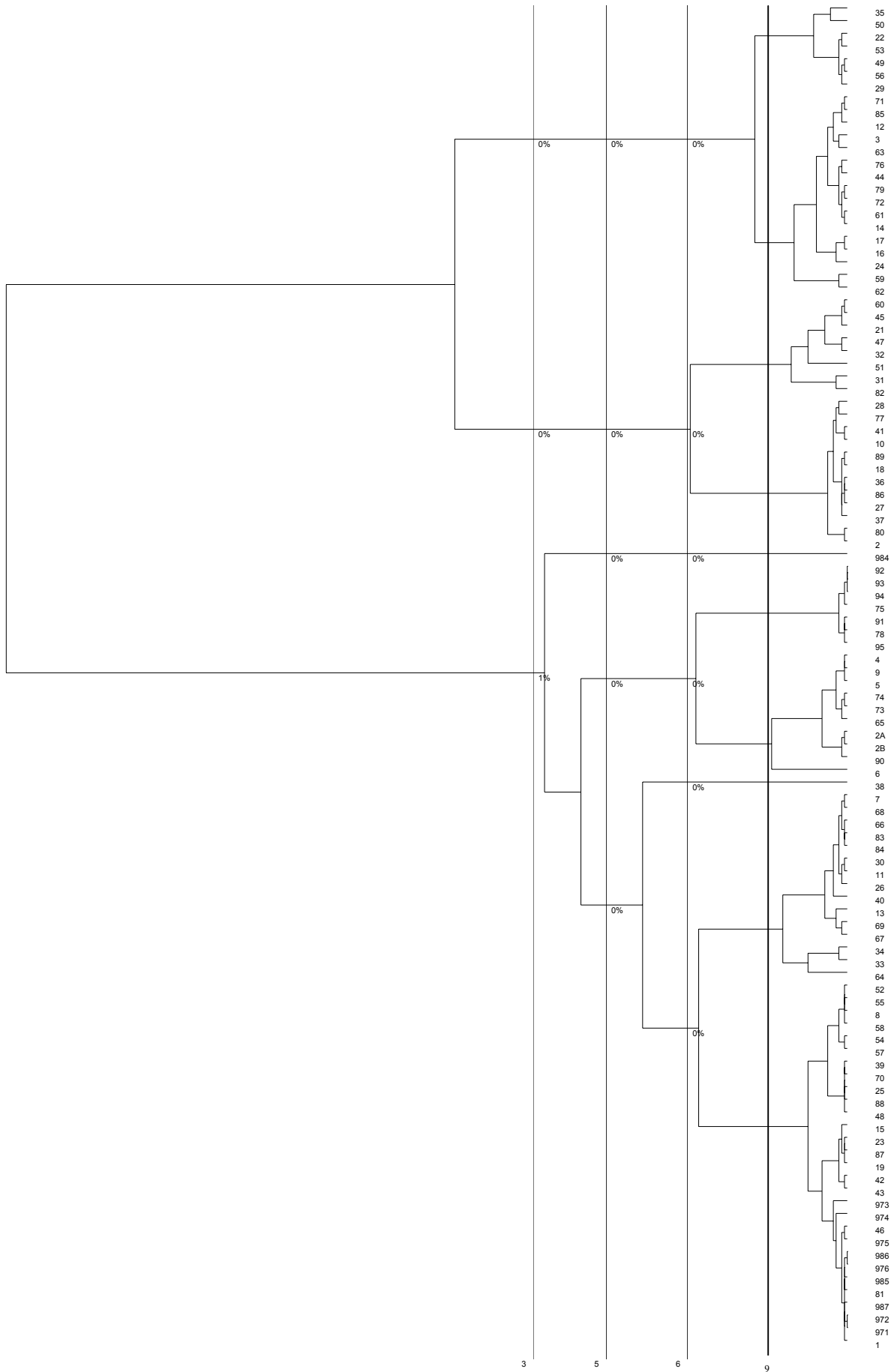


Figure 4.2 : Dendrogramme.

	Superficie	altitude max	Population 1999	nb d'essais OGM en 2000	nb d'exploitations agricoles	exploitation avec superficie	superficie agricole utilisée	exploitation en blé tendre	superficie en blé tendre	exploitation en bovin	effectifs bovins
Classe 1 Effectif 41 1, 972, 971, 976, 985, 46, 986, 987, 88, 81, 43, 68, 25, 39, 70, 975, 87, 7, 19, 67, 42, 57, 54, 974, 58, 23, 26, 52, 55, 30, 69, 65, 11, 84, 40, 973, 15, 33, 82, 34			-2,35					-2,97	-3,68		
Classe 2 38											
Classe 3 Effectif 21 2A, 2B, 90, 9, 48, 4, 95, 91, 78, 83, 94, 73, 66, 93, 92, 13, 5, 75, 74, 6		3,16			-5,83	-5,84	-7,02	-4,46	-4,26	-4,64	-4,61
Classe 4 984											
Classe 5 Effectif 21 10, 86, 2, 21, 18, 80, 16, 41, 37, 36, 27, 89, 45, 32, 28, 17, 60, 77, 47, 51, 31			-3,20	5,32			4,32	2,70	7,07		
Classe 6 Effectif 22 56, 29, 49, 14, 44, 61, 85, 76, 71, 53, 79, 12, 72, 24, 63, 3, 22, 62, 35, 59, 64, 50			-2,56		5,74	5,70	5,14	4,83		7,73	7,80

Tableau 4.1 : Valeurs tests en fonction des classes et des variables actives pour une coupure de 6 classes.

	Superficie	altitude max	Population 1999	nb d'essais OGM en 2000	nb d'exploitations agricoles	exploitation avec superficie	superficie agricole utilisée	exploitation en blé tendre	superficie en blé tendre	exploitation en bovin	effectifs bovins
Classe 1 Effectif 29 972, 971, 976, 985, 1, 986, 88, 987, 70, 43, 87, 25, 46, 58, 81, 975, 8, 39, 23, 57, 19, 54, 52, 55, 42, 48, 974, 15, 973			-2,83	-3							
Classe 2 Effectif 13 30, 68, 84, 83, 7, 11, 69, 26, 67, 40, 13, 33, 34							-2,85	-2,94	3,27	-2,86	-3,30
Classe 4 Effectif 10 9, 73, 4, 5, 2B, 65, 2A, 74, 66, 6		6,59			-2,83	-2,84	-4,22		-3,11		-2,68
Classe 5 Effectif 8 95, 94, 93, 92, 78, 91, 75, 90			2,59		-5,06	-5,05	-5,14	-3,41		-3,32	-3,09
Classe 7 Effectif 14 86, 18, 10, 89, 37, 36, 27, 2, 28, 80, 16, 41, 17, 77		-3,06					3,40		6,18		
Classe 8 Effectif 8 21, 47, 45, 32, 60, 31, 82, 51				8,82					2,55		

Tableau 4.2 : Valeurs tests en fonction des classes et des variables actives pour une coupure de 9 classes.

A la suite de cette classification, nous reprenons la projection des individus dans le premier plan factoriel issu de l'analyse en composantes principales avec la représentation des 6 classes (Figure 4.3). A l'exception des deux classes ne contenant qu'un seul individu, nous retrouvons distinctement les quatre autres classes. La première classe contenant le plus de départements (41) est mal représentée (*cf.* Figure 3.3), ces départements ne se distinguent des autres que légèrement par une faible production agricole restant globalement dans la moyenne (proche de l'origine). Nous retrouvons bien les caractéristiques des trois autres classes avec la classe 3 ayant une très faible production agricole contenant les départements de la région parisienne et du sud de la France, la classe 5 marquée par sa production en blé tendre avec les départements plutôt dans le nord de la France et la classe 6 caractérisée par sa production agricole importante, particulièrement en bovins, qui contient les départements les plus à l'ouest de la France.

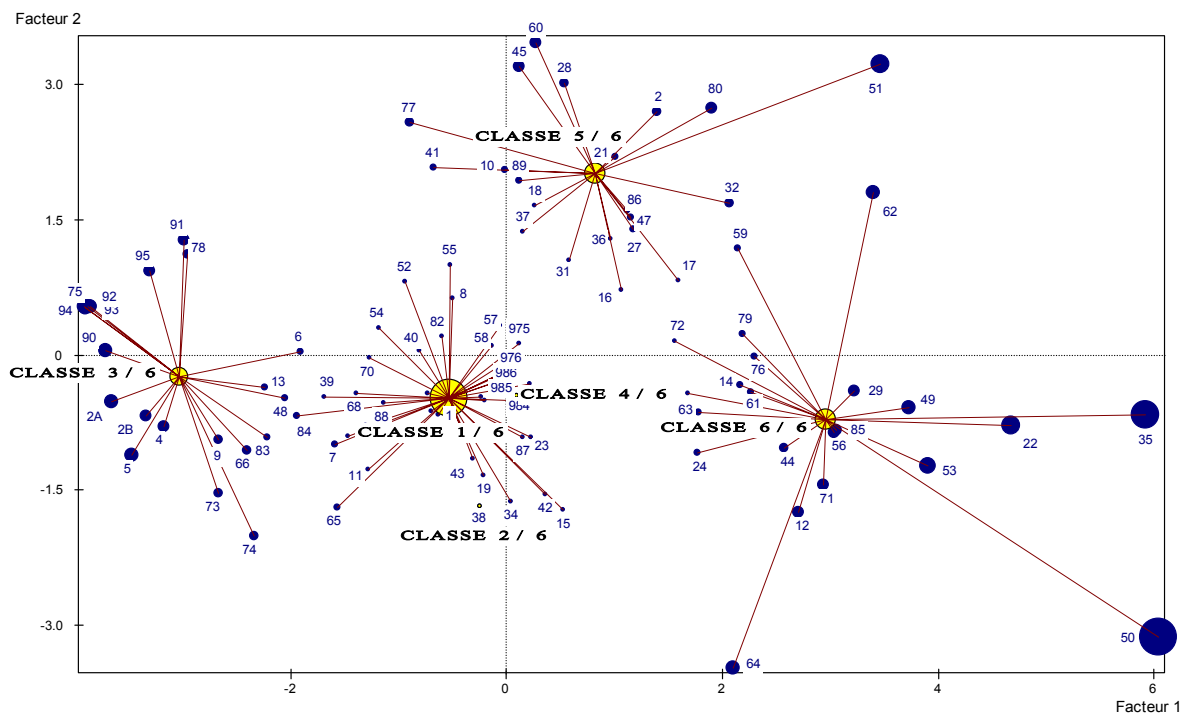


Figure 4.3 : Projection des individus dans le premier plan factoriel obtenu par une ACP.

Cette classification hiérarchique à la suite de l'ACP, nous a permis de caractériser 6 classes de départements par des comportements différents quant à leur production agricole. De plus la classification hiérarchique en 9 classes fait apparaître les DOM TOM dans une même classe ce qui confirme un comportement similaire. Ces départements ne seront plus conservés dans la suite de l'étude.

5. Analyse factorielle des correspondances

Nous allons étudier dans cette partie les caractéristiques agricoles des départements selon les caractéristiques de la population. Les variables considérées sont celles du tableau 5.1. Seuls les départements de métropole (ceux de Corse inclus) serviront lors de l'analyse, il y a donc 96 individus.

Superficie
0 à 14 ans
15 à 24 ans
25 à 39 ans
40 à 59 ans
60 ans ou plus
nb d'essais OGM en 2000
nb d'exploitations agricoles
exploitation avec superficie agricole
superficie agricole utilisée
exploitation en blé tendre
superficie en blé tendre
exploitation en bovin
effectifs bovins

Tableau 5.1 : Variables actives.

Nous avons de plus retenu quatre variables pour illustrer les analyses : *altitude max*, *actifs*, *étranger*, *diplômes supérieurs* et *PIB* pour les variables quantitatives et *montagneux* pour les variables nominales. L'histogramme des valeurs propres (Figure 5.1) montre que les trois premiers facteurs représentent un pourcentage important (93.57%) de l'information. Nous allons donc restreindre l'analyse à l'étude des plans de projection formés par les trois premiers facteurs.

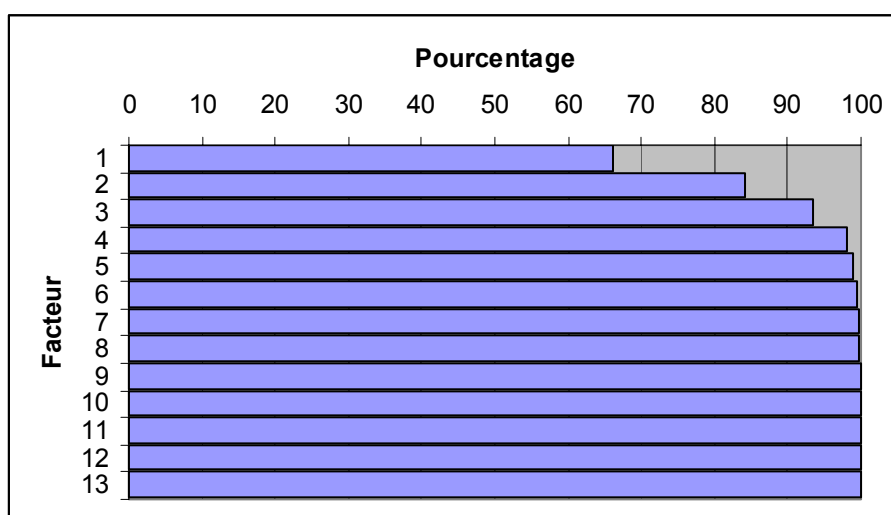


Figure 5.1 : Histogramme des valeurs propres (% cumulé) pour l'AFC.

La projection dans le premier plan factoriel (formé par les 2 premiers axes factoriels, cf. Figure 5.2) fait apparaître plusieurs regroupements d'individus autour de quelques variables. Notons tout d'abord que la variable *effectifs bovins* contribue fortement à ce plan de

projection (cf. en annexe Tableau A.3). La variable *exploitation en bovin* qui représente le nombre d'exploitations ayant des bovins y contribue moins ; elle reste cependant proche de la variable *effectifs bovins*, ce qui confirme que celles-ci sont corrélées. Ces deux variables font apparaître une classe de départements à forte production de bovins uniquement : l'Allier (3), le Cantal (15), la Creuse (23), l'Aveyron (12), la Corrèze (19), la Nièvre (58) et la Saône et Loire (71) qui se trouvent dans le centre de la France et la Manche (50), la Mayenne (53), la Vendée (85) et l'Orne (61) qui se trouvent plus au nord. Ces individus se trouvent éloignés des variables décrivant la population. Les départements français à forte production bovine sont donc des départements plutôt faiblement peuplés.

Les variables décrivant la population en fonction de l'âge sont proches les uns des autres. Ainsi les départements à forte population ont un nombre important de personnes dans chaque classe d'âge. Il est intéressant de noter que la variable illustrative *actifs* résume bien l'ensemble de ces variables dans cette projection car elle se retrouve projetée au centre des autres variables. Nous pouvons noter que certains départements (1, 30, 31, 33, 34, 38, 59, 67, 69, 84) contribuant légèrement au plan sont projetés proches de ces variables. Ce sont des départements hors de la région parisienne qui pour la plupart ont des villes importantes (telles que Bordeaux, Lille, Strasbourg, Lyon, Grenoble) et qui ont donc une population assez importante.

Des départements de la région parisienne (75, 78, 91, 92, 94, 95) forment un regroupement avec trois départements de la Côte d'Azur (les Alpes Maritimes (06), Les Bouches du Rhône (13) et le Var (83)). Ce regroupement est proche des variables décrivant la population en classe d'âge, ce qui montre qu'ils ont une population importante. Il est cependant bien centré sur les variables illustratives : *PIB*, *Diplômes supérieurs* et *étrangers* ce qui montre que dans une forte population il y a un grand nombre de diplômés et d'étrangers et ceci est lié à un PIB important. Attention, ça ne signifie pas qu'il y a des différences avec les autres départements au niveau des proportions, mais du nombre uniquement. Notons de plus que le département Rhône (69) se retrouve entre les deux regroupements détaillés précédemment.

Les départements consommateurs de surface agricole sont liés aux deux variables *superficie en blé tendre* et *surface agricole utilisée* qui contribuent au plan. Ces deux variables sont éloignées l'une de l'autre (alors qu'elles sont corrélées (0.70), cf. Tableau 3.2) ce qui donne un regroupement de départements (28, 10, 51, 89, 86, 2, 18, 32, 36, 55, 79) très large autour de ces variables. Notons de plus que les 2 variables *exploitation avec superficie* et *nb d'exploitations agricoles* très corrélées sont projetées sur le même point.

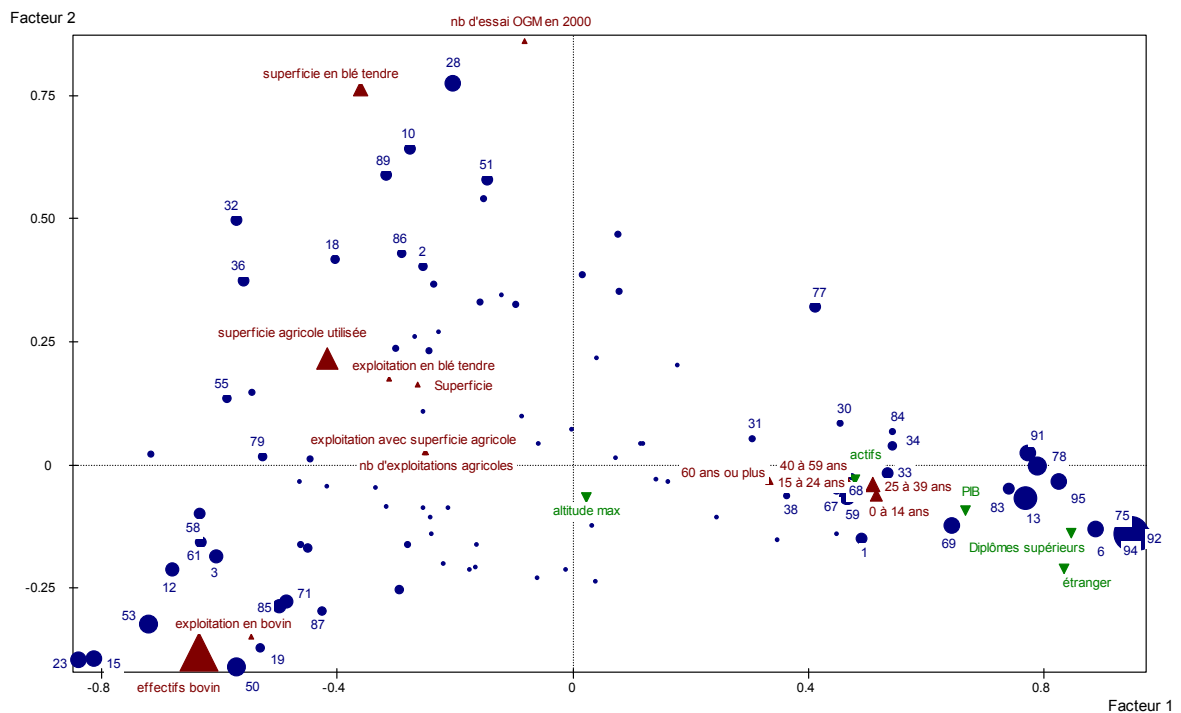


Figure 5.2 : Projection des individus, fréquences actives et illustratives dans le plan de projection formé par les 2 premiers axes factoriels.

Si nous reprenons la Figure 5.2 (projection des individus dans le premier plan factoriel) en ajoutant la variable illustrative *montagneux* (Figure 5.3), nous constatons que les départements à forte production agricole ne sont pas ou peu montagneux, alors que les départements très montagneux se retrouvent mal représentés dans ce plan (ils n'y contribuent que très faiblement). Il peut être intéressant d'effectuer une analyse des correspondances multiples en fonction de cette variable.

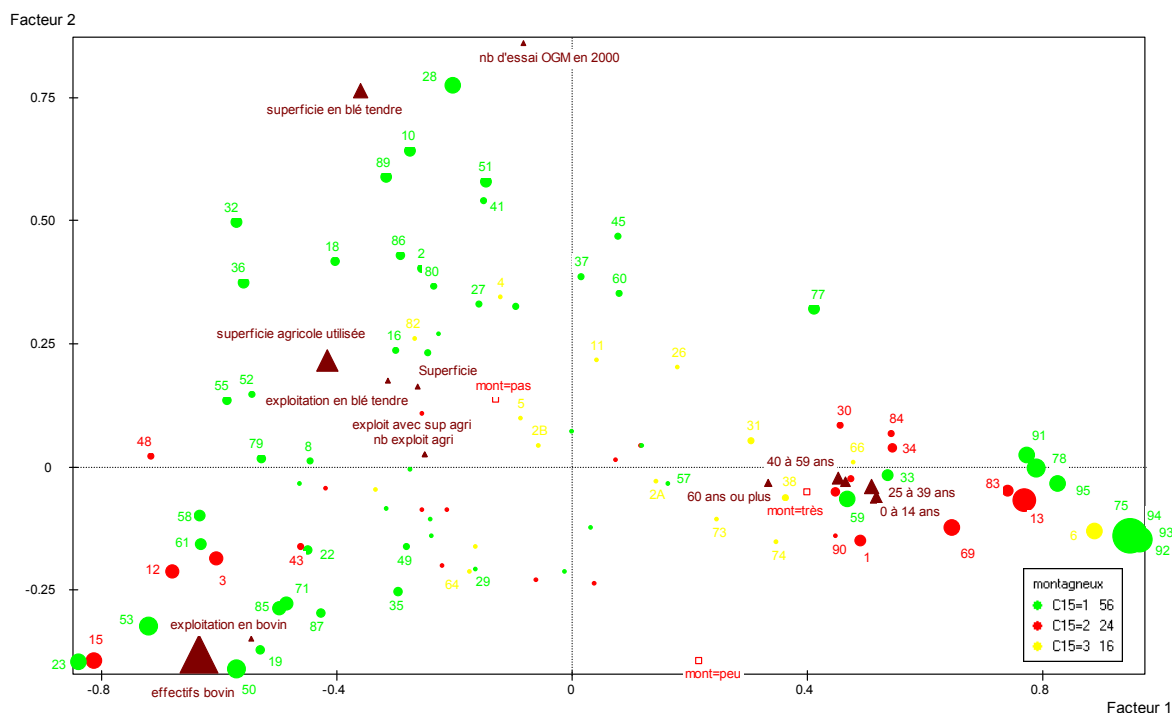


Figure 5.3 : Projection des individus, fréquences actives et illustratives dans le plan de projection formé par les 2 premiers axes factoriels en différenciant la couleur des individus en fonction de la variable montagneux.

L'ajout du troisième axe factoriel pour les deux plans de projection (axe 1 et 3 et axe 2 et 3) apporte une information assez semblable. Nous représentons donc uniquement les projections des individus et variables dans le plan de projection formé par les axes 2 et 3 (Figure 5.4).

L'axe 3 fait apparaître le département de l'Ain qui contribue fortement à cet axe. Il se retrouve éloigné des autres départements et des projections des variables. Les variables *l'âge de 0 à 14 ans*, *effectifs bovins*, *superficie en blé tendre* et *superficie agricole utilisée* sont celles qui contribuent le plus fortement à ce plan de projection. Nous pouvons donc penser que l'Ain se distingue particulièrement des autres départements dans son comportement face à l'ensemble de ces variables. Il est intéressant de noter que dans ce plan de projection les deux variables *effectifs bovins* et *superficie en blé tendre* sont à l'opposées. Les départements proches à forte production bovine (qui sont les mêmes que précédemment) sont donc plutôt à faible production en blé tendre, et inversement. Les départements ayant à la fois une production en bovins et en blé tendre importante ou non se retrouvent entre ces deux variables et rien ne permet de les distinguer dans ce plan de projection.

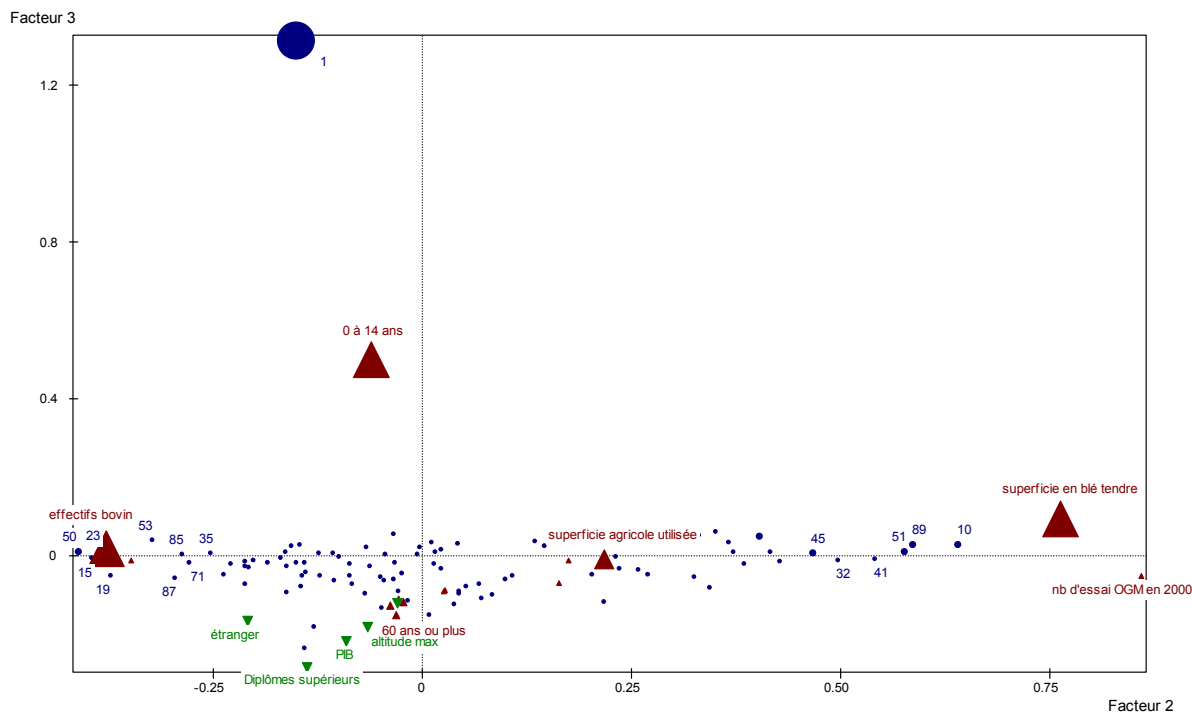


Figure 5.4 : Projection des individus, des variables actives et illustratives sur le plan de projection constitué des axes factoriels 2 et 3.

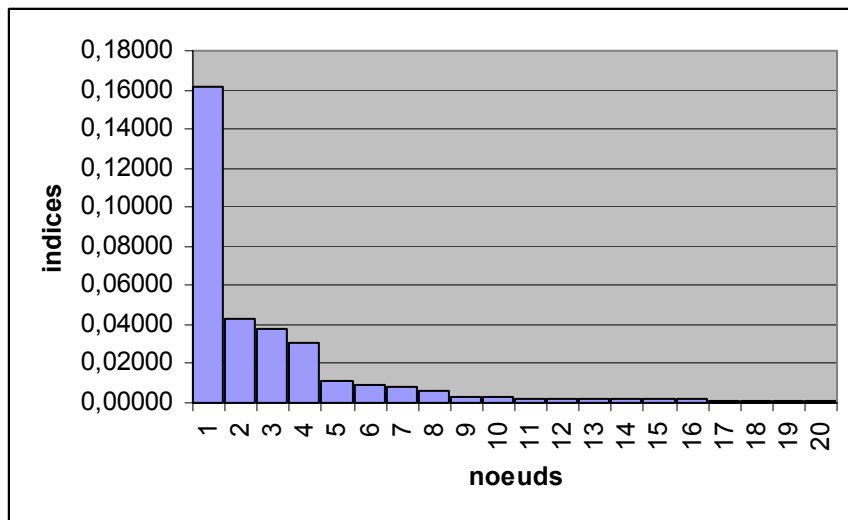


Figure 5.5 : Histogramme des indices des 20 premiers nœuds.

A la suite de cette analyse factorielle des correspondances, nous avons réalisé une classification hiérarchique. L'étude des indices des nœuds (Figure 5.5) et du dendrogramme (en Annexe Figure A.1) montre un saut d'indice important pour une coupure en 5 classes. Cette coupure permet d'obtenir un quotient d'inertie interclasse sur l'inertie totale de 80.12% ce qui montre qu'une part importante de l'information est conservée. Les 5 classes sont décrites en Annexe (Tableau A.4). Ainsi la première classe contient uniquement le département de l'Ain. La deuxième classe est formée de 26 départements assez disparates qui ont une faible production agricole et une population plutôt élevée. La troisième classe est caractérisée par une très faible production agricole, elle contient 11 départements de la région

parisienne et à forte population. La classe 4 contient 23 départements (plutôt du centre de la France) qui ont une production en blé tendre importante, avec une production en bovins plutôt faible. La classe 5 est formée de 35 départements de l'ouest de la France qui ont une production en bovins importante et plus généralement une forte production agricole.

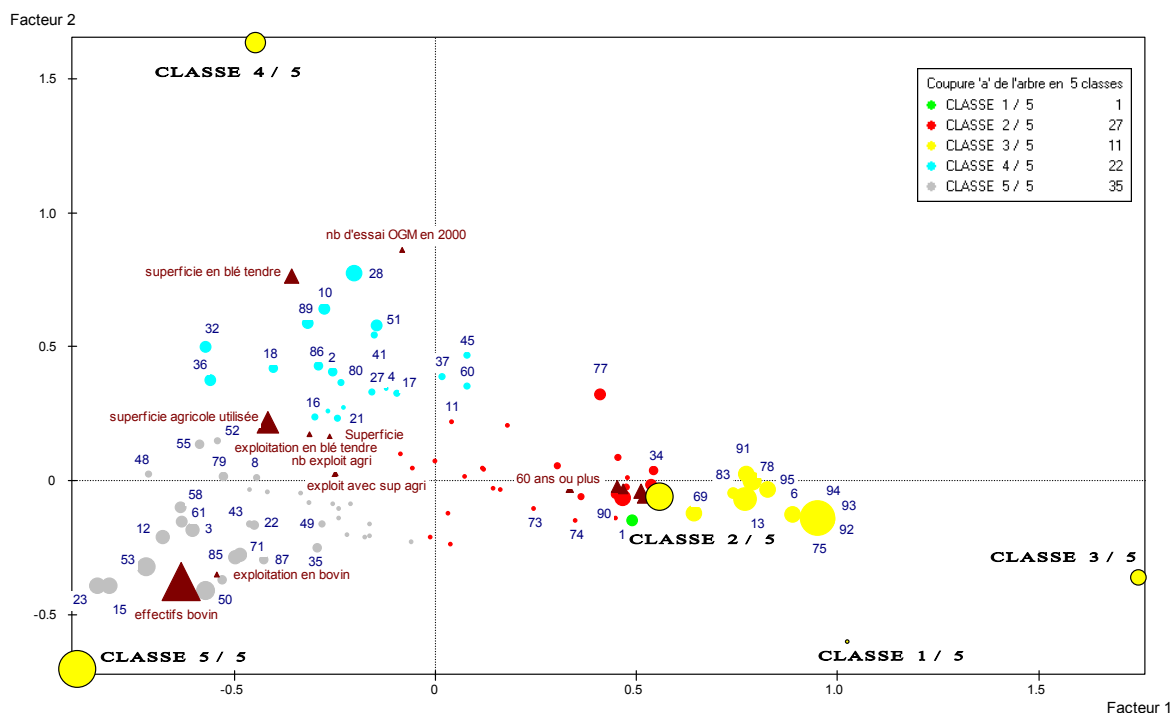


Figure 5.6 : Projection des individus, fréquences actives, illustratives et des centres des 4 classes dans le plan de projection formé par les 2 premiers axes factoriels.

La projection des individus et fréquences (actives et illustratives) dans le premier plan factoriel avec les centres des quatre classes obtenus par la classification hiérarchique (Figure 5.6), fait apparaître 4 classes bien distinctes. Cependant la distance des individus au centre des classes peut être éloignée (par exemple pour la classe 3). Des individus de la classe 3 peuvent être plus proche du centre de la classe 1. Nous remarquons que la classe 2, est assez mal représentée dans ce plan (taille des symboles petits selon la contribution) ; les individus de cette classe sont ceux proches des variables décrivant la population, qui sont des départements contenant des grandes villes. La classe 5 essentiellement constituée de départements de l'ouest de la France à forte production en bovins contient également des départements à forte production agricole. La troisième classe est assez éloignée de son centre et contient des départements qui contribuent faiblement au plan. Ces départements sont caractérisés par leur production en blé. De même les départements (région parisienne) de la classe 3 sont éloignés de leur centre de la classe.

Cette étude a fait apparaître des regroupements de départements se distinguant soit par une production importante en bovin en particulier (dans l'ouest de la France) ou consommateurs de surface agricole uniquement (plutôt dans le nord de la France), ou encore des regroupements à faible production agricole, mais fortement peuplés (avec 2 classes : des départements de la région parisienne, et des départements avec de grandes villes).

6. Analyse des correspondances multiples

A l'aide de cette analyse, nous cherchons à étudier la production agricole de manière plus grossière que précédemment en fonction du relief des départements.

Afin de réaliser une analyse des correspondances multiples sur les données agricoles, nous avons créé des variables qualitatives obtenues à partir des variables continues caractérisant les données agricoles. Pour chaque variable (*nb d'exploitations agricoles, exploitation avec superficie agricole, superficie agricole utilisée, exploitation en blé tendre, superficie en blé tendre, exploitation en bovin et effectifs bovins*), nous avons créé 2 classes (« peu » et « beaucoup ») en fonction de la médiane : si la valeur est en dessous de la médiane la classe est « peu », si elle est au dessus, la classe est « beaucoup ». Ces variables ainsi que la variable *montagneux* qui s'est révélée intéressante par l'analyse précédente, seront les variables nominales actives. Nous avons de plus conservé les variables qualitatives existantes (*augmentation de la population, frontière, mer, station de ski*) en tant que variables illustratives.

La figure 6.1 représente l'histogramme des valeurs propres. Les trois premiers axes donnent un pourcentage cumulé de 76.80% et de 88.88% pour les quatre premiers axes. L'apport d'information du quatrième facteur reste peu élevé, nous avons donc choisi de ne conserver que les plans de projection constitués par les trois premiers facteurs, en dépit du relativement faible pourcentage.

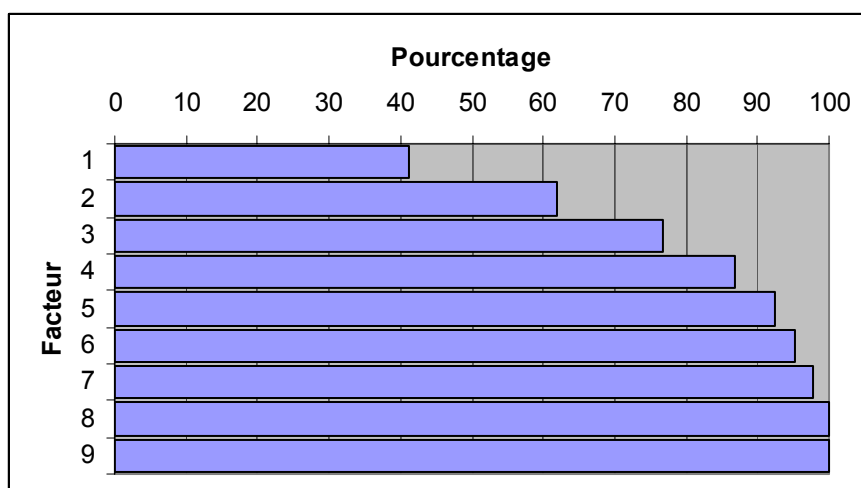


Figure 6.1 : Histogramme des valeurs propres (% cumulé) pour l'ACM.

Sur le premier plan de projection (facteur 2 en fonction du facteur 1, Figure 6.2) nous avons représenté les individus, les variables nominales actives et illustratives. Nous constatons clairement que les 2 classes de chacune des variables nominales sont largement séparées, avec à gauche les faibles valeurs et à droite les fortes valeurs (ceci est confirmé lorsque nous réalisons une classification en 2 classes cf. Figure A.2 en Annexe). Nous remarquons de nouveau à l'aide de la variable *montagneux* que les départements ont une activité agricole plus intense s'il n'y a pas de montagnes. Nous notons de plus que les modalités (« peu » et « beaucoup ») des variables *nb d'exploitations agricoles* et *exploitation avec superficie agricole* se retrouvent projetées sur les mêmes points, ce qui souligne de nouveau leur forte corrélation (cf. Tableau 3.2). Les variables illustratives ne permettent pas de distinguer des corrélations, ou d'interpréter des ressemblances.

Nous retrouvons les départements très montagneux de Corse (2A, 2B), des Alpes (4,5), l'Ariège et la Savoie comme étant des départements peu agricoles ainsi que les départements de la région parisienne (75, 92, 93, 94, 95), tandis que les départements de Bretagne, Pays de Loire et Normandie, pas montagneux, ont une forte productivité agricole. La distinction entre les productions en bovin ou en blé tendre n'apparaît cependant pas sur ce plan.

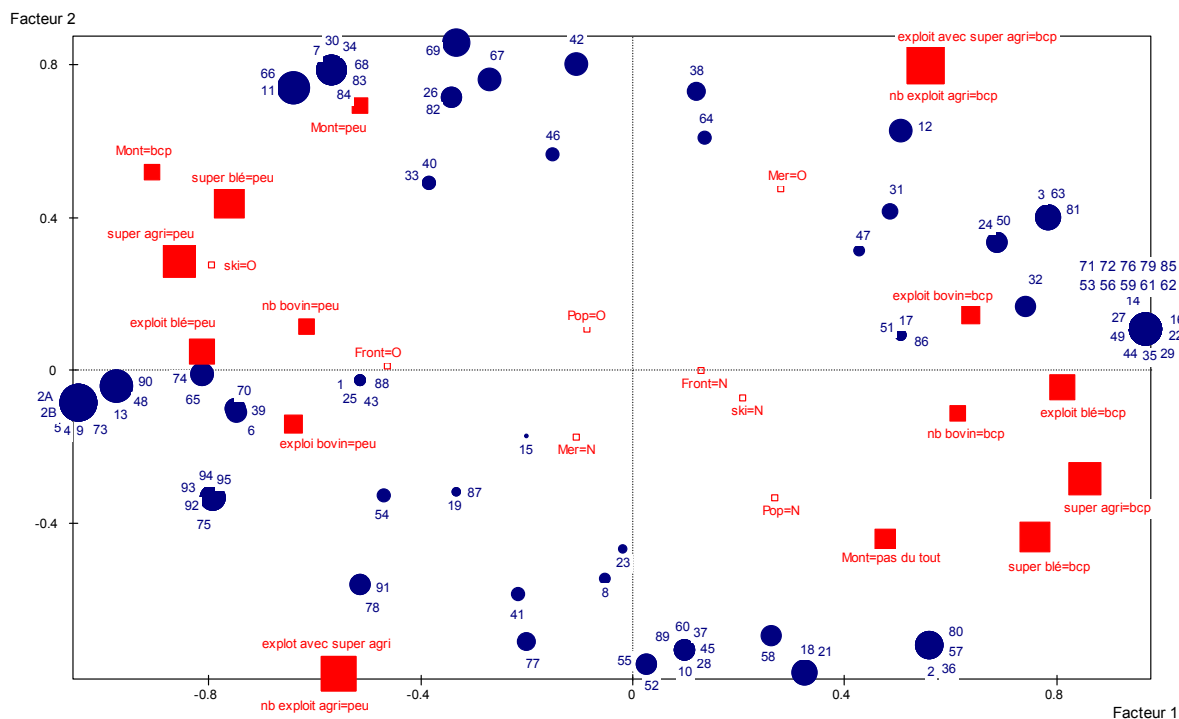


Figure 6.2 : Projection des individus, des nominales actives et illustratives sur le premier plan de projection.

Sur la Figure 6.3 nous représentons le plan factoriel formé par les facteurs 1 et 3. Les variables agricoles sont toujours discriminées selon l'axe 1. Dans ce plan nous constatons que les modalités (« peu », « beaucoup ») des variables liées à la production de blé sont proches des modalités (respectivement « peu », « beaucoup ») des variables caractérisant la superficie, tandis que les variables liées à la production bovine sont éloignées de celles-ci. Ceci nous permet donc de distinguer les types de productions agricoles (bovin ou blé). Nous distinguons ainsi que le groupe des départements de la Marne (51), Charente Maritime (17), Gars (31), Gers (32), Vienne (86) et du Lot et Garonne (47) sont davantage caractérisés par leur production en blé alors que les départements plus au centre (Puits de Dôme (63), Aisne (2), Aveyron (12), Somme (81)) ont une production plus importante en bovin. Cependant la majorité des départements à forte production agricole (départements de Bretagne, Pays de Loire et Normandie) ne se distinguent pas clairement par un type de production. En effet, leur production est aussi bien bovine que céréalière.

De la même façon, nous pouvons distinguer les départements à faible production agricole en 2 groupes ceux qui produisent peu de bovins (6, 11, 66, 26, 82) et ceux produisant peu de blé (13, 48, 90, 65, 74, 7, 84, 34, 83, 68, 30). Les départements de la région parisienne n'ont pas un type de production particulier plus faible ; et ils se retrouvent proche de l'axe 1.

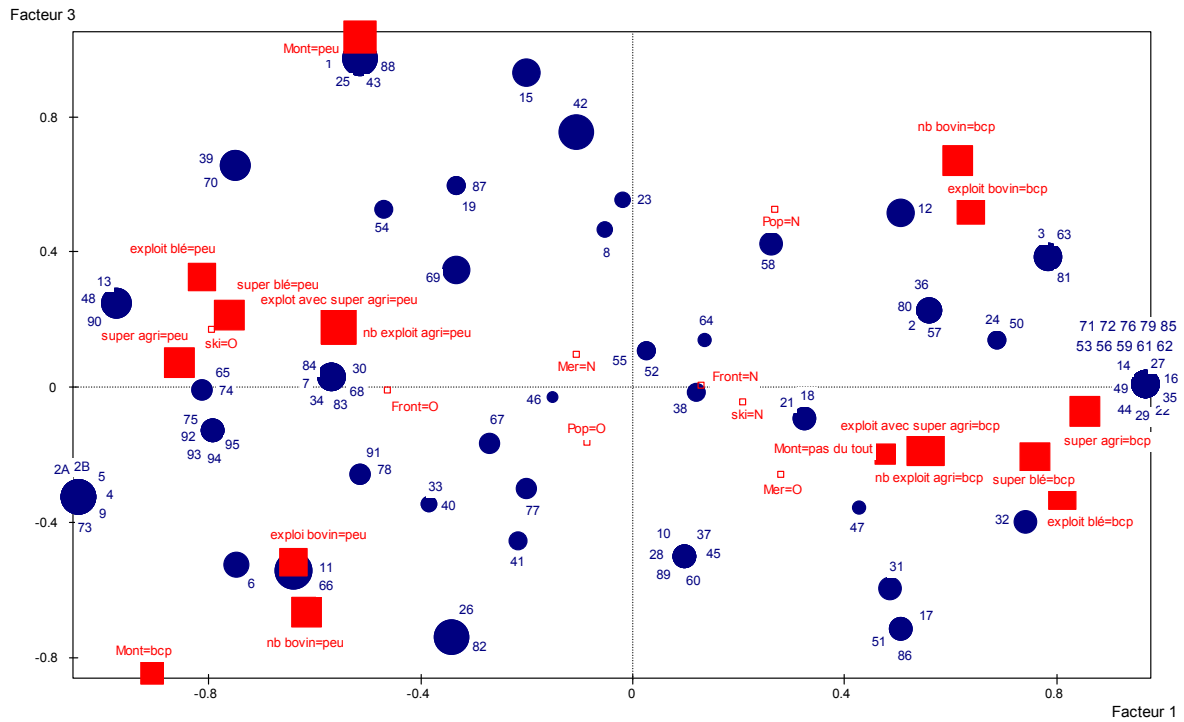


Figure 6.3 : Projection des individus, des nominales actives et illustratives sur le plan factoriel formé par l'axe 1 et 3.

Afin de confirmer nos analyses, nous avons réalisé une classification hiérarchique. L'étude du dendrogramme (en Annexe Figure A.3) et de la courbe des indices pour les vingt premiers noeuds (Figure 6.4) indique une coupure en 5 classes plus marquée qu'une coupure en 4 classes qui pouvait être pressentie par l'étude précédente. Le quotient d'inertie interclasse sur l'inertie totale est 64.73% pour cette coupure en 5 classes.

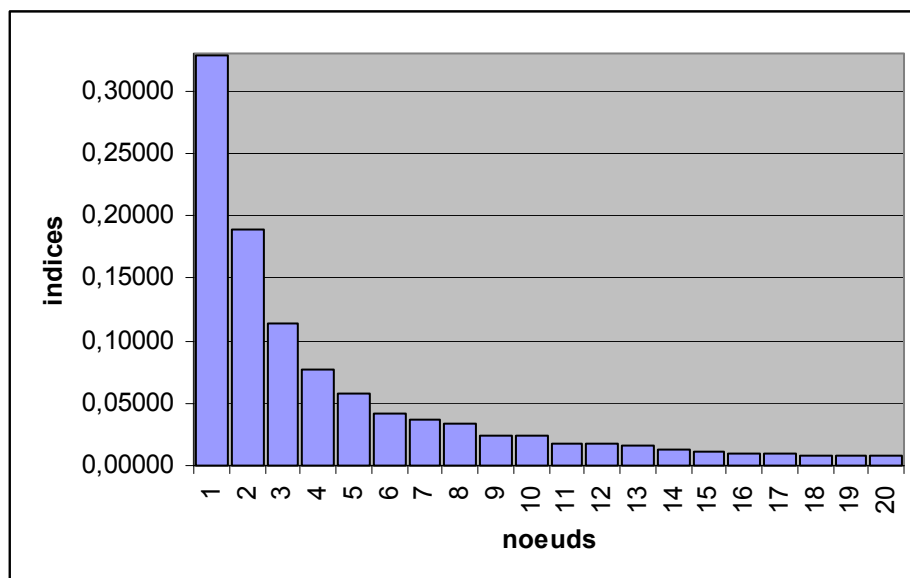


Figure 6.4. Histogramme des indices des 20 premiers noeuds.

Les 5 classes ainsi constituées sont caractérisées de différentes façons (cf. en Annexe Tableau A.5).

La première classe contient 13 départements situés à l'est de la France peu montagneux à très faible production agricole (*nb d'exploitations agricoles, exploitation avec superficie, superficie agricole utilisée, exploitation en blé tendre, superficie en blé tendre*, les variables sur la production en bovins ne caractérisent pas cette classe, ni positivement, ni négativement).

La deuxième classe est constituée par 12 départements qui se situent plutôt dans le sud de la France et qui présentent un nombre positif semblable (ou proche) d'*exploitations agricoles*, et d'*exploitations agricoles avec superficie*. Ce sont des départements peu montagneux qui ne sont pas marqués par une production agricole en bovin ou en blé tendre importante ce qui semble contradictoire avec le nombre d'exploitations. Ces départements sont en fait des départements à faible production agricole qui possèdent un nombre positif d'exploitations proche les uns des autres, mais qui reste faible par rapport à d'autres départements issus d'autres classes. L'ajout d'une classe pour ces variables aurait permis de distinguer cette nuance.

La troisième classe contient 15 départements très montagneux (des Alpes, Pyrénées et Corse). Ces départements n'ont pas une production agricole vraiment similaire, mais ils sont plutôt à faible production en blé tendre (*superficie en blé tendre*).

La classe 4 a un effectif de 26 départements qui se situent dans la partie nord de la France (région parisienne incluse). Ces départements sont avant tout caractérisés par leur faible nombre d'exploitations agricoles et par le fait qu'ils sont peu montagneux. Ils ont cependant un comportement semblable face à leur superficie en blé tendre qui est positif. Là encore ceci signifie qu'ils ont de la superficie agricole, mais elle reste faible par rapport aux autres départements, l'ajout d'une classe pour cette variable aurait également permis de distinguer la nuance.

La classe 5 est constituée de 30 départements se situant dans la partie ouest de la France. Ces départements présentent une ressemblance quant à leur forte production agricole (*nb d'exploitations agricoles, exploitation avec superficie, superficie utilisée*) aussi bien pour la production de blé (*exploitation en blé, superficie en blé tendre*) que de bovins (*exploitations en bovin, effectifs bovins*) ; ils sont de plus peu montagneux.

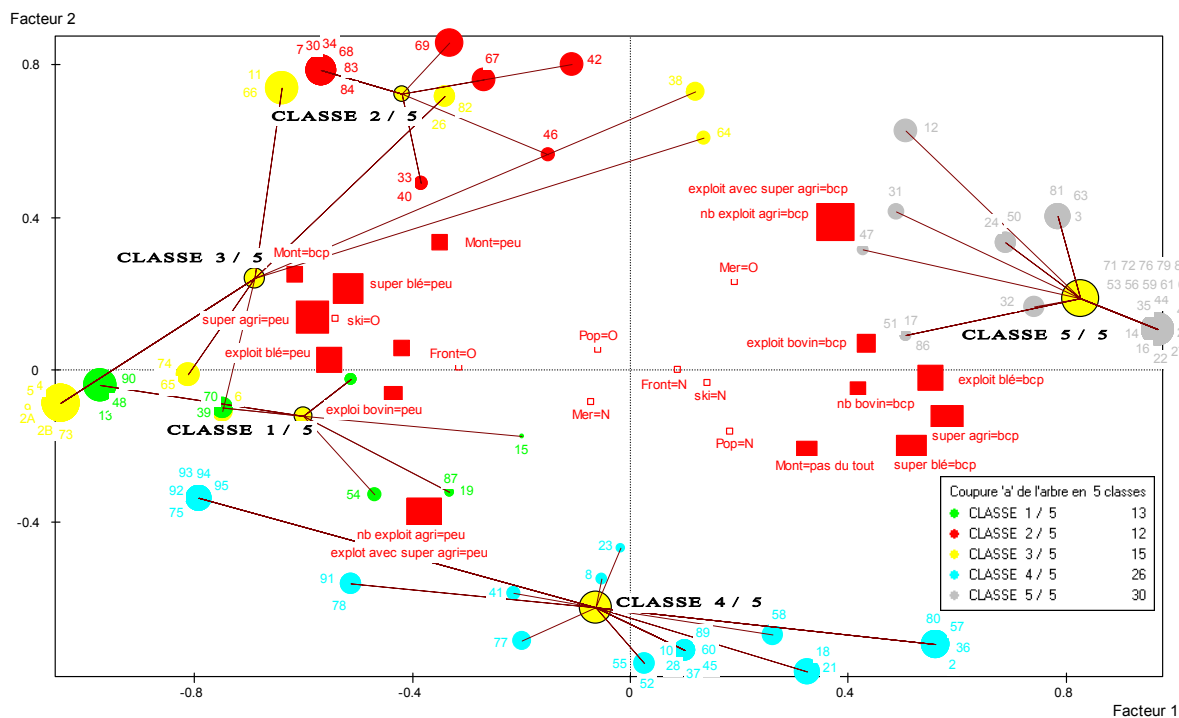


Figure 6.5 : Projection des individus, des nominales actives et illustratives ainsi que des 5 classes sur le premier plan de projection.

Afin d'afficher les parangons sur le premier plan de projection (Figure 6.5) les modalités et les centres des classes sont repositionnés en fonction du barycentre des individus (ce qui explique que les modalités ne se retrouvent pas exactement placés comme sur la Figure 6.2). Nous pouvons constater que des départements pour une classe donnée peuvent être proches de la projection du centre d'une autre classe, ceci s'explique par le fait que les départements d'une classe donnée sont mieux représentés dans un autre plan. Par exemple des départements de la classe 3 sont plus proches de la classe 2 ou d'autres de la classe 1. Ainsi les 3 classes 1, 2 et 3 ne sont pas représentées dans ce plan distinctement d'autres classes. La classe 4 des départements de la partie nord de la France (région parisienne) se retrouve bien proche des modalités « peu » des variables *nb d'exploitations agricoles* et *exploitation avec superficie*. Les départements de la région parisienne se retrouvent les plus éloignés du centre car ils ont également peu de production en bovins. Les départements de la classe 5 sont très éloignés des autres départements : ce sont les départements de la partie ouest de la France à forte production agricole. Cette classe est en effet proche des modalités « beaucoup » de toutes les variables caractérisant la production agricole.

Cette analyse factorielle des correspondances multiples a permis de voir le rôle important de la géographie (par la variable *montagneux*) sur la production agricole. La caractérisation des variables agricoles en 2 classes de modalité « peu » et « beaucoup » est souvent insuffisante pour une étude précise des caractéristiques des départements. Elle permet cependant une représentation rapide pour la séparation des départements en 2 classes (Figure 6.3 et Figure A.2). Les variables nominales illustratives ne jouent pas un rôle discriminant pour l'étude de la production agricole des départements.

7. Conclusion

Cette étude a permis de caractériser la production agricole des départements français. Nous avons ainsi constaté des différences importantes entre les départements français. Ces différences peuvent être dues à la géographie des départements, en particulier leur caractère montagneux et leur superficie, ou encore à la population (avec la région parisienne qui se distingue à ce niveau). Ainsi nous obtenons des regroupements des départements assez nets en fonction de la partie géographique sur le territoire français. Nous avons distingué les départements de la partie ouest de la France à forte production agricole en particulier pour la production en bovin, tandis que la partie centrale à une production plus orientée vers le blé tendre que les bovins. La partie du sud – sud-est constituée des Pyrénées, des Alpes et Corse a une faible production agricole due au relief de ces départements. La partie est et nord de la France se distingue moins, ces départements ont une production agricole plus faible et peu marquée pour un type de production.

Ces analyses peuvent être le prélude à des analyses plus détaillées pour répondre à des questions précises sur la production agricole des départements français.

8. Annexes

altitude max	Superficie	Région	nb de communes	nb de cantons	Préfecture	Département	Numéro
1723	5736	rhône-alpes	443	38	Bourg en Bresse	Ain	1
285	7378	Picardie	831	42	Laon	Aisne	2
1287	7327	auvergne	321	33	Moulins	Allier	3
3400	6944	Provence côte	237	32	Digne	Alpes de Haute	4
4102	5520	Provence côte	179	28	Gap	Alpes Hautes	5
3143	4294	Provence côte	163	40	Nice	Alpes Maritimes	6
1754	5523	Rhône-Alpes	349	33	Privas	Ardèche	7
504	5219	champagne	485	36	Charleville-	Ardennes	8
3080	4890	midi pyrenées	340	20	Foix	Ariège	9
369	6002	champagne	443	32	Troyes	Aube	10
2471	6232	Languedoc	439	34	Carcassonne	Aude	11
1471	8725	midi pyrenées	306	45	Rodez	Aveyron	12
1041	5112	Provence côte	119	44	Marseille	Bouches du Rhône	13
365	5536	basse Normandie	756	40	Caen	Calvados	14
1858	5779	auvergne	269	25	Aurillac	Cantal	15
366	5972	Poitou-Charentes	424	33	Angoulême	Charente	16
172	6848	Poitou-Charentes	484	43	La Rochelle	Charente Maritime	17
434	7228	centre	290	33	Bourges	Cher	18
980	5860	limousin	289	30	Tulle	Corrèze	19
2425	4014	corse	130	24	Ajaccio	Corse du sud	2A
2710	4666	corse	236	25	Bastia	Corse Haute	2B
243	6004	haute Normandie	688	36	Evreux	Eure	27

chômeurs	actifs	nb de chambres d'hôtels	station de ski	montagneux	frontalier	bord de mer	augmentation de la pop. entre 1990 et 1999	densité	Population 1990	Population 1999
106273	243873	165	0	1	1	1	0	89,867154	471019	515 478
858572	232354	77	1	0	0	1	1	72,555299	537259	535 313
19973	147083	145	1	1	1	1	1	47,033574	357710	344 615
8110	59313	138	0	2	0	1	0	20,115639	130883	139 683
5103	54959	238	0	2	0	1	0	22,034601	113300	121 631
62681	432283	705	0	2	0	0	0	235,64648	971829	1 011 866
15531	124168	148	1	1	1	1	0	51,812420	277581	286 160
20917	124431	44	1	0	0	1	1	55,589959	296357	290 124
8104	57061	75	0	2	0	1	0	28,087321	136455	137 347
18363	133096	54	1	0	1	1	0	48,666944	289207	292 099
22849	127253	118	1	2	1	0	0	49,735558	298712	309 952
8955	112897	177	1	1	1	1	1	30,249169	270141	263 924
156189	799061	432	1	1	1	0	0	359,03892	1759371	1 835 407
39391	291208	274	1	0	1	0	0	117,10603	618478	648 299
6925	66139	167	0	1	1	1	1	26,089634	158723	150 772
20051	150088	76	1	0	1	1	1	56,855994	341993	339 544
35959	236661	265	1	0	1	0	0	81,394421	527146	557 389
17734	138269	91	1	0	1	1	1	43,495849	321559	314 388
10160	99609	110	1	0	1	1	1	39,673037	237908	232 484
8150	48300	185	0	2	1	0	1	29,536123	118808	118 558
10219	56225	156	0	2	1	0	0	30,345263	131563	141 591
32376	253057	91	1	0	1	0	0	90,150399	513818	541 263

actif H	60 ans ou plus	40 à 59 ans	25 à 39 ans	15 à 24 ans	0 à 14 ans	F	H	étranger	nationalité française	NB de bénéficiaires du RMI au 31/12/2000	revenu imposable en 2000 euros/hab
134996	97434	138319	114387	63565	101773	259802	255 676	40 903	474575	3653	6845,3
131357	112949	137334	111968	69287	103775	273151	262162	13 940	521373	7691	6279,8
79882	99815	91141	65819	37290	50550	179061	165554	8 834	335781	5526	6637,9
32666	36872	37311	27465	14125	23910	71287	68396	5 783	133900	2310	6478,2
29725	29137	32518	25449	13269	21258	61913	59718	3 790	117841	1051	6708,2
228010	279513	262180	201694	111669	156810	537323	474543	83 962	927904	20117	8149,8
68508	73597	75954	57441	30951	48217	145748	140412	9 330	276830	3564	6283
71609	61973	73665	61567	37229	55690	147310	142814	12 306	277818	5964	5972,4
31479	40735	36069	26072	13614	20857	69943	67404	6 312	131035	3507	5921,9
72184	65030	76264	61415	37094	53396	149500	142599	13 864	278235	4409	6926,5
70401	88457	79646	59712	33054	49083	159596	150356	15 091	294861	9002	5771,1
62068	80164	67702	50131	27437	38490	134028	129896	6 372	257552	2095	6195,1
430220	391770	486195	393363	243000	321079	952153	883254	105 062	1730345	68288	6896,5
156131	131931	165437	136685	96001	118245	335587	312712	10 019	638280	10067	6576,5
36786	43895	40004	29380	16449	21044	76532	74240	1 226	149546	1814	5789,6
81319	88893	90491	67582	38408	54170	174372	165172	6 935	332609	5976	6505,1
127460	153965	146112	105896	63902	87514	287504	269885	7 218	550171	11249	6614,3
75523	81839	83785	62934	34705	51125	160884	153504	10 961	303427	5456	6851,9
53777	71086	61258	44036	24120	31984	119733	112751	112 543	119941	2270	6627,3
27166	29610	32239	24601	12311	19797	61156	57402	12 286	106272	3213	5776
32999	35146	37797	29999	16693	21956	72091	69500	13 387	128204	4014	4786,6
138430	103225	144131	117506	68660	107741	274601	266662	15 510	525753	6372	6921,7

BAC+2	Bac brevet profes	CAP BEP	BEPC	CEP	Aucun diplôme	Etudes en cours	Nationalit é française F	Nationalit é française H	chômeur F	actif F	chômeur H
32814	46462	102362	27703	67575	68432	42051	240622	233953	97856	108877	8417
24627	38090	96126	30512	79259	101049	44693	266917	254456	841497	100997	17075
17815	29682	70557	22499	64282	52090	24197	174928	160853	11506	67201	8467
8982	14225	26998	9971	18854	19794	9207	68689	65211	4441	26647	3669
8718	12981	23444	8635	16988	14112	8957	60193	57648	2902	25234	2201
67062	108716	166934	83563	129695	138966	80053	498340	429564	32387	204273	30294
17419	25342	52968	18551	45336	45560	20628	141235	135595	8957	55660	6574
13192	21394	53281	17704	44508	51896	23810	141609	136209	11563	52822	9354
7553	13613	25069	9935	21807	23673	8947	66924	64111	4459	25582	3645
13807	21276	56923	15880	46842	50409	24135	143103	135132	10242	60912	8121
16332	30500	54884	21868	42929	59229	21105	152559	142302	12424	56852	10425
16190	25299	49579	16914	49293	38971	18439	130992	126560	5283	50829	3672
120890	168922	300484	120541	193274	293210	182347	903865	826480	82807	368841	73382
34471	49986	117806	35778	85584	102021	70340	330908	307372	21130	135077	18261
7255	14236	27953	11720	27400	25228	10676	75969	73577	4024	29353	2901
16759	29056	71796	17931	50534	61001	25133	171044	161565	11652	68769	8399
27533	48471	113525	34236	85793	95241	41781	284158	266013	20222	109201	15737
16879	25557	65968	17399	53416	48806	23067	155716	147711	10131	62746	7603
12444	21519	47919	15883	42847	34384	16135	11606	108335	6054	45832	4106
5992	13885	16240	11371	13577	22773	8160	55849	50423	4025	21134	4125
6743	15337	19452	13217	16821	28864	11586	66676	61528	5298	23226	4921
26693	41308	103290	30740	70421	94743	44068	267545	258208	17714	114627	14662

effectifs bovins	exploitation en bovin	superficie en blé tendre	exploitation en blé tendre	superficie agricole utilisée	exploitation avec superficie agricole	nb d'exploitations agricoles	nb d'essais OGM en 2000	PIB en 1996 (million de F)	Diplômes supérieurs
197608	2962	27608	2654	254497	6145	6317	1	58997	26306
218884	2468	184399	3687	497639	5969	6067	4	58136	17182
532050	4603	47107	3197	502545	7293	7338	2	36579	12943
12763	239	1688	379	165807	2823	2947	0	15014	7742
33971	689	2890	641	95372	2248	2318	0	14224	6538
1512	102	21	11312	61	2538	2620	0	123724	80067
49669	1835	3228	1002	144333	6998	7086	0	27606	12139
281501	2394	67939	2427	309926	3733	3768	0	30339	8649
92055	1597	6080	844	138037	3092	3146	0	13407	5893
63646	839	120035	3268	380917	5674	5802	3	34366	10531
27947	467	1586	563	238738	9311	9370	0	28967	14022
493878	7470	12227	3532	530696	10636	10712	1	27832	10749
13936	158	402	131	162690	5716	5797	0	232546	134660
430834	6393	75941	3620	397753	9036	9099	1	73953	34068
473709	5829	4736	1359	368902	6629	6641	0	14192	5260
160808	2492	61909	4677	376006	8744	8837	3	40434	13164
120694	2290	90248	5471	442993	9866	9999	0	53683	23295
177114	1848	122367	2809	442438	5059	5138	1	35671	12171
316491	5218	2239	1069	233681	6513	6557	0	24458	9369
23970	596	0	0	52238	1381	1381	0	12502	6763
39728	657	162	11	103651	2195	2197	0	13281	7615
191393	3107	150807	3741	385464	7581	7710	0	58219	22259

Tableau A.1 : Extrait du tableau de données.

Le Tableau A.1 présente les données initiales obtenues à partir des différents sites internet. Seules les variables *montagneux* et *augmentation de la population* ont été créées à partir respectivement de la variable *altitude max* et des 2 variables *population en 1999* et *population en 1990*. Ainsi la variable *montagneux* prend la modalité « beaucoup » si *altitude max* est supérieure à 2000, « un peu » si elle est comprise entre 1000 et 2000 et « pas du tout » sinon. La variable *augmentation de la population* prend la valeur 0 si *population en 1999* est supérieure à *population en 1990*, 1 sinon. Les variables nominales *bord de mer*, *frontière* et

stations de ski prennent la valeur 0 si la réponse pour un département donné est oui, et 1 sinon.

Libellé de la variable	Effectif	Poids	Moyenne	Ecart-type	Minimum	Maximum
nb de cantons	100	100.00	36.440	10.238	0.000	69.000
nb de communes	102	102.00	371.343	206.614	1.000	907.000
Superficie	106	106.00	10442.600	42034.100	105.000	432000.000
altitude max	102	102.00	1158.390	1113.630	55.000	4807.000
Population 1999	105	105.00	624222.000	693336.000	6600.000	5944790.000
Population 1990	102	102.00	571284.000	445150.000	27.000	2531860.000
densité	103	103.00	500.179	2232.800	1.720	20246.200
nb de chambres d'hôtels	96	96.00	191.385	182.026	23.000	1495.000
actifs	100	100.00	281220.000	229207.000	31697.000	1128580.000
chômeurs	100	100.00	47178.300	89307.900	2538.000	858572.000
revenu imposable en 2000 euros/hab	100	100.00	6722.770	1262.280	3430.700	13508.200
NB de bénéficiaires du RMI au 31/12/2000	100	100.00	10968.500	12868.600	665.000	68288.000
nationalité française	100	100.00	567447.000	421905.000	70898.000	2445910.000
étranger	100	100.00	82935.200	488560.000	1226.000	4918730.000
H	100	100.00	340805.000	552595.000	36959.000	5389390.000
F	100	100.00	309577.000	237430.000	36549.000	1317760.000
0 à 14 ans	100	100.00	117781.000	124346.000	11132.000	1001770.000
15 à 24 ans	100	100.00	79014.200	65757.100	8276.000	395157.000
25 à 39 ans	100	100.00	131722.000	110113.000	14819.000	602056.000
40 à 59 ans	100	100.00	155560.000	117838.000	19133.000	620603.000
60 ans ou plus	100	100.00	126806.000	83822.500	9396.000	461236.000
actif H	100	100.00	147535.000	115334.000	17685.000	605966.000
chômeur H	100	100.00	17050.300	16771.800	1114.000	96093.000
actif F	100	100.00	133684.000	129984.000	14012.000	947033.000
chômeur F	100	100.00	30128.000	84541.400	1424.000	841497.000
Nationalité française H	100	100.00	274612.000	201066.000	35552.000	1177930.000
Nationalité française F	100	100.00	292834.000	221062.000	11606.000	1267990.000
Etudes en cours	100	100.00	55947.300	49749.300	5562.000	286122.000
Aucun diplôme	100	100.00	90614.300	64730.500	11713.000	395358.000
CEP	100	100.00	74371.500	47428.300	4068.000	315592.000
BEPC	100	100.00	35428.300	27661.000	4911.000	141225.000
CAP BEP	100	100.00	107336.000	74929.300	12891.000	431188.000
Bac brevet profés	100	100.00	52997.500	42916.800	7681.000	227747.000
BAC+2	100	100.00	36834.000	33916.200	4823.000	191363.000
Diplômes supérieurs	100	100.00	39273.600	61402.200	3204.000	513061.000
PIB en 1996 (million de F)	96	96.00	81585.900	100851.000	6573.000	808065.000
nb d'essais OGM en 2000	106	106.00	1.594	3.123	0.000	16.000
nb d'exploitations agricoles	96	96.00	6914.660	3732.840	0.000	18242.000
exploitation avec superficie agricole	96	96.00	6803.020	3684.450	0.000	18183.000
superficie agricole utilisée	96	96.00	289534.000	146942.000	0.000	559062.000
exploitation en blé tendre	96	96.00	2865.310	2217.240	0.000	11312.000
superficie en blé tendre	96	96.00	51005.300	52412.900	0.000	195012.000
exploitation en bovin	96	96.00	2937.590	2534.310	0.000	12444.000
effectifs bovins	96	96.00	211030.000	196406.000	0.000	764522.000

Tableau A.2 : Statistique des variables continues.

Libellé de la variable	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4
Superficie	0.47	0.33180	0.14	0.20	0.07	4.49
0 à 14 ans	9.89	0.52265	11.49	0.60	76.75	1.07
15 à 24 ans	6.66	0.24637	6.27	0.09	2.46	2.53
25 à 39 ans	11.11	0.28771	12.56	0.27	5.66	2.90
40 à 59 ans	13.24	0.22099	11.78	0.12	5.79	0.35
60 ans ou plus	10.87	0.15495	5.25	0.18	7.58	3.96
nb d'essais OGM en 2000	0.00	3.64264	0.00	0.00	0.00	0.00
nb d'exploitations agricoles	0.58	0.29510	0.16	0.01	0.14	4.66
exploitation avec superficie agricole	0.57	0.29852	0.16	0.01	0.14	4.64
superficie agricole utilisée	24.21	0.23779	18.35	18.19	0.04	23.31
exploitation en blé tendre	0.24	0.45611	0.10	0.12	0.00	0.05
superficie en blé tendre	4.27	0.88485	2.41	39.52	1.18	42.56
exploitation en bovin	0.25	0.52713	0.32	0.48	0.00	0.05
effectifs bovins	17.65	0.55536	31.01	40.23	0.18	9.43

Tableau A.3 : Contributions des variables actives de l'analyse factorielle des correspondances.

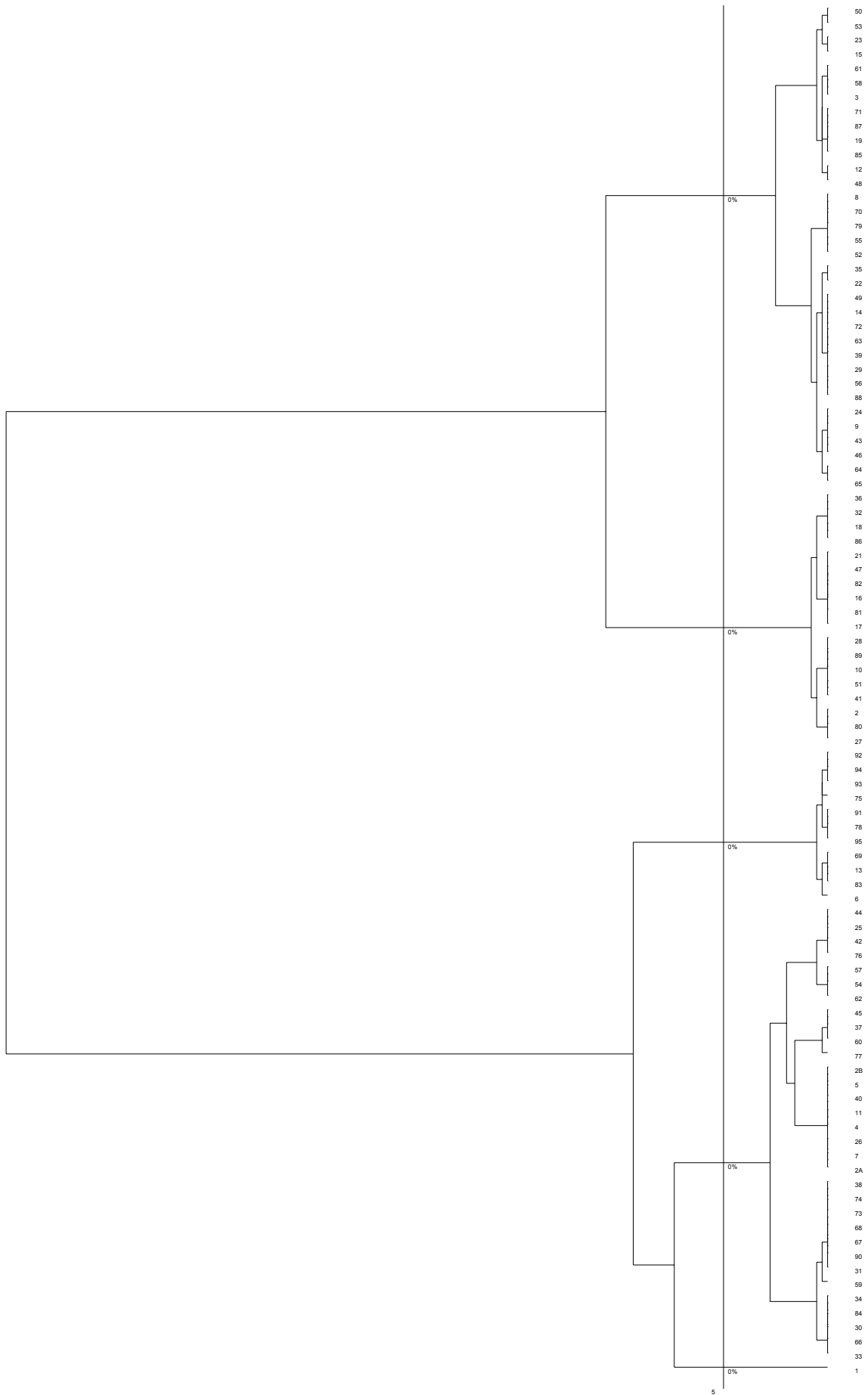


Figure A.1 : Dendrogramme après l'analyse factorielle des correspondances.

	nb d'exploitations agricoles	exploitation avec superficie agricole	superficie agricole utilisée	exploitation en blé tendre	superficie en blé tendre	exploitation en bovin	effectifs bovins
Classe 1 1							
Classe 2 Effectif : 26 7, 11, 2A, 2B, 26, 30, 31, 33, 34, 38, 40, 42, 44, 54, 57, 59, 62, 66, 67, 68, 73, 74, 76, 77, 84, 90			-2.55				
Classe 3 Effectif : 11 6, 13, 69, 75, 78, 83, 91, 92, 93, 94, 95	-4.13	-4.13	-5.52		-2.70	-3.64	-3.54
Classe 4 Effectif : 23 2, 4, 5, 10, 16, 17, 18, 21, 27, 28, 32, 36, 37, 41, 45, 47, 51, 60, 80, 81, 82, 86, 89			3.42		5.89	-2.40	-2.43
Classe 5 Effectif : 35 3, 8, 9, 12, 14, 15, 19, 22, 23, 24, 25, 29, 35, 39, 43, 46, 48, 49, 50, 52, 53, 55, 56, 58, 61, 63, 64, 65, 70, 71	2.50	2.49	3.01			6.61	6.21

Tableau A.4 : Valeurs tests en fonction des classes et des variables actives pour une coupure de 5 classes après l'analyse factorielles des correspondances.

Le tableau A.4 ne présente que les variables dont la valeur test dépasse 2 pour une des 5 classes.

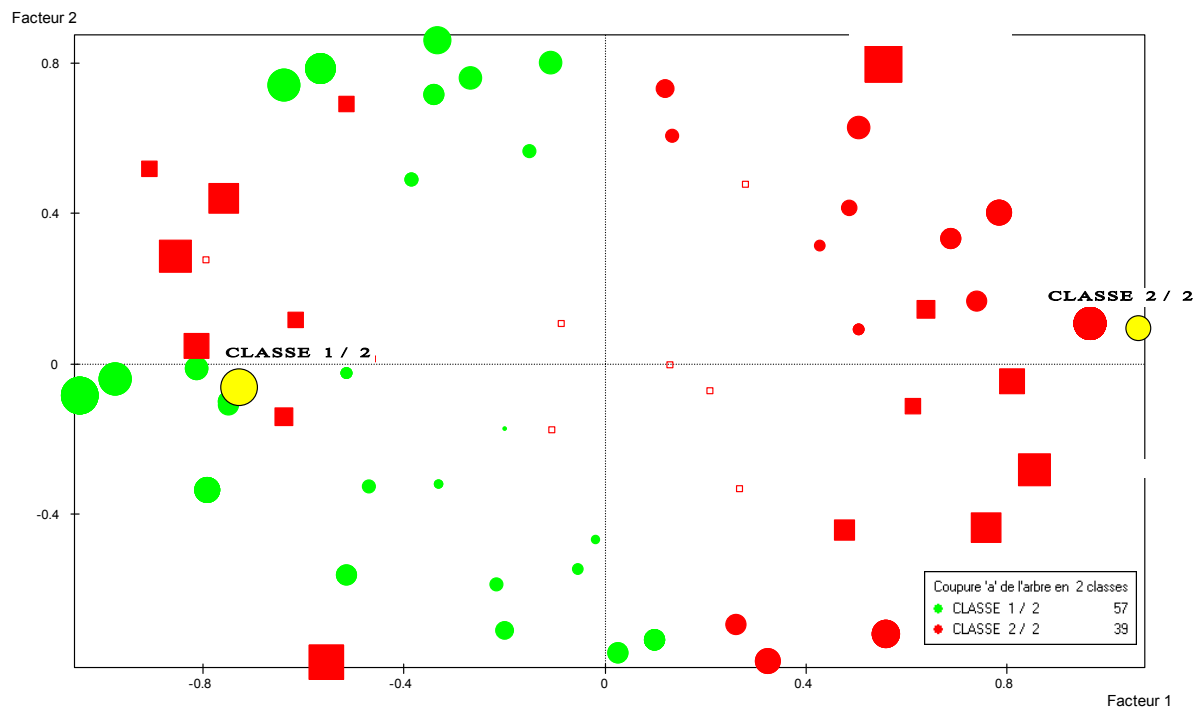


Figure A.2 : Projection des individus, des variables nominales actives et illustratives dans le premier plan de projection avec une classification en 2 classes.

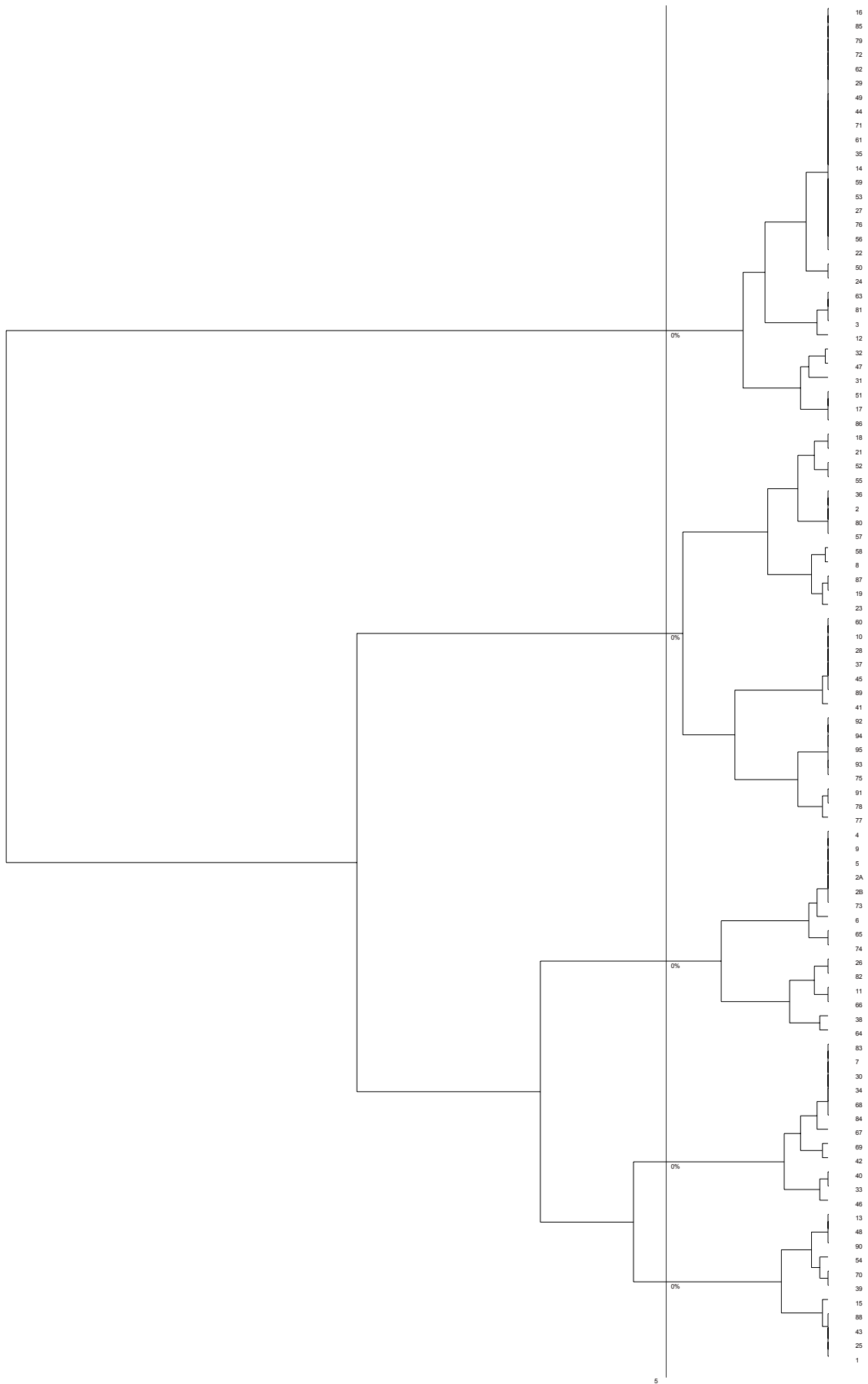


Figure A.3 : Dendrogramme après l'analyse factorielle des correspondances multiples.

Classe 5 3, 12, 14, 16, 17, 22, 24, 27, 29, 31, 32, 35, 44, 47, 49, 50, 51, 53, 56, 59, 61, 62, 63, 71, 72, 76, 79, 81, 85, 86	Classe 4 2, 8, 10, 18, 21, 23, 28, 36, 37, 41, 45, 52, 55, 57, 58, 60, 75, 77, 78, 80, 89, 91, 92, 93, 94, 95	Classe 3 4, 5, 6, 9, 11, 2A, 2B, 26, 38, 64, 65, 66, 73, 74, 82	Classe 2 7, 30, 33, 34, 40, 42, 46, 67, 68, 69, 83, 84	Classe 1 1, 13, 15, 19, 25, 39, 43, 48, 54, 70, 87, 88, 90	Modalités		
					pas	peu	bcp
3.22	5.41	-4.99		-3.10	montagneux		
	-3.64	-2.37	3.65	4.65			
	-2.67	8.20					
-7.03	6.32		-3.69	3.90	nb d'exploitations agricoles		
7.03	-6.32		3.69	-3.90			
-7.03	6.32		-3.69	3.90	exploitation avec superficie		
7.03	-6.32		3.69	-3.90			
-6.39		3.56	3.69	3.13	superficie agricole utilisée		
6.39		-3.56	-3.69	-3.13			
-7.03			2.90	3.90	exploitation en blé tendre		
7.03			-2.90	-3.90			
-5.30	-3.03	4.30	3.69	3.13	superficie en blé tendre		
5.30	3.03	-4.30	-3.69	-3.13			
-5.30	2.55				exploitation en bovin		
5.30	-2.55						
-3.82		2.90	2.90		peu		
3.82		-2.90	-2.90		bcp		

Tableau A.5 : Valeurs tests en fonction des classes et des variables actives pour une coupure de 5 classes après l'analyse factorielles des correspondances multiples.