ANALYSE DES DONNEES

Cours 2ème partie

IV. Analyse Factorielle des Correspondances Multiples

Le but de l'Analyse Factorielle des Correspondances Multiple (AFCM), comme celui de l'AFC, est de détecter des liens entre variables qualitatives, et de positionner les individus par rapport à ces liens. La méthode AFC vue dans le chapitre précédent ne traite formellement que du lien entre 2 variables. La méthode AFCM permet de généraliser cette étude à autant de variables qualitatives que l'on souhaite.

IV.1. Précautions avant de passer de 2 à plus de deux variables

Cette généralisation ne se fait pas sans un minimum de précautions, en raison, d'une part, du fait que dès que l'on a plus de deux variables, on est confronté aux problèmes d'interactions entre ces variables, et, d'autre part, de la non possibilité de généralisation de tous les aspects de l'interprétation d'une AFC.
Nous allons illustrer les problèmes d'interaction avec la présentation du paradoxe de Simpson.

IV.2. Le paradoxe de Simpson

Exemple 1 : Barouf à Bombach
Au cours d'un débat télévisé sur « la femme et les études scientifiques », on aborde la question de la réussite au bac S au cours de l'année précédente.
Un premier participant fait état d'un dossier qui fournit les statistiques ville par ville. A propos de la ville de Bombach, on y trouve les résultats suivants :

	Résultats au bac S		
	+ (succès)	- (échec)	Total
Garçons	24	36	60
Filles	36	24	60

Conclusion : les filles réussissent mieux que les garçons : la différence est de 20 points en faveur des filles.

De ce tableau, on déduit les proportions de réussite pour les garçons et pour les filles : Garçons : 24/60=40%, Filles : 36/60=60%.
Mais un deuxième participant fait état d'un dossier plus détaillé, qui fournit les résultats lycée par lycée.
Dans la ville de Bombach, les lycéens sont répartis en deux lycées : Anastase et Bénédicte, et les statistiques relatives à chacun de ces deux lycées figurent ci-contre.

	Anastase		
	Résultats au bac S		
	+ (succès)	- (échec)	Total
Garçons	15	35	50
Filles	1	9	10

	Bénédicte		
	Résultats au bac S		
	+ (succès)	- (échec)	Total
Garçons	9	1	10
Filles	35	15	50

Notons que bien entendu, en ajoutant case à case les deux tableaux, on retrouvera le tableau précédent.
Or, de chacun de ces deux tableaux, on déduit les proportions de réussite suivantes :
Anastase : Garçons : 15/50=30%, Filles : 1/10=10% ; Bénédicte : Garçons : 9/10=90%, Filles : 35/50=70%.
Conclusion : à l'intérieur de chaque lycée, les garçons réussissent mieux que les filles, la différence des pourcentages est la même dans les deux lycées : 20 points en faveur des garçons (donc valeur opposée à la valeur globale obtenue plus haut).
Cet exercice sur les proportions est extrait du polycopié « Documents de cours de statistique. Procédures statistiques fondamentales ». Enseignement de statistique de H. Rouanet pour le C1 de psychologie générale, licence de psychologie, Université Paris V.
Exemple 2 : la criminalité en Floride
Il s'agit de 4764 homicides jugés en Floride de 1973 à 1979. Ces données ont été publiées dans le New-York Times du 11 mars 1979. Elles ont été maintes fois utilisées par les statisticiens (cf site internet http://www.cict.fr/personnel/stpierre/expose-16-01-98/expose.html).

On dispose des données sous forme de tableau de contingence complète.
Calculer les pourcentages de peine de mort chez les meurtriers blancs et chez les meurtriers noirs, toutes victimes confondues, puis en distinguant les victimes noires et les victimes blanches. Conclure.

		sentence	
meurtrier	victime	peine de mort	autre peine
blanc	blanc	72	2074
	noir	0	111
noir	blanc	48	239
	noir	11	2209

IV.3. Les différents types de tableaux

IV.3.1. Le tableau des données brutes
Le premier tableau qui vient à l'idée est celui des données brutes, indiquant pour chaque individu les valeurs des variables qualitatives.
Exemple d'un tel tableau : une enquête commandée par une cave coopérative auprès de viticulteurs de la région Languedoc-Roussillon a été menée dans le but de mieux connaître les attentes de ces derniers du point de vue des services de cette coopérative.

Nous sélectionnons un sous-échantillon de 154 viticulteurs de telle manière qu'il n'y ait pas de données manquantes. Le tableau suivant rassemble les variables :
région: région de production du vin
statut: statut juridique (EARL, EI=exploitation individuelle, GAEC, SCEA)
adhérent: A=adhérent à la cave coopérative, P=prospect, c'est-à-dire non adhérent
valorisation: circuit de distribution de la production
vin: principale catégorie de vin produit.
Dans la première colonne figure le numéro du viticulteur. Ce numéro constitue un identificateur.

no	région	statut	adhérent	valorisation	vin
1	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
2	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
3	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
5	MINERVOIS-CORBIERES	GAEC	A	Cave coopérative	AOC
7	MINERVOIS-CORBIERES	EARL	A	Mixte	AOC
8	CARCASSONNE	EARL	A	Cave coopérative	VDP
9	CARCASSONNE	EI	A	Cave coopérative	VDP
10	CARCASSONNE	GAEC	A	Cave coopérative	VDP
11	CARCASSONNE	SCEA	A	Cave coopérative	AOC
12	CARCASSONNE	EI	P	Cave coopérative	VDP
15	MINERVOIS-CORBIERES	GAEC	A	Cave coopérative	VDP
16	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
17	NARBONNE-HERAULT	EI	A	Cave coopérative	VDP
18	NARBONNE-HERAULT	EI	A	Cave coopérative	VDP
19	NARBONNE-HERAULT	EI	A	Cave coopérative	VDP
21	NARBONNE-HERAULT	GAEC	P	Cave coopérative	VDP
22	MINERVOIS-CORBIERES	SCEA	A	Cave coopérative	AOC
25	NARBONNE-HERAULT	EARL	A	Cave coopérative	AOC
26	NARBONNE-HERAULT	SCEA	A	Cave particulière	AOC
27	NARBONNE-HERAULT	EI	A	Cave coopérative	VDP
29	NARBONNE-HERAULT	EI	A	Cave particulière	AOC
30	NARBONNE-HERAULT	EI	P	Cave particulière	VDP
31	NARBONNE-HERAULT	EI	A	Cave coopérative	VDP
32	NARBONNE-HERAULT	EI	P	Cave coopérative	VDP
33	NARBONNE-HERAULT	SCEA	P	Cave particulière	VDP
34	NARBONNE-HERAULT	SCEA	A	Cave coopérative	VDP
35	MINERVOIS-CORBIERES	SCEA	A	Cave coopérative	AOC
36	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
37	MINERVOIS-CORBIERES	EARL	P	Cave particulière	AOC
38	NARBONNE-HERAULT	GAEC	A	Cave coopérative	VDP
39	NARBONNE-HERAULT	GAEC	P	Cave particulière	VDP
40	CARCASSONNE	EI	P	Cave particulière	VDP
41	CARCASSONNE	SCEA	P	Cave particulière	VDP
42	CARCASSONNE	EARL	A	Cave coopérative	VDP
43	CARCASSONNE	EI	P	Cave coopérative	VDP
44	BRAM	EARL	A	Cave coopérative	AOC
45	NARBONNE-HERAULT	SCEA	A	Cave coopérative	VDP
46	NARBONNE-HERAULT	EI	A	Cave coopérative	VDP
47	NARBONNE-HERAULT	EI	P	Cave coopérative	VDP
48	NARBONNE-HERAULT	EI	A	Cave coopérative	AOC
49	NARBONNE-HERAULT	SCEA	P	Cave particulière	AOC
50	NARBONNE-HERAULT	EI	A	Cave coopérative	VDP
52	NARBONNE-HERAULT	EI	A	Cave coopérative	VDP
53	MINERVOIS-CORBIERES	EI	A	Cave coopérative	VDP
54	MINERVOIS-CORBIERES	EARL	A	Cave coopérative	AOC
56	MINERVOIS-CORBIERES	GAEC	A	Cave coopérative	VDP
57	MINERVOIS-CORBIERES	EI	A	Cave coopérative	VDP
58	MINERVOIS-CORBIERES	EI	A	Mixte	AOC
59	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
60	MINERVOIS-CORBIERES	EI	A	Mixte	AOC
61	MINERVOIS-CORBIERES	EARL	A	Cave particulière	VDP
62	MINERVOIS-CORBIERES	SCEA	A	Cave coopérative	AOC
63	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
64	MINERVOIS-CORBIERES	SCEA	A	Cave particulière	AOC
65	MINERVOIS-CORBIERES	SCEA	A	Cave particulière	AOC
66	MINERVOIS-CORBIERES	EI	P	Cave coopérative	AOC
67	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
68	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
69	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
70	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
71	MINERVOIS-CORBIERES	EI	P	Cave coopérative	AOC
72	MINERVOIS-CORBIERES	EARL	P	Cave coopérative	VDP
74	MINERVOIS-CORBIERES	GAEC	A	Cave coopérative	VDP
75	MINERVOIS-CORBIERES	EI	P	Cave coopérative	AOC
76	MINERVOIS-CORBIERES	EARL	P	Mixte	AOC
77	MINERVOIS-CORBIERES	EI	P	Cave coopérative	AOC
78	MINERVOIS-CORBIERES	EI	A	Mixte	VDP
79	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
84	MINERVOIS-CORBIERES	GAEC	A	Cave coopérative	AOC
85	MINERVOIS-CORBIERES	EI	P	Cave coopérative	AOC
86	BRAM	EARL	A	Cave coopérative	AO-VDQS
87	MINERVOIS-CORBIERES	SCEA	A	Cave particulière	AOC
88	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
89	MINERVOIS-CORBIERES	SCEA	A	Cave particulière	AOC
90	MINERVOIS-CORBIERES	GAEC	A	Cave coopérative	AOC
91	MINERVOIS-CORBIERES	EI	P	Cave particulière	AOC
94	MINERVOIS-CORBIERES	EI	P	Cave coopérative	AOC

no	région	statut	adhérent	valorisation	vin
95	MINERVOIS-CORBIERES	EI	P	Cave coopérative	AOC
97	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
98	MINERVOIS-CORBIERES	EI	P	Cave coopérative	AOC
99	MINERVOIS-CORBIERES	EI	P	Cave coopérative	AOC
100	MINERVOIS-CORBIERES	SCEA	P	Cave coopérative	AOC
101	MINERVOIS-CORBIERES	EI	A	Cave coopérative	VDP
102	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
103	MINERVOIS-CORBIERES	EARL	P	Cave particulière	AOC
104	MINERVOIS-CORBIERES	GAEC	A	Cave coopérative	AOC
105	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
106	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
107	NARBONNE-HERAULT	EI	A	Cave coopérative	VDP
108	NARBONNE-HERAULT	EARL	A	Cave coopérative	AOC
109	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
110	MINERVOIS-CORBIERES	GAEC	A	Cave particulière	AOC
112	MINERVOIS-CORBIERES	EI	A	Cave coopérative	VDP
113	NARBONNE-HERAULT	EI	A	Cave coopérative	VDP
114	NARBONNE-HERAULT	EI	A	Cave coopérative	VDP
115	NARBONNE-HERAULT	EI	A	Cave coopérative	VDP
116	NARBONNE-HERAULT	EI	P	Cave coopérative	VDP
117	NARBONNE-HERAULT	GAEC	A	Cave coopérative	VDP
118	NARBONNE-HERAULT	SCEA	P	Cave particulière	AOC
119	NARBONNE-HERAULT	EI	P	Cave coopérative	VDP
120	NARBONNE-HERAULT	EI	A	Cave coopérative	VDP
124	NARBONNE-HERAULT	GAEC	P	Cave particulière	VDP
125	NARBONNE-HERAULT	EI	A	Cave coopérative	VDP
126	NARBONNE-HERAULT	EI	A	Cave coopérative	VDP
127	NARBONNE-HERAULT	EI	P	Cave particulière	VDP
128	NARBONNE-HERAULT	EI	P	Cave coopérative	VDP
129	NARBONNE-HERAULT	SCEA	A	Cave particulière	VDP
131	CARCASSONNE	EI	A	Cave coopérative	VDP
135	CARCASSONNE	EARL	A	Cave coopérative	VDP
136	CARCASSONNE	EI	A	Cave coopérative	AO-VDQS
137	CARCASSONNE	EI	A	Cave coopérative	AO-VDQS
138	CARCASSONNE	EI	A	Cave coopérative	AO-VDQS
139	BRAM	EI	A	Cave coopérative	VDP
140	CARCASSONNE	EI	A	Cave coopérative	AO-VDQS
142	CARCASSONNE	EARL	A	Mixte	AOC
143	CARCASSONNE	EI	A	Cave coopérative	AOC
144	CARCASSONNE	EI	A	Cave particulière	AOC
146	CARCASSONNE	EI	A	Cave coopérative	VDP
147	CARCASSONNE	EI	A	Cave coopérative	AOC
148	CARCASSONNE	GAEC	A	Cave coopérative	VDP
149	CARCASSONNE	EI	P	Cave coopérative	AOC
150	CARCASSONNE	EI	P	Cave coopérative	VDP
151	CARCASSONNE	EI	P	Cave coopérative	AOC
152	CARCASSONNE	EI	P	Cave coopérative	AOC
153	CARCASSONNE	EI	A	Cave coopérative	AOC
154	CARCASSONNE	EI	A	Cave coopérative	VDP
156	CARCASSONNE	EI	A	Cave particulière	AO-VDQS
157	CARCASSONNE	EARL	A	Mixte	VDP
159	BRAM	GAEC	A	Cave coopérative	AOC
160	BRAM	EARL	A	Cave coopérative	AO-VDQS
161	BRAM	EI	A	Cave coopérative	AO-VDQS
162	BRAM	EI	A	Mixte	AO-VDQS
163	BRAM	GAEC	A	Cave coopérative	VDP
164	BRAM	EARL	A	Cave particulière	AOC
165	NARBONNE-HERAULT	EI	P	Cave coopérative	AOC
166	NARBONNE-HERAULT	GAEC	A	Cave coopérative	VDP
167	CARCASSONNE	EI	P	Cave particulière	VDP
168	NARBONNE-HERAULT	GAEC	A	Cave coopérative	VDP
169	NARBONNE-HERAULT	GAEC	P	Cave particulière	AOC
170	NARBONNE-HERAULT	EI	P	Cave coopérative	VDP
171	NARBONNE-HERAULT	GAEC	P	Cave coopérative	VDP
172	NARBONNE-HERAULT	EARL	A	Cave particulière	AOC
173	CARCASSONNE	EI	A	Cave particulière	AOC
174	MINERVOIS-CORBIERES	EI	P	Cave particulière	AOC
175	BRAM	EI	A	Cave coopérative	VDP
176	BRAM	GAEC	A	Cave coopérative	VDP
177	BRAM	EI	A	Cave coopérative	VDP
178	BRAM	EARL	A	Cave coopérative	AO-VDQS
179	BRAM	GAEC	A	Cave coopérative	VDP
180	MINERVOIS-CORBIERES	EI	P	Cave coopérative	AOC
182	MINERVOIS-CORBIERES	EI	P	Cave coopérative	AOC
183	MINERVOIS-CORBIERES	GAEC	A	Cave coopérative	AOC
184	MINERVOIS-CORBIERES	EARL	A	Cave coopérative	AOC
185	CARCASSONNE	EARL	P	Mixte	VDP

IV.3.2. Le tableau disjonctif complet

C'est un tableau dans lequel chaque ligne correspond à un individu, et chaque colonne à une modalité. Les cases du tableau contiennent 1 si l'individu de la ligne prend la modalité de la colonne, et 0 sinon.
Un extrait du tableau disjonctif complet issu du tableau de données précédent est le suivant :

no	BRAM	CARC	MC	NH	EARL	EI	GAEC	SCEA	A	P	Mixte	Coop	Part	VDQS	AO-VDQS	AOC	VDP
1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0
2	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0
3	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0
5	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0	1	0
7	0	0	1	0	1	0	0	0	1	0	1	0	0	0	0	1	0
8	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	1
9	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1
…																	
185	0	1	0	0	1	0	0	0	0	1	1	0	0	0	0	0	1

Ces deux tableaux contiennent l'information la plus complète qui soit (information par individu). Le second est une « quantification » du premier, c'est l'avantage qu'il présente. On remarquera qu'il occupe plus de place que le précédent.
Exercice : quelles valeurs ont les sommes des lignes ? Des colonnes ?

IV.3.3. Le tableau des données groupées

Quand l'identité des individus n'est pas d'intérêt, ou quand on n'aura pas besoin d'identifier les individus par la suite, on peut simplifier le tableau en comptant le nombre d'individus correspondant à chaque profil. Le tableau précédent peut ainsi être simplifié comme ci-contre.

Mais sa présentation peut être optimisée, en créant un tableau dans lequel les noms des modalités des variables ne sont pas répétés. C'est ce que l'on fait dans le paragraphe qui suit.

région	statut	adhérent	valorisation	vin	Effectif
BRAM	EARL	A	coop	AO-VDQS	3
BRAM	EARL	A	coop	AOC	1
BRAM	EARL	A	part	AOC	1
BRAM	EI	A	mixte	AO-VDQS	1
BRAM	EI	A	coop	AO-VDQS	1
BRAM	EI	A	coop	VDP	3
BRAM	GAEC	A	coop	AOC	1
BRAM	GAEC	A	coop	VDP	3
CARC	EARL	A	mixte	AOC	1
CARC	EARL	A	mixte	VDP	1
CARC	EARL	A	coop	VDP	3
CARC	EARL	P	mixte	VDP	1
CARC	EI	A	coop	VDQS	4
CARC	EI	A	coop	AOC	3
CARC	EI	A	coop	VDP	4
CARC	EI	A	part	AO-VDQS	1
CARC	EI	A	part	AOC	2
CARC	EI	P	coop	AOC	3
CARC	EI	P	coop	VDP	3
CARC	EI	P	part	VDP	2
CARC	GAEC	A	coop	VDP	2
CARC	SCEA	A	coop	AOC	1
CARC	SCEA	P	part	VDP	1
MC	EARL	A	mixte	AOC	1
MC	EARL	A	coop	AOC	2
MC	EARL	A	part	VDP	1
MC	EARL	P	mixte	AOC	1
MC	EARL	P	coop	VDP	1
MC	EARL	P	part	AOC	2
MC	EI	A	mixte	AOC	2
MC	EI	A	mixte	VDP	1
MC	EI	A	coop	AOC	18
MC	EI	A	coop	VDP	4
MC	EI	P	coop	AOC	11
MC	EI	P	part	AOC	2
MC	GAEC	A	coop	AOC	5
MC	GAEC	A	coop	VDP	3
MC	GAEC	A	part	AOC	1
MC	SCEA	A	coop	AOC	3
MC	SCEA	A	part	AOC	4
MC	SCEA	P	coop	AOC	1
NH	EARL	A	coop	AOC	2
NH	EARL	A	part	AOC	1
NH	EI	A	coop	AOC	1
NH	EI	A	coop	VDP	15
NH	EI	A	part	AOC	1
NH	EI	P	coop	AOC	1
NH	EI	P	coop	VDP	6
NH	EI	P	part	VDP	2
NH	GAEC	A	coop	VDP	4
NH	GAEC	P	coop	VDP	2
NH	GAEC	P	part	AOC	1
NH	GAEC	P	part	VDP	2
NH	SCEA	A	coop	VDP	2
NH	SCEA	A	part	AOC	1
NH	SCEA	A	part	VDP	1
NH	SCEA	P	part	AOC	2
NH	SCEA	P	part	VDP	1

IV.3.4. Le tableau de contingence complète

Ce tableau contient les mêmes informations qu'un tableau de données groupées, mais dans une présentation plus synthétique.

		A										P					
		mixte			Coop			part			mixte		coop		part	
		AO-VDQS	AOC	VDP	AO-VDQS	AOC	VDP	AO-VDQS	AOC	VDP	AOC	VDP	AOC	VDP	AOC	VDP
BRAM	**EARL**	0	0	0	3	1	0	0	1	0	0	0	0	0	0	0
	EI	1	0	0	1	0	3	0	0	0	0	0	0	0	0	0
	GAEC	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0
	SCEA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CARC	**EARL**	0	1	1	0	0	3	0	0	0	0	1	0	0	0	0
	EI	0	0	0	4	3	4	1	2	0	0	0	3	3	0	2
	GAEC	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
	SCEA	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
MC	**EARL**	0	1	0	0	2	0	0	0	1	1	0	0	1	2	0
	EI	0	2	1	0	18	4	0	0	0	0	0	11	0	2	0
	GAEC	0	0	0	0	5	3	0	1	0	0	0	0	0	0	0
	SCEA	0	0	0	0	3	0	0	4	0	0	0	1	0	0	0
NH	**EARL**	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0
	EI	0	0	0	0	1	15	0	1	0	0	0	1	6	0	2
	GAEC	0	0	0	0	0	4	0	0	0	0	0	0	2	1	2
	SCEA	0	0	0	0	0	2	0	1	1	0	0	0	0	2	1

Exercice : que valent les sommes des lignes ? des colonnes ?
Remarque 1 : quand il n'y a pas beaucoup d'individus et beaucoup de modalités de variables, ce tableau risque d'être plus volumineux que celui des données groupées, voire celui des données brutes.
Remarque 2 : parfois, les données se présentent uniquement sous forme de données groupées ou de tableau de contingence complète. C'est le cas lorsqu'elles ont été recueillies par estimation. Par exemple, le tableau suivant restitue le tonnage transporté par bateaux selon les types de bateaux. Les données n'ont jamais été recueillies tonne par tonne, mais des estimations ont été faites pour arriver à ces valeurs. Elles concernent le trafic annuel par bateaux.

typenavire	pétrolier				Vraquier				classique				porte-conteneur				autre			
Typepays immatriculation	age0-4ans	age5-9ans	age10-14ans	age +15ans	age0-4ans	age5-9ans	age10-14ans	age +15ans	age0-4ans	age5-9ans	age10-14ans	age +15ans	age0-4ans	age5-9ans	age10-14ans	age +15ans	age0-4ans	age5-9ans	age10-14ans	age+15ans
developpe	8.2663	10.8381	35.3615	37.3822	9.8878	22.9562	13.9673	22.3339	2.7760	5.6231	6.2637	9.0634	2.3403	3.2238	2.9373	3.4387	3.5265	6.3835	4.8880	7.5218
libimmatr	12.2924	6.6584	23.7653	59.7205	6.0698	19.6879	15.9526	36.1074	2.8525	4.6824	7.8847	11.4906	1.7933	1.4574	1.0836	1.0836	1.9359	2.7843	2.7724	4.4573
e-orienta	1.1820	1.6390	2.3718	2.6870	1.8158	4.2036	4.4323	3.8461	1.7035	2.2447	2.9771	8.9949	0.2268	0.2761	0.1504	0.1684	0.7751	0.8127	0.9150	2.8796
asiesocia	0.2817	0.3955	0.6420	1.3895	0.5569	1.7889	1.7045	4.3878	0.2794	0.8294	1.7548	5.8668	0.1682	0.5349	0.0485	0.3268	0.0250	0.0793	0.3607	0.6214
endevelop	3.8131	5.9002	11.0779	19.3462	11.1607	20.1013	12.7208	16.0211	1.1571	3.6598	8.9073	13.1860	1.1545	0.8751	1.1832	0.8956	0.6014	2.2941	2.4488	3.2478

Sur ce dernier tableau, on est aussi confronté au problème de l'unité dans laquelle on mesure les quantités dénombrées. Les unités sont ici exprimées en millions de tonnes de port en lourd (MTPL). Si l'on change d'unité de comptage, on change certainement la valeur de la statistique du khi-deux. Mais ce qui est rassurant, c'est qu'on ne changera pas les interprétations descriptives des liens éventuels entre les variables.

IV.3.5. Le tableau de Burt

C'est le tableau qui sera utilisé comme un tableau de contingence. Utilisé par Burt dans un article de 1950 dans *British Journal of Psychology*, ce tableau a gardé le nom de celui qui l'a utilisé dans les premières fois. L'AFCM a été introduite plus tôt, en 1941, par Guttman, dont on a cité le nom à la fin du chapitre sur l'AFC (effet Guttman).

Ce tableau est un « super tableau de contingence », construit comme une juxtaposition de plusieurs tableaux de contingence. Les lignes et les colonnes correspondent chacune à une modalité de variable, et à l'intersection d'une ligne et d'une colonne se trouve l'effectif des individus qui prennent conjointement la modalité de la ligne et celle de la colonne. Exemple des données de l'enquête :

	BRAM	CARC	MC	NH	EARL	EI	GAEC	SCEA	A	P	mixte	coop	part	AO-VDQS	AOC	VDP
BRAM	14	0	0	0	5	5	4	0	14	0	1	12	1	5	3	6
CARC	0	32	0	0	6	22	2	2	22	10	3	23	6	5	10	17
MC	0	0	63	0	8	38	9	8	45	18	5	48	10	0	53	10
NH	0	0	0	45	3	26	9	7	28	17	0	33	12	0	10	35
EARL	5	6	8	3	22	0	0	0	17	5	5	12	5	3	12	7
EI	5	22	38	26	0	91	0	0	61	30	4	77	10	7	44	40
GAEC	4	2	9	9	0	0	24	0	19	5	0	20	4	0	8	16
SCEA	0	2	8	7	0	0	0	17	12	5	0	7	10	0	12	5
A	14	22	45	28	17	61	19	12	109	0	7	88	14	10	52	47
P	0	10	18	17	5	30	5	5	0	45	2	28	15	0	24	21
mixte	1	3	5	0	5	4	0	0	7	2	9	0	0	1	5	3
coop	12	23	48	33	12	77	20	7	88	28	0	116	0	8	53	55
part	1	6	10	12	5	10	4	10	14	15	0	0	29	1	18	10
AO-VDQS	5	5	0	0	3	7	0	0	10	0	1	8	1	10	0	0
AOC	3	10	53	10	12	44	8	12	52	24	5	53	18	0	76	0
VDP	6	17	10	35	7	40	16	5	47	21	3	55	10	0	0	68

Exercice : que valent les sommes des lignes ? des colonnes ?

IV.4. Comparaison des résultats de l'AFC et de l'AFCM lorsque p=2, et conséquences dans les différences d'interprétation

L'AFCM est une AFC simple opérée soit sur le tableau disjonctif complet, soit sur le tableau de Burt.

Dans le premier cas, les lignes sont les individus de départ et les colonnes sont les modalités. L'AFC positionne donc ces deux informations. Dans le second cas, les lignes et les colonnes sont égales aux modalités des variables. On n'a donc plus d'informations sont les comportements individuels, mais seulement sur les liens entre modalités.

IV.4.1. Les valeurs propres

Considérant l'exemple traité dans le chapitre AFC, la série des valeurs propres issues de l'AFC et de celles issues de l'AFCM sont les suivantes :

rappel tableau de contingence :

CSP\HEB	CAMP	HOTEL	LOCA	RESI	Total
AGRI	239	155	129	0	523
CADR	1003	1556	1821	1521	5901
INAC	682	1944	967	1333	4926
OUVR	2594	1124	2176	1038	6932
Total	4518	4779	5093	3892	18282

valeurs propres AFC (r=c=4➔ 3 axes)

λ_1=0,098243 (86,855%)

λ_2=0,013863 (12,256%)

λ_3=0,0010054 (0,889%)

$\Sigma_{i=1}^{3} \lambda_i$=0,113114

valeurs propres AFCM (r+c-2=6 axes)

μ_1= (1+ λ_1)/2=0,6567 (21,89%)

μ_2= (1+ λ_2)/2=0,5589 (18,63%)

μ_3= (1+ λ_3)/2=0,5159 (17,20%)

μ_i= (1- λ_{7-i})/2, sans intérêt

$\Sigma_{i=1}^{6} \mu_i$=(r+c)/2-1=3.

De façon plus générale, les liens entre les valeurs propres issues de l'AFC et celles issues de l'AFCM sont, lorsque p=2 :
pour i=1,…,min(r,c)-1, μ_i= (1+λ_i)/2,
pour i=min(r,c), max(r,c)-1, μ_i= 1/2,
pour i=max(r,c), …, r+c-2, μ_i= (1+$\lambda_{r+c-1-i}$)/2.

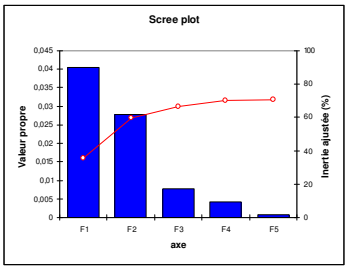
Pour faciliter la généralisation de 2 à p variables, nous remplacerons r et c par m_1 et m_2, de sorte que les nombres de modalités des variables X_1, X_2, …, X_p seront m_1, m_2, …, m_p.
Les valeurs propres sont au nombre de m_1+…+m_p-p, et leur somme égale (m_1+…+m_p-p)/p.

On voit avec ces propriétés que l'utilisation des valeurs propres ne pourra plus se faire de la même façon en AFCM qu'en AFC.
Nous pouvons mettre en comparaison les éléments d'interprétation à partir des valeurs propres de l'AFC et de l'AFCM.

analyse	AFC	AFCM
somme des valeurs propres	égale $n\chi^2$, donc est utilisée pour évaluer l'importance du lien entre les deux variables.	égale $(m_1+\ldots+m_p-p)/p$. Est donc uniquement dépendante du nombre de variables et du nombre total de modalités. Ne peut servir à évaluer l'importance du lien entre les variables.
nombre de valeurs propres=nombre d'axes	avec min(r,c)-1 axes, on restitue 100% de l'inertie, c'est-à-dire du lien entre les 2 variables	pour p=2, on a vu que le nombre d'axes pouvait être le double d'avec la méthode AFC, sur les mêmes données. Bien sûr, les seules 3 premières apportent 100% de l'information, les autres sont des "redondances".
pourcentage d'inertie apporté par les axes	il égale le pourcentage du lien apporté par l'axe, donc le pourcentage de l'information à laquelle on s'intéresse	n'égale plus le pourcentage d'information; on ne peut s'en servir pour savoir quel pourcentage du lien on a expliqué.
choix du nombre d'axes à retenir	méthode de Kaiser adaptée, ou méthode de l'éboulis	la méthode de Kaiser ne peut plus convenir, puisque la moyenne des valeurs propres ne dépend pas de l'importance de l'information totale; la méthode de l'éboulis seule convient, à condition de ne faire le choix qu'avec la première moitié, voire les [k/p] premiers axes, où k est le nombre total d'axes ([x] dénote la partie entière de x).

Exemple : les valeurs propres obtenues avec les données de viticulture sont les suivantes :

Valeurs propres et pourcentages d'inertie :											
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
Valeur propre	0,361	0,333	0,271	0,252	0,220	0,178	0,166	0,134	0,111	0,099	0,075
Inertie (%)	16,421	15,157	12,296	11,446	9,981	8,072	7,560	6,091	5,049	4,518	3,410
% cumulé	16,421	31,577	43,873	55,320	65,300	73,372	80,932	87,023	92,072	96,590	100,000
Inertie ajustée	0,041	0,028	0,008	0,004	0,001						
Inertie ajustée (%)	35,526	24,330	6,793	3,669	0,523						
% cumulé	35,526	59,855	66,649	70,318	70,841						



A l'aide des remarques faites ci-dessus, faire un choix du nombre d'axes à retenir.

IV.5. Les résultats pour les modalités

Ils se déclinent comme pour toute analyse factorielle, en coordonnées, contributions, et cosinus carrés. Une des différences avec l'AFC est que les modalités de TOUTES les variables figurent dans un même tableau, alors qu'avec 2 variables, l'AFC nous donnait une série de tableaux pour les modalités lignes, et une autre série pour les modalités colonnes.
Signalons des propriétés que l'on retrouve aussi pour toutes les analyses factorielles vues, à savoir que les coordonnées des modalités sont centrées, les moyennes étant à calculer en pondérant par les effectifs, que l'on retrouve dans la colonne des Poids). Chaque ensemble des coordonnées d'une même variable est centré (à vérifier).

Coordonnées principales (Variables) :						Contributions (Variables) :								Cosinus carrés (Variables) :					
	F1	F2	F3	F4	F5		Poids	Poids (relatif)	F1	F2	F3	F4	F5		F1	F2	F3	F4	F5
BRAM	1,993	-0,178	0,477	1,402	0,413	BRAM	14	0,018	0,200	0,002	0,015	0,142	0,014	BRAM	0,397	0,003	0,023	0,197	0,017
CARC	0,502	-0,095	0,556	-1,165	0,326	CARC	32	0,042	0,029	0,001	0,048	0,224	0,020	CARC	0,066	0,002	0,081	0,356	0,028
MC	-0,315	0,819	-0,603	0,192	-0,111	MC	63	0,082	0,022	0,165	0,110	0,012	0,005	MC	0,069	0,465	0,252	0,026	0,009
NH	-0,536	-1,024	0,300	0,123	-0,204	NH	45	0,058	0,046	0,184	0,019	0,004	0,011	NH	0,119	0,433	0,037	0,006	0,017
EARL	0,924	0,846	1,095	0,128	-1,061	EARL	22	0,029	0,067	0,061	0,127	0,002	0,146	EARL	0,142	0,119	0,200	0,003	0,188
EI	0,020	-0,072	-0,374	-0,546	0,267	EI	91	0,118	0,000	0,002	0,061	0,140	0,038	EI	0,001	0,007	0,202	0,430	0,103
GAEC	0,009	-0,836	-0,391	1,284	-0,725	GAEC	24	0,031	0,000	0,065	0,018	0,204	0,075	GAEC	0,000	0,129	0,028	0,304	0,097
SCEA	-1,316	0,469	1,136	0,943	0,967	SCEA	17	0,022	0,106	0,015	0,105	0,078	0,094	SCEA	0,215	0,027	0,160	0,110	0,116
A	0,307	0,025	-0,140	0,263	0,069	A	109	0,142	0,037	0,000	0,010	0,039	0,003	A	0,228	0,002	0,047	0,167	0,011
P	-0,744	-0,061	0,339	-0,636	-0,166	P	45	0,058	0,089	0,001	0,025	0,094	0,007	P	0,228	0,002	0,047	0,167	0,011
mixte	1,118	1,504	0,928	-0,847	-2,399	mixte	9	0,012	0,040	0,079	0,037	0,033	0,306	mixte	0,078	0,140	0,053	0,045	0,357
coop	0,161	-0,178	-0,412	-0,019	0,098	coop	116	0,151	0,011	0,014	0,094	0,000	0,007	coop	0,079	0,097	0,517	0,001	0,029
part	-0,991	0,246	1,359	0,341	0,352	part	29	0,038	0,102	0,007	0,257	0,017	0,021	part	0,228	0,014	0,428	0,027	0,029
AO-VDQS	2,465	0,186	0,973	-0,110	1,845	AO-VDQS	10	0,013	0,218	0,001	0,045	0,001	0,201	AO-VDQS	0,422	0,002	0,066	0,001	0,236
AOC	-0,333	0,789	-0,248	0,119	0,068	AOC	76	0,099	0,030	0,184	0,022	0,006	0,002	AOC	0,108	0,606	0,060	0,014	0,004
VDP	0,010	-0,909	0,134	-0,117	-0,347	VDP	68	0,088	0,000	0,219	0,006	0,005	0,048	VDP	0,000	0,653	0,014	0,011	0,095

Les contributions des modalités. Calculs. Ils se font comme pour ceux des modalités en AFC, en pondérant par les effectifs.

Exercice : retrouver certaines de ces contributions.

Utilisation. Une autre différence est que c'est la somme des contributions des modalités de toutes les variables qui égale 100% (ou 1, dans nos sorties), alors qu'en AFC, il fallait additionner les contributions des modalités d'une seule variable pour avoir 100%.
La moyenne des contributions de toutes les modalités est donc $1/(m_1+\ldots+m_p)$.
Dans l'exemple que nous suivons, les modalités contribuant plus que la moyenne sont relevées dans le tableau suivant, avec répartition dans le tableau selon les signes des coordonnées.

signes coordonnées	-	+
axe 1	SCEA, P, part	BRAM, EARL, AO-VDQS
axe 2	NH, GAEC, VDP	MC, mixte, AOC

Interprétation de ces contributions.
L'axe 1 est construit sur la base du fait que les viticulteurs de BRAM sont plutôt en EARL, produisent plus fréquemment que les autres de l'AO-VDQS, et sont moins fréquemment des SCEA, P, et part.
L'axe 2 est construit sur la base du fait que les viticulteurs de NH sont plutôt en GAEC, produisent plus fréquemment que les autres du VDP et moins de l'AOC, sont peu en valorisation mixte, contrairement à ceux du MC.

Les cosinus carrés des modalités. Calcul. De la même manière que dans les analyses précédentes, ils sont égaux au rapport entre le carré de la coordonnée sur l'axe et la somme des carrés de toutes les coordonnées, pour une modalité donnée. Pour les retrouver, il faut disposer des coordonnées sur tous les axes.

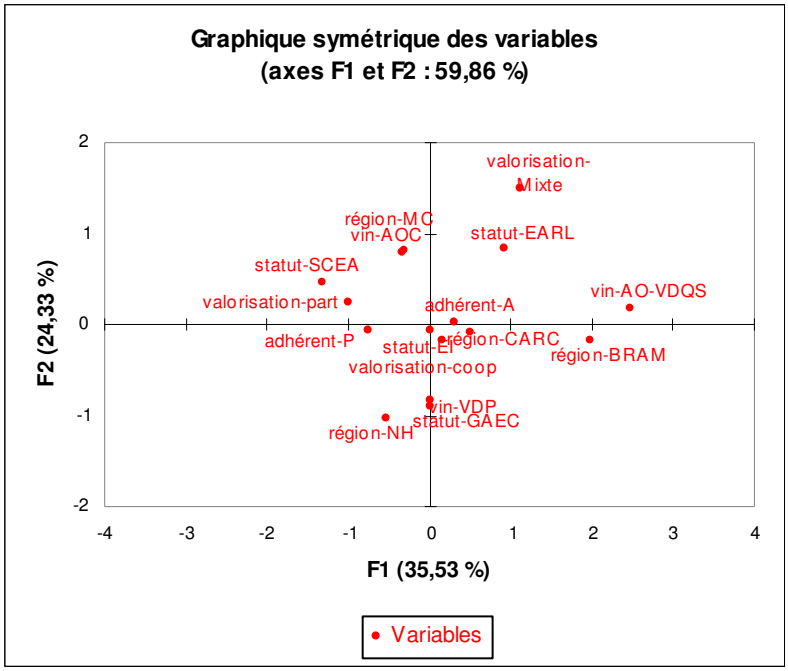
Exercice. Retrouver les cosinus carrés de certaines modalités sur certains axes (au choix).

Coordonnées principales (Variables) :						
	F6	F7	F8	F9	F10	F11
région-BRAM	1,120	0,441	0,236	-1,001	0,988	0,388
région-CARC	-0,521	-1,161	-0,218	-0,169	0,089	0,273
région-MC	0,073	-0,024	0,076	0,005	-0,055	-0,496
région-NH	-0,080	0,722	-0,024	0,425	-0,294	0,379
statut-EARL	0,204	0,248	-1,360	0,090	-0,320	-0,162
statut-EI	0,082	0,257	0,148	0,117	0,264	0,004
statut-GAEC	0,335	-1,360	0,475	0,253	-0,343	0,145
statut-SCEA	-1,175	0,223	0,296	-1,100	-0,513	-0,018
adhérent-A	-0,421	0,029	-0,059	0,171	0,120	0,000
adhérent-P	1,019	-0,069	0,144	-0,413	-0,290	-0,001
valorisation-Mixte	-0,849	0,545	2,037	-0,107	0,007	0,293
valorisation-coop	-0,011	0,032	-0,179	-0,163	-0,171	0,038
valorisation-part	0,306	-0,295	0,084	0,685	0,682	-0,243
vin-AO-VDQS	0,586	0,160	0,867	0,850	-1,349	-0,513
vin-AOC	0,102	-0,051	-0,097	0,089	0,016	0,428
vin-VDP	-0,200	0,033	-0,019	-0,225	0,181	-0,402

Utilisation et interprétation. Pour les cosinus carrés, la moyenne est, comme dans toutes les analyses factorielles vues, 1/(nb d'axes). C'est le calcul du nombre d'axes qui change d'une analyse à l'autre.
Le tableau qui suit indique les modalités qui sont reconstituées plus que la moyenne, et les signes des coordonnées de ces modalités.

signes coordonnées	-	+
axe 1	NH, SCEA, P, part, AOC	BRAM, EARL, A, AO-VDQS
axe 2	NH, GAEC, coop, VDP	MC, EARL, mixte, AOC

On peut remarquer que, pour les 2 premiers axes, toutes les modalités qui contribuent sont aussi bien reconstituées.
Sur l'axe 1, on peut ajouter par rapport à ce qui a été dit avec les contributions que les viticulteurs du NH ont un profil voisin de ceux en SCEA, et produisent plutôt de l'AOC. A l'opposé, les viticulteurs de BRAM sont plutôt Adhérents.
Sur l'axe 2, les viticulteurs du NH en GAEC sont aussi souvent en coop, ils s'opposent en cela à ceux du MC qui sont plutôt en EARL.
On remarque que les modalités ne sont pas toutes bien reconstituées. Il manque l'interprétation des modalités CARCASSONNE et EI. Cela signifie que les viticulteurs de Carcassonne, ainsi que ceux qui sont EI (entreprise individuelle, les plus nombreux), ont les profils les plus près de la moyenne.
Le nuage des modalités ci-contre est relativement "rond", on ne voit pas d'effet Guttman, ni de modalité ayant de caractéristique vraiment atypique.



IV.6. Les résultats pour les individus, lorsqu'ils sont présents

Lorsque l'on ne connaît des données qu'un tableau de données groupées, les résultats précédents sont les seuls que nous aurons à analyser. D'autres résultats numériques et graphiques existent (cf sorties SPAD, XLSTAT, …), mais nous ne les analyserons pas systématiquement.

Lorsque l'on connaît les profils de chaque individu, on peut avoir, en plus des résultats des modalités, ceux de ces individus. C'est le cas dans l'exemple que nous suivons. Une utilité de l'analyse de ces individus est dans l'intérêt qu'a la position de ces individus par rapport à celle des modalités (comme dans l'analyse des positions comparées des modalités lignes et colonnes en AFC). D'autre part, même si l'analyse individuelle n'a pas d'intérêt propre, la connaissance de la position de ces individus permet de détecter des individus atypiques, et permet d'opérer une classification de ces individus. La classification, que nous verrons dans le prochain chapitre, regroupe les individus en classes qu'on cherche à avoir les plus "homogènes". Cette mise en classes a plusieurs desseins. Le premier est une simplification de la lecture des profils possibles d'individus. Les autres buts dépendent du contexte dans lequel l'étude a été

faite. Par exemple, en marketing, on peut isoler des "segments" de marché en choisissant les classes dont le profil correspond à des critères préalablement choisis.

Mais revenons aux résultats pour les individus, qui, dans l'exemple, sont dans le tableau qui suit.

Coordonnées principales (Observations) :												Contributions (Observations) :								Cosinus carrés (Observations) :							
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11		Poids	Poids (relatif)	F1	F2	F3	F4	F5		F1	F2	F3	F4	F5
1	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	1	1	0,006	0,000	0,004	0,011	0,000	0,001	1	0,004	0,294	0,597	0,000	0,035
2	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	2	1	0,006	0,000	0,004	0,011	0,000	0,001	2	0,004	0,294	0,597	0,000	0,035
3	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	3	1	0,006	0,000	0,004	0,011	0,000	0,001	3	0,004	0,294	0,597	0,000	0,035
5	-0,057	0,214	-0,689	0,733	-0,257	0,037	-0,674	0,118	0,213	-0,275	0,084	5	1	0,006	0,000	0,001	0,011	0,014	0,002	5	0,002	0,027	0,275	0,311	0,038
7	0,566	1,380	0,397	-0,058	-1,466	-0,423	0,366	0,326	0,149	-0,147	0,046	7	1	0,006	0,006	0,037	0,004	0,000	0,064	7	0,064	0,381	0,032	0,001	0,430
8	0,633	-0,108	0,474	-0,363	-0,391	-0,450	-0,402	-1,004	-0,178	-0,064	-0,185	8	1	0,006	0,007	0,000	0,005	0,003	0,005	8	0,170	0,005	0,095	0,056	0,065
9	0,333	-0,426	-0,091	-0,632	0,176	-0,508	-0,397	-0,179	-0,162	0,306	-0,063	9	1	0,006	0,002	0,004	0,000	0,010	0,001	9	0,085	0,139	0,006	0,306	0,024
10	0,329	-0,690	-0,097	0,098	-0,248	-0,388	-1,190	0,000	-0,080	-0,079	0,040	10	1	0,006	0,002	0,009	0,000	0,000	0,002	10	0,048	0,212	0,004	0,004	0,027
11	-0,226	0,350	0,343	0,056	0,652	-0,961	-0,455	-0,141	-0,703	-0,291	0,527	11	1	0,006	0,001	0,002	0,003	0,000	0,013	11	0,019	0,045	0,043	0,001	0,156
12	-0,017	-0,455	0,093	-0,990	0,076	0,175	-0,445	-0,068	-0,512	0,046	-0,064	12	1	0,006	0,000	0,004	0,000	0,025	0,000	12	0,000	0,122	0,005	0,575	0,003
15	0,057	-0,373	-0,543	0,639	-0,434	-0,106	-0,633	0,160	0,024	-0,170	-0,522	15	1	0,006	0,000	0,003	0,007	0,011	0,006	15	0,002	0,079	0,166	0,230	0,106
16	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	16	1	0,006	0,000	0,004	0,011	0,000	0,001	16	0,004	0,294	0,597	0,000	0,035
17	-0,013	-0,747	-0,189	-0,118	-0,050	-0,299	0,526	-0,073	0,194	0,063	0,014	17	1	0,006	0,000	0,011	0,001	0,000	0,000	17	0,000	0,545	0,035	0,014	0,002
18	-0,013	-0,747	-0,189	-0,118	-0,050	-0,299	0,526	-0,073	0,194	0,063	0,014	18	1	0,006	0,000	0,011	0,001	0,000	0,000	18	0,000	0,545	0,035	0,014	0,002
19	-0,013	-0,747	-0,189	-0,118	-0,050	-0,299	0,526	-0,073	0,194	0,063	0,014	19	1	0,006	0,000	0,011	0,001	0,000	0,000	19	0,000	0,545	0,035	0,014	0,002
21	-0,366	-1,042	-0,011	0,253	-0,574	0,504	-0,315	0,217	-0,074	-0,582	0,116	21	1	0,006	0,002	0,021	0,000	0,002	0,010	21	0,057	0,458	0,000	0,027	0,139
22	-0,498	0,666	-0,102	0,597	0,465	-0,679	0,102	0,020	-0,599	-0,382	-0,035	22	1	0,006	0,004	0,009	0,000	0,009	0,006	22	0,110	0,197	0,005	0,158	0,096
25	0,174	0,158	0,229	0,244	-0,440	-0,097	0,480	-0,940	0,367	-0,412	0,499	25	1	0,006	0,001	0,000	0,001	0,002	0,006	25	0,015	0,012	0,026	0,029	0,095
26	-0,954	0,175	0,926	0,713	0,534	-0,602	0,308	0,109	0,162	0,007	0,399	26	1	0,006	0,016	0,001	0,021	0,013	0,008	26	0,281	0,009	0,264	0,157	0,088
27	-0,013	-0,747	-0,189	-0,118	-0,050	-0,299	0,526	-0,073	0,194	0,063	0,014	27	1	0,006	0,000	0,011	0,001	0,000	0,000	27	0,000	0,545	0,035	0,014	0,002
29	-0,510	-0,013	0,345	0,120	0,235	-0,005	0,324	0,029	0,892	0,499	0,415	29	1	0,006	0,005	0,000	0,003	0,000	0,002	29	0,147	0,000	0,067	0,008	0,031
30	-0,746	-0,630	0,676	-0,333	-0,042	0,535	0,318	0,182	0,353	0,344	-0,192	30	1	0,006	0,010	0,008	0,011	0,003	0,000	30	0,250	0,179	0,205	0,050	0,001
31	-0,013	-0,747	-0,189	-0,118	-0,050	-0,299	0,526	-0,073	0,194	0,063	0,014	31	1	0,006	0,000	0,011	0,001	0,000	0,000	31	0,000	0,545	0,035	0,014	0,002
32	-0,362	-0,777	-0,005	-0,476	-0,150	0,384	0,478	0,038	-0,156	-0,197	0,014	32	1	0,006	0,002	0,012	0,000	0,006	0,001	32	0,092	0,423	0,000	0,159	0,016
33	-1,190	-0,443	1,256	0,261	0,257	-0,062	0,301	0,263	-0,377	-0,149	-0,208	33	1	0,006	0,025	0,004	0,038	0,002	0,002	33	0,383	0,053	0,427	0,018	0,018
34	-0,457	-0,560	0,392	0,475	0,249	-0,895	0,509	0,008	-0,536	-0,430	-0,002	34	1	0,006	0,004	0,006	0,004	0,006	0,002	34	0,084	0,126	0,061	0,091	0,025
35	-0,498	0,666	-0,102	0,597	0,465	-0,679	0,102	0,020	-0,599	-0,382	-0,035	35	1	0,006	0,004	0,009	0,000	0,009	0,006	35	0,110	0,197	0,005	0,158	0,096
36	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	36	1	0,006	0,000	0,004	0,011	0,000	0,001	36	0,004	0,294	0,597	0,000	0,035
37	-0,485	0,914	0,747	0,057	-0,392	0,809	-0,094	-0,630	0,274	0,021	-0,347	37	1	0,006	0,004	0,016	0,013	0,000	0,005	37	0,077	0,275	0,183	0,001	0,051
38	-0,016	-1,012	-0,195	0,611	-0,474	-0,179	-0,267	0,106	0,276	-0,322	0,117	38	1	0,006	0,000	0,020	0,001	0,010	0,007	38	0,000	0,520	0,019	0,190	0,114
39	-0,749	-0,895	0,669	0,396	-0,465	0,655	-0,475	0,361	0,435	-0,041	-0,089	39	1	0,006	0,010	0,016	0,011	0,004	0,006	39	0,177	0,253	0,141	0,050	0,068
40	-0,400	-0,309	0,774	-0,846	0,185	0,325	-0,605	0,076	-0,003	0,587	-0,269	40	1	0,006	0,003	0,002	0,014	0,018	0,001	40	0,064	0,038	0,240	0,287	0,014
41	-0,845	-0,121	1,355	-0,253	0,483	-0,271	-0,622	0,156	-0,733	0,094	-0,285	41	1	0,006	0,013	0,000	0,044	0,002	0,007	41	0,180	0,004	0,462	0,016	0,059
42	0,633	-0,108	0,474	-0,363	-0,391	-0,450	-0,402	-1,004	-0,178	-0,064	-0,185	42	1	0,006	0,007	0,000	0,005	0,003	0,005	42	0,170	0,005	0,095	0,056	0,065
43	-0,017	-0,455	0,093	-0,990	0,076	0,175	-0,445	-0,068	-0,512	0,046	-0,064	43	1	0,006	0,000	0,004	0,000	0,025	0,000	43	0,000	0,122	0,005	0,575	0,003
44	1,015	0,451	0,297	0,754	-0,177	0,472	0,342	-0,797	-0,488	0,401	0,505	44	1	0,006	0,019	0,004	0,002	0,015	0,001	44	0,290	0,057	0,025	0,160	0,009
45	-0,457	-0,560	0,392	0,475	0,249	-0,895	0,509	0,008	-0,536	-0,430	-0,002	45	1	0,006	0,004	0,006	0,004	0,006	0,002	45	0,084	0,126	0,061	0,091	0,025
46	-0,013	-0,747	-0,189	-0,118	-0,050	-0,299	0,526	-0,073	0,194	0,063	0,014	46	1	0,006	0,000	0,011	0,001	0,000	0,000	46	0,000	0,545	0,035	0,014	0,002
47	-0,362	-0,777	-0,005	-0,476	-0,150	0,384	0,478	0,038	-0,156	-0,197	0,014	47	1	0,006	0,002	0,012	0,000	0,006	0,001	47	0,092	0,423	0,000	0,159	0,016
48	-0,127	-0,159	-0,336	-0,024	0,127	-0,155	0,485	-0,116	0,383	-0,042	0,620	48	1	0,006	0,000	0,000	0,003	0,000	0,000	48	0,016	0,026	0,115	0,001	0,016
49	-1,304	0,145	1,110	0,355	0,434	0,082	0,260	0,220	-0,188	-0,253	0,398	49	1	0,006	0,031	0,000	0,030	0,003	0,006	49	0,466	0,006	0,338	0,034	0,052
50	-0,013	-0,747	-0,189	-0,118	-0,050	-0,299	0,526	-0,073	0,194	0,063	0,014	50	1	0,006	0,000	0,011	0,001	0,000	0,000	50	0,000	0,545	0,035	0,014	0,002
52	-0,013	-0,747	-0,189	-0,118	-0,050	-0,299	0,526	-0,073	0,194	0,063	0,014	52	1	0,006	0,000	0,011	0,001	0,000	0,000	52	0,000	0,545	0,035	0,014	0,002
53	0,061	-0,109	-0,536	-0,090	-0,010	-0,226	0,160	-0,019	-0,057	0,215	-0,625	53	1	0,006	0,000	0,000	0,007	0,000	0,000	53	0,004	0,014	0,347	0,010	0,000
54	0,247	0,797	-0,118	0,272	-0,400	-0,025	0,114	-0,885	0,115	-0,260	-0,141	54	1	0,006	0,001	0,012	0,000	0,002	0,005	54	0,033	0,345	0,008	0,040	0,087
56	0,057	-0,373	-0,543	0,639	-0,434	-0,106	-0,633	0,160	0,024	-0,170	-0,522	56	1	0,006	0,000	0,003	0,007	0,011	0,006	56	0,002	0,079	0,166	0,230	0,106
57	0,061	-0,109	-0,536	-0,090	-0,010	-0,226	0,160	-0,019	-0,057	0,215	-0,625	57	1	0,006	0,000	0,000	0,007	0,000	0,000	57	0,004	0,014	0,347	0,010	0,000
58	0,265	1,062	-0,168	-0,326	-0,899	-0,481	0,370	1,150	0,165	0,223	0,168	58	1	0,006	0,001	0,022	0,001	0,003	0,024	58	0,018	0,286	0,007	0,027	0,205
59	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	59	1	0,006	0,000	0,004	0,011	0,000	0,001	59	0,004	0,294	0,597	0,000	0,035
60	0,265	1,062	-0,168	-0,326	-0,899	-0,481	0,370	1,150	0,165	0,223	0,168	60	1	0,006	0,001	0,022	0,001	0,003	0,024	60	0,018	0,286	0,007	0,027	0,205
61	-0,022	0,356	0,709	0,321	-0,469	-0,018	-0,004	-0,699	0,436	0,386	-0,952	61	1	0,006	0,000	0,002	0,012	0,003	0,006	61	0,000	0,047	0,187	0,038	0,082
62	-0,498	0,666	-0,102	0,597	0,465	-0,679	0,102	0,020	-0,599	-0,382	-0,035	62	1	0,006	0,004	0,009	0,000	0,009	0,006	62	0,110	0,197	0,005	0,158	0,096
63	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	63	1	0,006	0,000	0,004	0,011	0,000	0,001	63	0,004	0,294	0,597	0,000	0,035
64	-0,881	0,813	0,578	0,741	0,574	-0,529	-0,058	0,164	-0,090	0,159	-0,240	64	1	0,006	0,014	0,013	0,008	0,014	0,010	64	0,254	0,217	0,110	0,180	0,108
65	-0,881	0,813	0,578	0,741	0,574	-0,529	-0,058	0,164	-0,090	0,159	-0,240	65	1	0,006	0,014	0,013	0,008	0,014	0,010	65	0,254	0,217	0,110	0,180	0,108
66	-0,403	0,449	-0,499	-0,355	0,066	0,601	0,071	0,050	-0,219	-0,150	-0,020	66	1	0,006	0,003	0,004	0,006	0,003	0,000	66	0,137	0,171	0,211	0,106	0,004
67	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	67	1	0,006	0,000	0,004	0,011	0,000	0,001	67	0,004	0,294	0,597	0,000	0,035
68	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	68	1	0,006	0,000	0,004	0,011	0,000	0,001	68	0,004	0,294	0,597	0,000	0,035
69	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	69	1	0,006	0,000	0,004	0,011	0,000	0,001	69	0,004	0,294	0,597	0,000	0,035
70	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	70	1	0,006	0,000	0,004	0,011	0,000	0,001	70	0,004	0,294	0,597	0,000	0,035
71	-0,403	0,449	-0,499	-0,355	0,066	0,601	0,071	0,050	-0,219	-0,150	-0,020	71	1	0,006	0,003	0,004	0,006	0,003	0,000	71	0,137	0,171	0,211	0,106	0,004
72	0,012	0,179	0,212	-0,180	-0,678	0,515	0,108	-0,732	-0,424	-0,415	-0,747	72	1	0,006	0,000	0,001	0,001	0,001	0,014	72	0,000	0,014	0,020	0,014	0,200
74	0,057	-0,373	-0,543	0,639	-0,434	-0,106	-0,633	0,160	0,024	-0,170	-0,522	74	1	0,006	0,000	0,003	0,007	0,011	0,006	74	0,002	0,079	0,166	0,230	0,106
75	-0,403	0,449	-0,499	-0,355	0,066	0,601	0,071	0,050	-0,219	-0,150	-0,020	75	1	0,006	0,003	0,004	0,006	0,003	0,000	75	0,137	0,171	0,211	0,106	0,004
76	0,216	1,350	0,581	-0,416	-1,566	0,260	0,318	0,437	-0,201	-0,407	0,045	76	1	0,006	0,001	0,035	0,008	0,004	0,073	76	0,009	0,337	0,062	0,032	0,454
77	-0,403	0,449	-0,499	-0,355	0,066	0,601	0,071	0,050	-0,219	-0,150	-0,020	77	1	0,006	0,003	0,004	0,006	0,003	0,000	77	0,137	0,171	0,211	0,106	0,004
78	0,379	0,474	-0,021	-0,420	-1,076	-0,624	0,412	1,193	-0,024	0,328	-0,439	78	1	0,006	0,003	0,004	0,000	0,005	0,034	78	0,036	0,056	0,000	0,044	0,291
79	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	79	1	0,006	0,000	0,004	0,011	0,000	0,001	79	0,004	0,294	0,597	0,000	0,035
84	-0,057	0,214	-0,689	0,733	-0,257	0,037	-0,674	0,118	0,213	-0,275	0,084	84	1	0,006	0,000	0,001	0,011	0,014	0,002	84	0,002	0,027	0,275	0,311	0,038
85	-0,403	0,449	-0,499	-0,355	0,066	0,601	0,071	0,050	-0,219	-0,150	-0,020	85	1	0,006	0,003	0,004	0,006	0,003	0,000	85	0,137	0,171	0,211	0,106	0,004
86	1,946	0,243	0,767	0,663	0,582	0,702	0,446	-0,271	-0,032	-0,465	-0,182	86	1	0,006	0,068	0,001	0,014	0,011	0,010	86	0,608	0,009	0,094	0,071	0,054
87	-0,881	0,813	0,578	0,741	0,574	-0,529	-0,058	0,164	-0,090	0,159	-0,240	87	1	0,006	0,014	0,013	0,008	0,014	0,010	87	0,254	0,217	0,110	0,180	0,108
88	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	88	1	0,006	0,000	0,004	0,011	0,000	0,001	88	0,004	0,294	0,597	0,000	0,035
89	-0,881	0,813	0,578	0,741	0,574	-0,529	-0,058	0,164	-0,090	0,159	-0,240	89	1	0,006	0,014	0,013	0,008	0,014	0,010	89	0,254	0,217	0,110	0,180	0,108
90	-0,057	0,214	-0,689	0,733	-0,257	0,037	-0,674	0,118	0,213	-0,275	0,084	90	1	0,006	0,000	0,001	0,011	0,014	0,002	90	0,002	0,027	0,275	0,311	0,038
91	-0,786	0,596	0,182	-0,211	0,175	0,751	-0,089	0,194	0,290	0,391	-0,225	91	1	0,006	0,011	0,007	0,001	0,001	0,001	91	0,312	0,180	0,017	0,023	0,015
94	-0,403	0,449	-0,499	-0,355	0,066	0,601	0,071	0,050	-0,219	-0,150	-0,020	94	1	0,006	0,003	0,004	0,006	0,003	0,000	94	0,137	0,171	0,211	0,106	0,004
95	-0,403	0,449	-0,499	-0,355	0,066	0,601	0,071	0,050	-0,219	-0,150	-0,020	95	1	0,006	0,003	0,004	0,006	0,003	0,000	95	0,137	0,171	0,211	0,106	0,004
97	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	97	1	0,006	0,000	0,004	0,011	0,000	0,001	97	0,004	0,294	0,597	0,000	0,035
98	-0,403	0,449	-0,499	-0,355	0,066	0,601	0,071	0,050	-0,219	-0,150	-0,020	98	1	0,006	0,003	0,004	0,006	0,003	0,000	98	0,137	0,171	0,211	0,106	0,004
99	-0,403	0,449	-0,499	-0,355	0,066	0,601	0,071	0,050	-0,219	-0,150	-0,020	99	1	0,006	0,003	0,004	0,006	0,003	0,000	99	0,137	0,171	0,211	0,106	0,004
100	-0,847	0,637	0,082	0,239	0,365	0,004	0,054	0,131	-0,949	-0,642	-0,036	100	1	0,006	0,013	0,008	0,000	0,001	0,004	100	0,270	0,153	0,003	0,021	0,050
101	0,061	-0,109	-0,536	-0,090	-0,010	-0,226	0,160	-0,019	-0,057	0,215	-0,625	101	1	0,006	0,000	0,000	0,007	0,000	0,000	101	0,004	0,014	0,347	0,010	0,000
102	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	102	1	0,006	0,000	0,004	0,011	0,000	0,001	102	0,004	0,294	0,597	0,000	0,035
103	-0,485	0,914	0,747	0,057	-0,392	0,809	-0,094	-0,630	0,274	0,021	-0,347	103	1	0,006	0,004	0,016	0,013	0,000	0,005	103	0,077	0,275	0,183	0,001	0,051
104	-0,057	0,214	-0,689	0,733	-0,257	0,037	-0,674	0,118	0,213	-0,275	0,084	104	1	0,006	0,000	0,001	0,011	0,014	0,002	104	0,002	0,027	0,275	0,311	0,038
105	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	105	1	0,006	0,000	0,004	0,011	0,000	0,001	105	0,004	0,294	0,597	0,000	0,035
106	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	106	1	0,006	0,000	0,004	0,011	0,000	0,001	106	0,004	0,294	0,597	0,000	0,035
107	-0,013	-0,747	-0,189	-0,118	-0,050	-0,299	0,526	-0,073	0,194	0,063	0,014	107	1	0,006	0,000	0,011	0,001	0,000	0,000	107	0,000	0,545	0,035	0,014	0,002
108	0,174	0,158	0,229	0,244	-0,440	-0,097	0,480	-0,940	0,367	-0,412	0,499	108	1	0,006	0,001	0,000	0,001	0,002	0,006	108	0,015	0,012	0,026	0,029	0,095
109	-0,053	0,479	-0,683	0,004	0,166	-0,083	0,119	-0,061	0,131	0,110	-0,019	109	1	0,006	0,000	0,004	0,011	0,000	0,001	109	0,004	0,294	0,597	0,000	0,035
110	-0,440	0,361	-0,009	0,876	-0,149	0,188	-0,834	0,262	0,722	0,267	-0,121	110	1	0,006	0,003	0,003	0,000	0,020	0,001	110	0,077	0,052	0,000	0,304	0,009
112	0,061	-0,109	-0,536	-0,090	-0,010	-0,226	0,160	-0,019	-0,057	0,215	-0,625	112	1	0,006	0,000	0,000	0,007	0,000	0,000	112	0,004	0,014	0,347	0,010	0,000
113	-0,013	-0,747	-0,189	-0,118	-0,050	-0,299	0,526	-0,073	0,194	0,063	0,014	113	1	0,006	0,000	0,011	0,001	0,000	0,000	113	0,000	0,545	0,035	0,014	0,002
114	-0,013	-0,747	-0,189	-0,118	-0,050	-0,299	0,526	-0,073	0,194	0,063	0,014	114	1	0,006	0,000	0,011	0,001	0,000	0,000	114	0,000	0,545	0,035	0,014	0,002
115	-0,013	-0,747	-0,189	-0,118	-0,050	-0,299	0,526	-0,073	0,194	0,063	0,014	115	1	0,006	0,000	0,011	0,001	0,000	0,000	115	0,000	0,545	0,035	0,014	0,002
116	-0,362	-0,777	-0,005	-0,476	-0,150	0,384	0,478	0,038	-0,156	-0,197	0,014	116	1	0,006	0,002	0,012	0,000	0,006	0,001	116	0,092	0,423	0,000	0,159	0,016

117	-0,016	-1,012	-0,195	0,611	-0,474	-0,179	-0,267	0,106	0,276	-0,322	0,117	117	1	0,006	0,000	0,020	0,001	0,010	0,007	117	0,000	0,520	0,019	0,190	0,114
118	-1,304	0,145	1,110	0,355	0,434	0,082	0,260	0,220	-0,188	-0,253	0,398	118	1	0,006	0,031	0,000	0,030	0,003	0,006	118	0,466	0,006	0,338	0,034	0,052
119	-0,362	-0,777	-0,005	-0,476	-0,150	0,384	0,478	0,038	-0,156	-0,197	0,014	119	1	0,006	0,002	0,012	0,000	0,006	0,001	119	0,092	0,423	0,000	0,159	0,016
120	-0,013	-0,747	-0,189	-0,118	-0,050	-0,299	0,526	-0,073	0,194	0,063	0,014	120	1	0,006	0,000	0,011	0,001	0,000	0,000	120	0,000	0,545	0,035	0,014	0,002
124	-0,749	-0,895	0,669	0,396	-0,465	0,655	-0,475	0,361	0,435	-0,041	-0,089	124	1	0,006	0,010	0,016	0,011	0,004	0,006	124	0,177	0,253	0,141	0,050	0,068
125	-0,013	-0,747	-0,189	-0,118	-0,050	-0,299	0,526	-0,073	0,194	0,063	0,014	125	1	0,006	0,000	0,011	0,001	0,000	0,000	125	0,000	0,545	0,035	0,014	0,002
126	-0,013	-0,747	-0,189	-0,118	-0,050	-0,299	0,526	-0,073	0,194	0,063	0,014	126	1	0,006	0,000	0,011	0,001	0,000	0,000	126	0,000	0,545	0,035	0,014	0,002
127	-0,746	-0,630	0,676	-0,333	-0,042	0,535	0,318	0,182	0,353	0,344	-0,192	127	1	0,006	0,010	0,008	0,011	0,003	0,000	127	0,250	0,179	0,205	0,050	0,001
128	-0,362	-0,777	-0,005	-0,476	-0,150	0,384	0,478	0,038	-0,156	-0,197	0,014	128	1	0,006	0,002	0,012	0,000	0,006	0,001	128	0,092	0,423	0,000	0,159	0,016
129	-0,840	-0,413	1,072	0,619	0,357	-0,745	0,349	0,152	-0,027	0,111	-0,207	129	1	0,006	0,013	0,003	0,028	0,010	0,004	129	0,214	0,052	0,349	0,116	0,039
131	0,333	-0,426	-0,091	-0,632	0,176	-0,508	-0,397	-0,179	-0,162	0,306	-0,063	131	1	0,006	0,002	0,004	0,000	0,010	0,001	131	0,085	0,139	0,006	0,306	0,024
135	0,633	-0,108	0,474	-0,363	-0,391	-0,450	-0,402	-1,004	-0,178	-0,064	-0,185	135	1	0,006	0,007	0,000	0,005	0,003	0,005	135	0,170	0,005	0,095	0,056	0,065
136	1,150	-0,046	0,232	-0,629	1,112	-0,135	-0,335	0,305	0,483	-0,665	-0,144	136	1	0,006	0,024	0,000	0,001	0,010	0,037	136	0,336	0,001	0,014	0,101	0,315
137	1,150	-0,046	0,232	-0,629	1,112	-0,135	-0,335	0,305	0,483	-0,665	-0,144	137	1	0,006	0,024	0,000	0,001	0,010	0,037	137	0,336	0,001	0,014	0,101	0,315
138	1,150	-0,046	0,232	-0,629	1,112	-0,135	-0,335	0,305	0,483	-0,665	-0,144	138	1	0,006	0,024	0,000	0,001	0,010	0,037	138	0,336	0,001	0,014	0,101	0,315
139	0,829	-0,454	-0,121	0,392	0,213	0,271	0,388	0,069	-0,661	0,876	0,021	139	1	0,006	0,012	0,004	0,000	0,004	0,001	139	0,270	0,081	0,006	0,060	0,018
140	1,150	-0,046	0,232	-0,629	1,112	-0,135	-0,335	0,305	0,483	-0,665	-0,144	140	1	0,006	0,024	0,000	0,001	0,010	0,037	140	0,336	0,001	0,014	0,101	0,315
142	0,838	1,063	0,843	-0,599	-1,279	-0,705	-0,191	0,165	0,045	-0,056	0,607	142	1	0,006	0,013	0,022	0,017	0,009	0,048	142	0,128	0,206	0,130	0,066	0,299
143	0,219	0,162	-0,237	-0,537	0,353	-0,365	-0,439	-0,222	0,027	0,201	0,543	143	1	0,006	0,001	0,001	0,001	0,007	0,004	143	0,038	0,021	0,045	0,230	0,099
144	-0,165	0,309	0,444	-0,394	0,462	-0,214	-0,599	-0,077	0,536	0,742	0,337	144	1	0,006	0,000	0,002	0,005	0,004	0,006	144	0,013	0,047	0,096	0,076	0,104
146	0,333	-0,426	-0,091	-0,632	0,176	-0,508	-0,397	-0,179	-0,162	0,306	-0,063	146	1	0,006	0,002	0,004	0,000	0,010	0,001	146	0,085	0,139	0,006	0,306	0,024
147	0,219	0,162	-0,237	-0,537	0,353	-0,365	-0,439	-0,222	0,027	0,201	0,543	147	1	0,006	0,001	0,001	0,001	0,007	0,004	147	0,038	0,021	0,045	0,230	0,099
148	0,329	-0,690	-0,097	0,098	-0,248	-0,388	-1,190	0,000	-0,080	-0,079	0,040	148	1	0,006	0,002	0,009	0,000	0,000	0,002	148	0,048	0,212	0,004	0,004	0,027
149	-0,131	0,133	-0,053	-0,896	0,253	0,319	-0,487	-0,111	-0,323	-0,059	0,542	149	1	0,006	0,000	0,000	0,000	0,021	0,002	149	0,010	0,011	0,002	0,485	0,039
150	-0,017	-0,455	0,093	-0,990	0,076	0,175	-0,445	-0,068	-0,512	0,046	-0,064	150	1	0,006	0,000	0,004	0,000	0,025	0,000	150	0,000	0,122	0,005	0,575	0,003
151	-0,131	0,133	-0,053	-0,896	0,253	0,319	-0,487	-0,111	-0,323	-0,059	0,542	151	1	0,006	0,000	0,000	0,000	0,021	0,002	151	0,010	0,011	0,002	0,485	0,039
152	-0,131	0,133	-0,053	-0,896	0,253	0,319	-0,487	-0,111	-0,323	-0,059	0,542	152	1	0,006	0,000	0,000	0,000	0,021	0,002	152	0,010	0,011	0,002	0,485	0,039
153	0,219	0,162	-0,237	-0,537	0,353	-0,365	-0,439	-0,222	0,027	0,201	0,543	153	1	0,006	0,001	0,001	0,001	0,007	0,004	153	0,038	0,021	0,045	0,230	0,099
154	0,333	-0,426	-0,091	-0,632	0,176	-0,508	-0,397	-0,179	-0,162	0,306	-0,063	154	1	0,006	0,002	0,004	0,000	0,010	0,001	154	0,085	0,139	0,006	0,306	0,024
156	0,766	0,100	0,913	-0,485	1,220	0,015	-0,495	0,449	0,993	-0,124	-0,350	156	1	0,006	0,011	0,000	0,020	0,006	0,044	156	0,124	0,002	0,176	0,050	0,315
157	0,952	0,475	0,989	-0,693	-1,456	-0,849	-0,150	0,208	-0,144	0,049	0,001	157	1	0,006	0,016	0,004	0,023	0,012	0,063	157	0,164	0,041	0,177	0,087	0,384
159	0,711	-0,131	-0,274	1,215	-0,033	0,534	-0,446	0,206	-0,390	0,386	0,730	159	1	0,006	0,009	0,000	0,002	0,038	0,000	159	0,147	0,005	0,022	0,430	0,000
160	1,946	0,243	0,767	0,663	0,582	0,702	0,446	-0,271	-0,032	-0,465	-0,182	160	1	0,006	0,068	0,001	0,014	0,011	0,010	160	0,608	0,009	0,094	0,071	0,054
161	1,646	-0,075	0,202	0,394	1,149	0,644	0,450	0,553	-0,016	-0,095	-0,060	161	1	0,006	0,049	0,000	0,001	0,004	0,039	161	0,524	0,001	0,008	0,030	0,256
162	1,964	0,508	0,717	0,065	0,083	0,246	0,702	1,764	0,018	0,018	0,126	162	1	0,006	0,069	0,005	0,012	0,000	0,000	162	0,464	0,031	0,062	0,001	0,001
163	0,825	-0,719	-0,127	1,121	-0,210	0,390	-0,405	0,248	-0,579	0,491	0,124	163	1	0,006	0,012	0,010	0,000	0,032	0,001	163	0,195	0,148	0,005	0,361	0,013
164	0,632	0,598	0,978	0,898	-0,068	0,622	0,182	-0,653	0,021	0,942	0,300	164	1	0,006	0,007	0,007	0,023	0,021	0,000	164	0,092	0,082	0,220	0,185	0,001
165	-0,476	-0,189	-0,152	-0,382	0,027	0,528	0,437	-0,005	0,033	-0,302	0,620	165	1	0,006	0,004	0,001	0,001	0,004	0,000	165	0,165	0,026	0,017	0,106	0,001
166	-0,016	-1,012	-0,195	0,611	-0,474	-0,179	-0,267	0,106	0,276	-0,322	0,117	166	1	0,006	0,000	0,020	0,001	0,010	0,007	166	0,000	0,520	0,019	0,190	0,114
167	-0,400	-0,309	0,774	-0,846	0,185	0,325	-0,605	0,076	-0,003	0,587	-0,269	167	1	0,006	0,003	0,002	0,014	0,018	0,001	167	0,064	0,038	0,240	0,287	0,014
168	-0,016	-1,012	-0,195	0,611	-0,474	-0,179	-0,267	0,106	0,276	-0,322	0,117	168	1	0,006	0,000	0,020	0,001	0,010	0,007	168	0,000	0,520	0,019	0,190	0,114
169	-0,863	-0,307	0,523	0,490	-0,289	0,798	-0,517	0,318	0,624	-0,145	0,517	169	1	0,006	0,013	0,002	0,007	0,006	0,002	169	0,239	0,030	0,088	0,077	0,027
170	-0,362	-0,777	-0,005	-0,476	-0,150	0,384	0,478	0,038	-0,156	-0,197	0,014	170	1	0,006	0,002	0,012	0,000	0,006	0,001	170	0,092	0,423	0,000	0,159	0,016
171	-0,366	-1,042	-0,011	0,253	-0,574	0,504	-0,315	0,217	-0,074	-0,582	0,116	171	1	0,006	0,002	0,021	0,000	0,002	0,010	171	0,057	0,458	0,000	0,027	0,139
172	-0,209	0,305	0,910	0,388	-0,332	0,053	0,320	-0,796	0,876	0,129	0,293	172	1	0,006	0,001	0,002	0,020	0,004	0,003	172	0,015	0,033	0,292	0,053	0,039
173	-0,165	0,309	0,444	-0,394	0,462	-0,214	-0,599	-0,077	0,536	0,742	0,337	173	1	0,006	0,000	0,002	0,005	0,004	0,006	173	0,013	0,047	0,096	0,076	0,104
174	-0,786	0,596	0,182	-0,211	0,175	0,751	-0,089	0,194	0,290	0,391	-0,225	174	1	0,006	0,011	0,007	0,001	0,001	0,001	174	0,312	0,180	0,017	0,023	0,015
175	0,829	-0,454	-0,121	0,392	0,213	0,271	0,388	0,069	-0,661	0,876	0,021	175	1	0,006	0,012	0,004	0,000	0,004	0,001	175	0,270	0,081	0,006	0,060	0,018
176	0,825	-0,719	-0,127	1,121	-0,210	0,390	-0,405	0,248	-0,579	0,491	0,124	176	1	0,006	0,012	0,010	0,000	0,032	0,001	176	0,195	0,148	0,005	0,361	0,013
177	0,829	-0,454	-0,121	0,392	0,213	0,271	0,388	0,069	-0,661	0,876	0,021	177	1	0,006	0,012	0,004	0,000	0,004	0,001	177	0,270	0,081	0,006	0,060	0,018
178	1,946	0,243	0,767	0,663	0,582	0,702	0,446	-0,271	-0,032	-0,465	-0,182	178	1	0,006	0,068	0,001	0,014	0,011	0,010	178	0,608	0,009	0,094	0,071	0,054
179	0,825	-0,719	-0,127	1,121	-0,210	0,390	-0,405	0,248	-0,579	0,491	0,124	179	1	0,006	0,012	0,010	0,000	0,032	0,001	179	0,195	0,148	0,005	0,361	0,013
180	-0,403	0,449	-0,499	-0,355	0,066	0,601	0,071	0,050	-0,219	-0,150	-0,020	180	1	0,006	0,003	0,004	0,006	0,003	0,000	180	0,137	0,171	0,211	0,106	0,004
182	-0,403	0,449	-0,499	-0,355	0,066	0,601	0,071	0,050	-0,219	-0,150	-0,020	182	1	0,006	0,003	0,004	0,006	0,003	0,000	182	0,137	0,171	0,211	0,106	0,004
183	-0,057	0,214	-0,689	0,733	-0,257	0,037	-0,674	0,118	0,213	-0,275	0,084	183	1	0,006	0,000	0,001	0,011	0,014	0,002	183	0,002	0,027	0,275	0,311	0,038
184	0,247	0,797	-0,118	0,272	-0,400	-0,025	0,114	-0,885	0,115	-0,260	-0,141	184	1	0,006	0,001	0,012	0,000	0,002	0,005	184	0,033	0,345	0,008	0,040	0,087
185	0,602	0,445	1,173	-1,051	-1,556	-0,165	-0,198	0,319	-0,494	-0,211	0,001	185	1	0,006	0,007	0,004	0,033	0,029	0,072	185	0,061	0,033	0,232	0,187	0,409

Les contributions des individus. Calcul. Notons tout d'abord que la variance d'une composante principale égale la valeur propre de l'axe correspondant. Comme pour le calcul des autres contributions, la contribution d'un individu à l'inertie (ou variance) d'un axe s'obtient par le rapport entre la part qu'occupe cet individu dans le calcul de cette variance et la variance totale. Ainsi, si on note c_{ij} la coordonnée de l'individu i sur l'axe j, et λ_j la valeur propre j, on a : $\text{contr}(\text{ind}_i,\text{axe}_j)=c_{ij}^2/(n\lambda_j)$.
Exercice. Retrouver certaines des contributions du tableau ci-dessus.

Utilisation des contributions. On relèvera toujours les individus dont la contribution est la plus élevée. Le seuil choisi pour cela peut encore être la moyenne de ces contributions, qui égale 1/n.

Exemple : individus contribuant plus que la moyenne aux axes 1 et 2. Moyenne=1/n=1/154≈0,0065.

signes coordonnées	-	+
axe 1	26, 30, 33, 39, 41, 49, 64, 65, 87, 89, 91, 100, 118, 124, 127, 129, 169, 174	8, 42, 44, 86, 135, 136, 137, 138, 139, 140, 142, 156, 157, 159, 160, 161, 162, 164, 175, 176, 177, 178, 179, 185
axe 2	10, 17, 18, 19, 21, 27, 30, 31, 32, 38, 39, 46, 47, 50, 52, 107, 113, 114, 115, 116, 117, 119, 120, 124, 125, 126, 127, 128, 148, 163, 166, 168, 170, 171, 176, 179	7, 22, 35, 37, 54, 58, 60, 62, 64, 65, 76, 87, 89, 91, 100, 103, 142, 164, 174, 184

Interprétation des contributions. Les individus côté négatif sur l'axe 1 ont contribué à la construction de cet axe par le fait qu'ils sont à majorité en SCEA, valorisation cave particulière, produisant de l'AOC, Prospects, et plutôt du NH; ils ont contribué par leurs caractéristiques opposées à ceux du côté positif, beaucoup de BRAM, en EARL, Adhérents, et produisant de l'AO-VDQS.

Exercice. Effectuer une interprétation sur ce modèle pour l'axe 2.

Les cosinus carrés des individus. Calcul. Le calcul se fait toujours sur le même modèle, par le rapport entre la norme projetée sur l'axe et la norme totale. A vérifier en exercice.

Utilisation des cosinus carrés. La moyenne des cosinus carrés égale toujours 1/(nb d'axes).

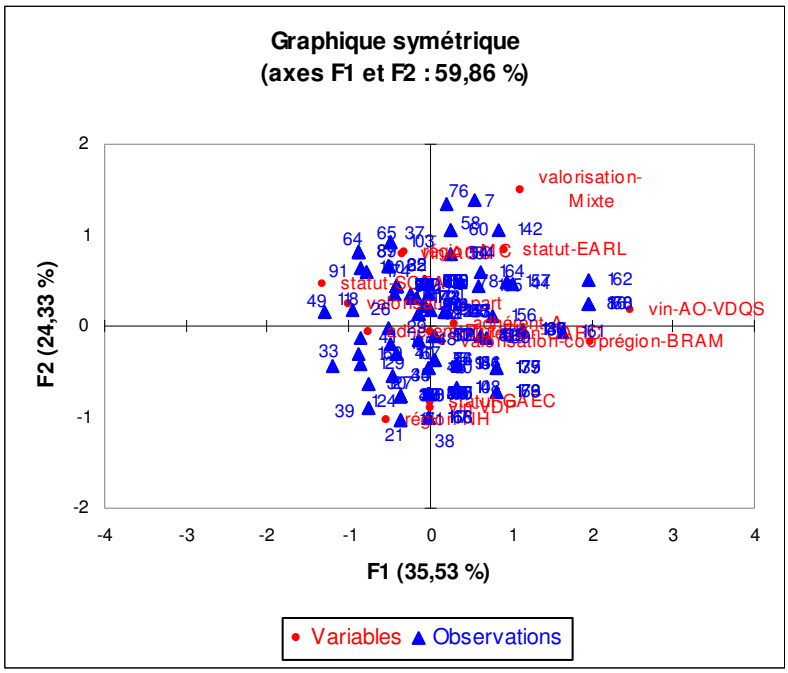
Exemple : individus reconstitués plus que la moyenne sur les axes 1 et 2. Moyenne=1/n=1/11=0,0909.

signes coordonnées	-	+
axe 1	22, 26, 29, 30, 32, 33, 35, 39, 41, 47, 49, 62, 64, 65, 66, 71, 77, 85, 87, 89, 91, 94, 95, 98, 99, 100, 116, 118, 119, 124, 127, 128, 129, 165, 169, 170, 174, 180, 182	8, 42, 44, 86, 135, 136, 137, 138, 139, 140, 142, 156, 157, 159, 160, 161, 162, 163, 164, 175, 176, 177, 178, 179, 185
axe 2	9, 10, 12, 17, 18, 19, 21, 27, 30, 31, 32, 34, 38, 39, 43, 45, 46, 47, 50, 52, 107, 113, 114, 115, 116, 117, 119, 120, 124, 125, 126, 127, 128, 131, 146, 148, 150, 154, 163, 166, 168, 170, 171, 176, 179	1, 2, 3, 7, 16, 22, 35, 36, 37, 54, 58, 59, 60, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 75, 76, 77, 79, 85, 87, 88, 89, 91, 94, 95, 97, 98, 99, 100, 102, 103, 105, 106, 109, 142, 164, 174, 180, 182, 184

Interprétation des cosinus carrés. Dans cet exemple, tous les individus contribuant fortement ont aussi un cosinus carré assez élevé pour que leur position soit interprétable.
Sur l'axe 1, on a une séparation des viticulteurs plutôt du NH, SCEA, P, part et AOC, qui ont une coordonnée négative, des viticulteurs plutôt de BRAM, EARL, A et AO-VDQS. Ce sont des profils dominants. Il n'est pas du tout certain que tous les individus cités côté négatif aient toutes les caractéristiques citées, par exemple soient de NH. De même que tous les individus cité côté >0 ne sont pas forcément de BRAM, mais en tout cas ont un profil voisin de ceux de BRAM.
De même, l'axe 2 oppose un groupe de viticulteurs cités côté <0 comme ayant des profils proches de ceux du NH, GAEC, coop et VDP, à un groupe de viticulteurs plutôt du MC, EARL, mixte et AOC.
Rappel des variables citées sur ces deux axes (pour faciliter la lecture):

signes coordonnées	-	+
axe 1	NH, SCEA, P, part, AOC	BRAM, EARL, A, AO-VDQS
axe 2	NH, GAEC, coop, VDP	MC, EARL, mixte, AOC

Visualisation du nuage de points.
Nuage "rond", pas d'individu à comportement atypique.
On peut même difficilement différencier ces individus sur le graphique pour en faire des groupes bien nets. Les cosinus carrés sont heureusement là pour aider à savoir quels individus interpréter sur chaque axe.



V. Classification

Les trois chapitres précédents présentaient des méthodes visant à simplifier la lisibilité d'un fichier de données via la réduction de dimension de l'espace des variables.
La classification s'intéresse à la simplification des profils des individus, en les regroupant.
Ce regroupement se fait avec deux contraintes principales :

- les individus d'un même groupe doivent être les plus près possible les uns des autres;
- les individus de deux groupes différents doivent être les plus différents possible, c'est-à-dire les plus "éloignés" possible.

Ces contraintes sous-entendent la définition de la notion de "proximité" ou d'"éloignement" entre les individus. Il faudra donc introduire la notion d'indice de similarité ou de dissimilarité.
Il est à noter que les méthodes de classification peuvent s'adapter à la classification de tout ensemble d'objets, variables, individus, éléments référencés cartographiquement, … à condition de pouvoir évaluer la proximité entre chacun de ces objets. Ce cours est dédié à la classification des individus. Le pas pour classifier d'autres éléments n'est pas grand. Les considérations exposées étant facilement généralisables.

V.1. Similarités et dissimilarités

Une similarité est une quantité qui est d'autant plus élevée que les individus sont semblables.

Exemple : nombre de points communs entre deux individus. Surtout utilisé pour des variables qualitatives. Dans l'exemple sur la viticulture,
sim(1,2)=5, ce qui signifie que 1 et 2 ont un même profil; sim(1, 8)=2 ➔ l'individu 1 a moins de points communs avec 8 qu'avec 2 (en tout cas parmi les points connus).

no	région	statut	adhérent	valorisation	vin
1	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
2	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
3	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC
5	MINERVOIS-CORBIERES	GAEC	A	Cave coopérative	AOC
7	MINERVOIS-CORBIERES	EARL	A	Mixte	AOC
8	CARCASSONNE	EARL	A	Cave coopérative	VDP
9	CARCASSONNE	EI	A	Cave coopérative	VDP
10	CARCASSONNE	GAEC	A	Cave coopérative	VDP
11	CARCASSONNE	SCEA	A	Cave coopérative	AOC
12	CARCASSONNE	EI	P	Cave coopérative	VDP
15	MINERVOIS-CORBIERES	GAEC	A	Cave coopérative	VDP
16	MINERVOIS-CORBIERES	EI	A	Cave coopérative	AOC

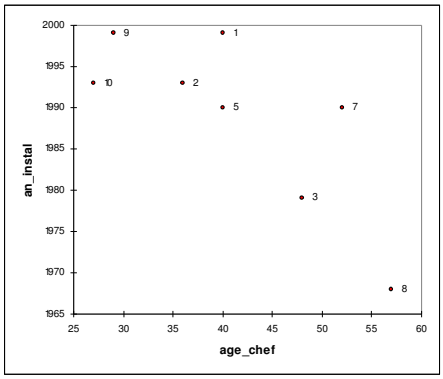
Il existe des indices de similarité beaucoup plus évolués que celui qui vient d'être présenté. Chacun répond à des propriétés intéressantes.

Une dissimilarité, au contraire, est une quantité qui est d'autant plus élevée que les individus sont différents.

Exemple : toute distance est une dissimilarité. Prenons la distance euclidienne, la plus classique (la longueur du plus court chemin entre 2 points).

Reprenant les mêmes individus que dans l'exemple précédent, on considère deux variables quantitatives, que sont l'âge du chef d'exploitation et l'année d'installation de ce chef. On peut calculer la distance entre chacun des points du nuage ainsi obtenu.
$d(1,2)= \sqrt{((1999-1993)^2+(40-36)^2)} \approx 7{,}211$
$d(1,8)= \sqrt{((1999-1968)^2+(40-57)^2)} \approx 35{,}355$.
L'individu 1 est plus distant de 8 que de 2.
On le voit bien sur la représentation graphique.

no	age_chef	an_instal
1	40	1 999
2	36	1 993
3	48	1 979
5	40	1 990
7	52	1 990
8	57	1 968
9	29	1 999
10	27	1 993



Remarque 1. Toute dissimilarité n'est pas forcément une distance. Une distance d vérifie les propriétés suivantes:

- $d(A,B) \geq 0$, pour tout A et B,
- $d(A,B)=0 \Rightarrow A=B$,
- $d(A,B)+d(B,C) \leq d(A,C)$

Remarque 2. On peut, par des transformations adéquates, obtenir une similarité à partir d'une dissimilarité, et vice-versa.

V.2. Variances intra et inter

Soit X une série statistique $(X_1, X_2, \ldots, X_n)$ mesurée sur un échantillon $(e_1, e_2, \ldots, e_n)$ et Y une variable qualitative qui a pour modalités $y_1, y_2, \ldots, y_k$ mesurée sur le même échantillon $(e_1, e_2, \ldots, e_n)$. Les valeurs prises par Y seront notées $(Y_1, Y_2, \ldots, Y_n)$. On note m_i le nombre d'individus de $(e_1, e_2, \ldots, e_n)$ prenant la valeur y_i de Y.
Notons que la connaissance des valeurs de Y pour l'échantillon $(e_1, e_2, \ldots, e_n)$ revient à constituer une *partition* de l'ensemble $E=\{e_1, e_2, \ldots, e_n\}$.
Les partitions $\mathcal{P}=\{A_1, A_2, \ldots, A_k\}$ que nous considérerons auront les propriétés suivantes :

- pour tout i, $Ai \neq \emptyset$
- pour tout $i \neq j$, $Ai \cap Aj = \emptyset$
- $A_1 \cup A_2 \cup \ldots \cup A_k = E$.

Selon les domaines dans lesquels est utilisée la classification, une *partie* de E sera aussi appelée un *groupe*, un *segment* (exemple : segment de marché), une *classe*, un *taxon* (utilisé en taxonomie), un *secteur*, une *catégorie*, une *tranche*, …

Soit $\overline{X_i}$ la moyenne des valeurs x_j pour les individus qui prennent la valeur y_i de Y, et $\overline{X}$ la moyenne de X pour les n individus.
Alors $\overline{X} = (\sum_{i=1}^{k} m_i \overline{X_i})/n$. Les $\overline{X_i}$ sont les moyennes intra-classes, $\overline{X}$ est la moyenne totale.
Par rapport à cette moyenne, on peut définir la variance totale, et les variances intra-classes et inter-classes comme suit :

Définitions.
La **variance totale** de X est la moyenne des carrés des écarts à la moyenne : $V(X)= \sum_{j=1}^{n}(X_j - \overline{X})^2/n$.
La **variance intra-classes** $W_Y(X)$ de X par rapport à Y est la moyenne des variances de X dans chaque classe de valeurs de Y :
$$W_Y(X)=(\sum_{i=1}^{k} m_i V_i)/n, \text{ où } V_i =\sum(X_{j/Yj=yi} - \overline{X_i})^2 /m_i. \quad \text{(W comme within, terme anglais)}$$
La **variance inter-classes** $B_Y(X)$ de X par rapport à Y est la variance des moyennes de X dans chaque classe de valeurs de Y :
$$B_Y(X) =\sum_{i=1}^{k} m_i (\overline{X_i} - \overline{X})^2 /n. \quad \text{(B comme between)}$$

Propriété. Pour toute variable quantitative X et toute variable qualitative Y, on a $V(X)= W_Y(X) + B_Y(X)$.

Ces définitions sont données ici dans le cas d'une seule variable quantitative. Elles sont généralisables au cas de plusieurs variables, et on parlera plutôt dans ce cas d'inerties totale, inter et intra. La propriété précédente, qu'on écrit aussi $I=I_W+I_B$, reste valable pour X dans n'importe quelle dimension.

Pour en revenir au problème de la classification, nous avions fixé comme contrainte celle de mettre dans des mêmes classes des individus proches, ce qui revient, en termes d'inertie, à trouver une partition d'inertie intra la plus faible possible.
De même, nous avions fixé comme but de mettre dans des groupes différents des individus éloignés, ce qui revient à faire en sorte d'avoir une inertie inter la plus grande possible.
Ces deux contraintes n'en sont finalement qu'une, vue la propriété $I=I_W+I_B$, qui précise que la somme des inerties intra et inter reste constante, et donc que augmenter l'inertie intra revient à diminuer l'inertie inter et inversement.

V.3. Pourquoi des méthodes approchées de classification

Si l'on savait trouver, pour tout ensemble de données, *LA* partition de variance intra minimum, il n'y aurait pas besoin de méthodes approchées, et notre cours s'arrêterait là. Or, voyons avec quelques calculs que cela n'est pas possible, et ne le sera pas avant d'avoir des ordinateurs dont la puissance est à la mesure des masses de données à traiter.

Le problème est dans le nombre de partitions possibles d'un ensemble, même modeste.
Posons $N_{k,n}$ le nombre de partitions contenant exactement k parties d'un ensemble E_n à n éléments.

Exercice. Que valent $N_{1,n}$, $N_{n,n}$, $N_{n-1,n}$, et $N_{2,n}$?

Remarquer la relation de récurrence $N_{k,n} = N_{k-1,n-1} + k\, N_{k,n-1}$.

Avec ces données, nous pouvons remplir le tableau suivant des $N_{k,n}$.

k \ n	2	3	4	5	6	7	8	9	10	11	12
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											

Le nombre de partitions total d'un ensemble à n éléments sera noté N_n : $N_n= \sum_k N_{k,n}$.
Ce nombre augmente plus qu'exponentiellement en fonction de n puisque, si $N_1=1$, et $N_2=2$, on a $N_{20} \approx 5.10^{13}$.
Dès que n est élevé, il n'est donc pas possible de passer en revue toutes les partitions possibles pour en déduire une "meilleure" au sens de l'inertie intra. C'est ce qui justifie l'utilisation de méthodes approchées, qui auront tendance à trouver une solution proche de la solution exacte, c'est-à-dire d'inertie intra minimum.

V.4. Les méthodes hiérarchiques

V.4.1. Présentation des méthodes

Elles consistent à établir, à partir d'un ensemble d'individus et de la connaissance d'un indice de similarité, ou de dissimilarité, une hiérarchie de partitions allant de n parties à 1 partie.
Définissons d'abord ce qu'est une hiérarchie de partitions.
Définition. Soit $E=\{e_1, e_2, \ldots, e_n\}$ un ensemble d'individus. Soit $\{\mathscr{P}_i\}_{i=1,\ldots,l}$ un ensemble de l partitions de E, telles que, sans perte de généralité, i<j => nombre de parties($\mathscr{P}_i$) < nombre de parties($\mathscr{P}_j$). On dit alors que cet ensemble de partitions de E est hiérarchisé si, lorsque i<j, toute partie de $\mathscr{P}_i$ peut être obtenue comme réunion de parties de $\mathscr{P}_j$.

Exemple. Prenons n=5 : $E=\{e_1, e_2, e_3, e_4, e_5\}$. Considérons les ensembles de partitions suivants. Indiquer si ces ensembles $\{P_1, P_2, P_3, P_4, P_5\}$ sont hiérarchisés ou pas.

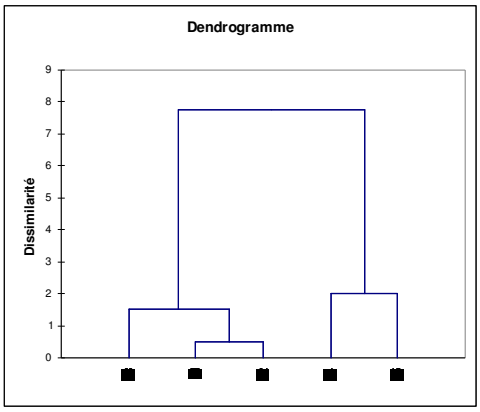
$P_1 = \{\{e_1, e_2, e_3, e_4, e_5\}\}$ $P_2 = \{\{e_1\},\{e_2, e_3, e_4, e_5\}\}$ $P_3 = \{\{e_1\},\{e_2, e_5\},\{e_3, e_4\}\}$ $P_4 = \{\{e_1\},\{e_2, e_5\},\{e_3\},\{e_4\}\}$ $P_5 = \{\{e_1\},\{e_2\},\{e_3\},\{e_4\},\{e_5\}\}$	$P_1 = \{\{e_1, e_2, e_3, e_4, e_5\}\}$ $P_2 = \{\{e_1, e_2, e_5\},\{e_3, e_4\}\}$ $P_3 = \{\{e_1, e_2\},\{e_3, e_5\},\{e_4\}\}$ $P_4 = \{\{e_1\},\{e_2, e_5\},\{e_3\},\{e_4\}\}$ $P_5 = \{\{e_1\},\{e_2\},\{e_3\},\{e_4\},\{e_5\}\}$	$P_1 = \{\{e_1, e_2, e_3, e_4, e_5\}\}$ $P_2 = \{\{e_1, e_5\},\{e_2, e_3, e_4\}\}$ $P_3 = \{\{e_1, e_5\},\{e_2, e_3\},\{e_4\}\}$ $P_4 = \{\{e_1\},\{e_5\},\{e_2, e_3\},\{e_4\}\}$ $P_5 = \{\{e_1\},\{e_2\},\{e_3\},\{e_4\},\{e_5\}\}$

Avec un ensemble de partitions hiérarchisées, qu'on appelle encore hiérarchie de partitions, on peut construire un arbre hiérarchique, ou dendrogramme.

Exemple. la hiérarchie des partitions suivante

$P_1 = \{\{e_1, e_2, e_3, e_4, e_5\}\}$
$P_2 = \{\{e_1, e_2, e_3\},\{e_4, e_5\}\}$
$P_3 = \{\{e_1, e_2, e_3\},\{e_4\},\{e_5\}\}$
$P_4 = \{\{e_1, e_2\},\{e_3\},\{e_4\},\{e_5\}\}$
$P_5 = \{\{e_1\},\{e_2\},\{e_3\},\{e_4\},\{e_5\}\}$

donne la représentation graphique ci-contre.



En plus de l'information qu'apporte la partition hiérarchisée, on peut ajouter sur le dendrogramme une information sur la proximité entre les individus, et entre les classes. En effet, on voit sur l'échelle verticale du dendrogramme un indicateur de dissimilarité, c'est-à-dire de proximité entre les individus. Par exemple, e_1 et e_2 sont plus proches entre eux que e_4 et e_5.

On peut distinguer deux ensembles de méthodes de classification hiérarchique.
Les méthodes **ascendantes** ont comme point de départ la partition des singletons, et par étapes successives, réunissent les classes deux à deux pour avoir en fin d'étapes la partition à un seul ensemble. On peut se souvenir de ce qualificatif "ascendante" par le fait que sur le dendrogramme la première étape est illustrée par le bas, et les étapes font "remonter" dans la représentation graphique. Le terme anglais est "agglomerative", plus parlant et indépendant de l'orientation du graphique.
Les méthodes **descendantes** procèdent à l'inverse. Leur point de départ est la partition à 1 ensemble, et à la fin du processus on obtient la partition des singletons. Le terme anglais, "divisive", est, une fois de plus, plus parlant que "descendant". Un avantage mnémotechnique est que les initiales sont les mêmes pour chaque groupe de méthodes.

Dans ce cours, on s'intéressera aux méthodes ascendantes, moins lourdes en calcul et plus généralisées dans les logiciels de statistique. Pour ces méthodes, le passage d'une partition de k classes à k-1 classes se fait par la réunion des deux classes les plus "près" en un certain sens.
On a vu les notions de similarité et de dissimilarité entre individus. Il faut aussi définir des notions de proximité entre classes d'individus. Plusieurs méthodes émergent parmi le nombre considérable d'indices existant dans la littérature. Voici quelques indices des plus courants entre deux parties A et B de E :

- saut minimum : plus petite dissimilarité entre un élément de A et un élément de B;
- saut maximum ou diamètre : plus grande dissimilarité entre un élément de A et un élément de B;
- saut moyen : moyenne des dissimilarités entre un élément de A et un élément de B;
- distance entre les barycentres de A et de B = $d(g_A, g_B)$;
- critère de Ward = distance entre les barycentres de A et de B pondérée par les poids de A et B = $d(g_A, g_B)(P_AP_B/(P_A+P_B))$; les poids sont en général les effectifs de A et B.

V.4.2. Application à la classification de données quantitatives

Reprenons l'exemple des exploitations en viticulture, pour lesquelles on considère les caractéristiques suivantes :

l'année d'installation (an_instal), l'âge du chef d'exploitation (age_chef), la surface agricole utile totale (sau_totale), la sau de vignes de plus de 3 ans (sau_vign>3ans), la sau de vignes de moins de 3 ans (sau_vign<3ans), et le nombre de salariés équivalent temps plein (salariés).

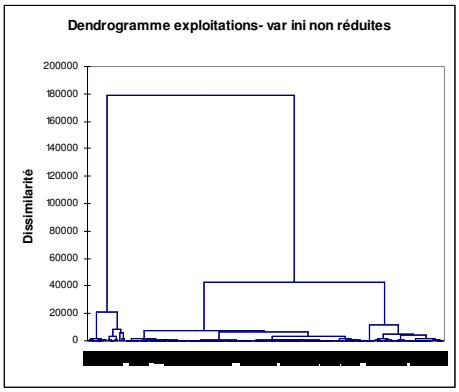
no	an_instal	age_chef	sau_totale	sau_vign>3ans	sau_vign<3ans	salariés
1	1 999	40	25,00	21,00	4,00	0,00
2	1 993	36	22,50	19,70	2,80	0,00
3	1 979	48	69,00	53,58	3,60	1,00
5	1 990	40	55,00	45,00	10,00	0,00
7	1 990	52	52,00	4,50	44,00	3,00
8	1 968	57	70,00	19,00	5,00	2,00
9	1 999	29	30,00	16,50	3,50	1,00
10	1 993	27	42,00	36,00	6,00	0,00
11	2 005	23	80,00	25,40	1,60	0,00
12	1 985	25	40,00	25,00	0,60	0,00
15	1 987	42	70,00	10,00	58,00	1,00
16	1 998	37	22,00	3,00	19,00	0,00
17	1 988	46	29,00	25,00	4,00	0,00
18	1 991	38	35,00	27,80	7,20	1,00
19	1 978	49	24,00	22,00	2,00	0,00
21	1 994	34	40,00	34,00	6,00	0,00
22	1 998	35	43,55	37,80	5,75	1,00
25	1 988	50	30,00	26,50	3,50	1,00
26	2 002	50	37,00	35,00	2,00	2,00
27	1 978	49	23,00	19,50	1,20	0,00
29	1 975	55	28,00	25,00	3,00	0,00
30	1 989	38	52,00	44,00	8,00	1,00
31	1 996	33	25,70	21,40	4,30	0,00
32	1 990	37	38,00	30,00	4,70	0,00
33	1 994	37	220,00	137,00	15,00	15,00
34	2 002	33	29,00	26,57	1,99	0,00
35	1 976	52	22,50	2,50	19,00	0,00
36	2 002	30	20,00	18,00	2,00	0,00
37	1 993	35	100,00	98,50	1,50	5,00
38	2 000	38	55,00	49,50	5,50	1,00
39	2 001	30	30,00	26,50	3,50	1,00
40	1 975	48	36,00	29,00	0,00	1,00
41	1 975	28	104,00	29,00	1,00	3,00
42	2 001	45	57,00	40,00	4,00	2,00
43	1 998	32	21,00	19,50	1,50	1,00
44	1 983	47	70,00	28,00	1,00	0,00
45	1 969	56	120,00	16,00	6,00	1,00
46	2 001	23	12,00	10,00	2,00	0,00
47	1 974	53	180,00	26,00	1,00	1,00
48	1 993	42	60,00	54,00	3,80	0,00
49	1 990	41	90,00	63,00	12,00	7,00
50	2 001	25	23,00	10,00	13,00	1,00
52	2 000	33	24,00	14,00	2,00	0,00
53	1 976	54	19,00	3,00	16,00	0,00
54	1 981	53	62,00	5,00	57,00	1,00
56	1 988	40	52,00	38,00	12,00	1,00
57	1 985	41	19,72	2,00	17,72	0,00
58	1 995	60	35,00	18,00	0,00	0,00
59	1 986	35	31,00	4,00	0,00	2,00
60	1 990	56	80,00	80,00	0,00	2,00
61	1 989	43	37,00	35,00	2,00	0,00
62	2 000	35	36,00	30,00	6,00	1,00
63	1 985	39	13,50	12,00	1,50	0,00
64	1 976	57	30,00	27,00	3,00	2,00
65	1 998	43	150,00	94,00	3,00	5,00
66	1 981	50	30,00	26,00	4,00	1,00
67	1 981	58	42,00	34,50	1,50	1,00
68	1 978	58	25,00	15,00	1,40	0,00
69	1 995	34	17,00	17,00	0,00	0,00
70	2 000	40	23,00	22,00	1,00	0,00
71	1 984	47	26,00	16,00	0,00	0,00
72	1 985	45	28,00	23,00	5,00	2,00
74	2 004	48	30,00	27,00	3,00	2,00
75	1 983	52	21,00	19,00	2,00	0,00
76	1 981	47	70,00	7,50	0,00	2,00
77	1 986	43	11,00	11,00	0,00	0,00
78	1 986	45	50,00	48,00	2,00	2,00
79	1 989	51	29,00	20,00	0,00	0,00
84	2 002	53	35,00	26,50	3,50	0,00
85	1 992	39	18,00	16,00	0,00	0,00
86	1 984	41	115,00	14,20	1,20	1,00
87	2 004	37	64,00	45,00	8,00	2,00
88	1 981	46	26,00	24,00	2,00	0,00
89	1 989	43	220,00	65,00	11,00	6,00
90	1 982	49	20,00	19,00	1,00	0,00
91	1 993	30	15,00	15,00	0,00	0,00
94	1 990	41	24,00	22,00	2,00	0,00

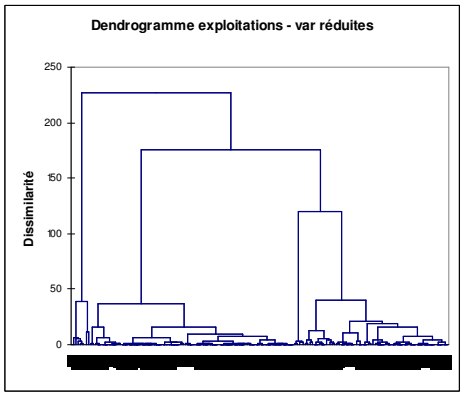
no	an_instal	age_chef	sau_totale	sau_vign>3ans	sau_vign<3ans	salariés
95	1 994	36	32,00	28,00	4,00	0,00
97	2 000	36	34,00	31,00	3,00	0,00
98	2 000	31	30,00	4,50	15,00	0,00
99	1 994	37	15,00	3,00	12,00	0,00
100	2 001	43	59,00	9,00	50,00	1,00
101	1 984	45	28,00	3,00	25,00	0,75
102	2 001	36	22,64	18,64	4,00	0,00
103	1 985	43	42,00	38,00	4,00	1,00
104	1 984	49	34,00	29,00	5,00	0,00
105	2 002	30	29,00	26,00	3,00	0,00
106	1 996	33	26,00	24,00	2,00	0,00
107	1 969	52	30,00	27,00	3,00	0,50
108	2 002	25	35,00	27,50	2,50	0,00
109	1 998	36	0,00	24,00	2,00	0,00
110	1 992	42	68,00	61,00	7,00	2,00
112	1 998	43	23,00	18,00	5,00	0,00
113	1 995	38	22,00	19,40	1,00	0,00
114	1 994	41	23,00	17,50	3,00	0,00
115	1 995	41	21,80	20,90	0,90	1,00
116	1 975	47	21,00	13,00	0,00	0,00
117	1 967	56	55,00	42,50	0,00	0,00
118	2 002	34	26,00	24,00	2,00	1,00
119	1 981	45	22,00	22,00	0,00	0,00
120	1 990	45	26,00	21,00	5,00	1,00
124	1 994	38	165,00	75,00	5,00	2,00
125	2 000	23	18,00	14,00	3,00	0,00
126	2 000	27	31,00	20,00	4,00	0,00
127	2 000	37	23,00	18,00	1,80	0,75
128	2 003	36	19,00	14,00	0,00	0,00
129	1 967	45	130,00	95,00	5,00	12,00
131	2 000	36	20,50	12,00	8,00	1,00
135	1 991	40	52,00	36,00	1,00	1,00
136	2 002	30	16,00	11,00	5,00	0,00
137	1 987	54	14,00	11,00	3,00	0,00
138	1 995	69	55,00	36,00	4,36	2,00
139	1 986	41	46,00	0,00	2,00	1,00
140	2 001	29	16,00	14,00	2,00	0,00
142	1 980	53	135,00	32,00	12,00	0,00
143	1 980	45	106,00	18,00	3,00	0,00
144	2 002	38	39,00	33,00	6,00	1,00
146	2 003	24	28,00	28,00	0,00	0,00
147	1 996	41	27,00	17,00	3,00	0,00
148	1 991	36	60,00	14,50	3,00	0,00
149	2 002	27	27,00	0,00	0,00	0,00
150	2 000	28	25,00	0,00	0,00	0,00
151	2 004	27	17,00	0,00	0,00	0,00
152	2 002	26	15,00	15,00	0,00	0,00
153	1 996	41	17,00	17,00	0,00	0,00
154	1 990	36	30,00	27,00	0,00	0,00
156	2 000	28	9,00	8,00	1,00	0,00
157	1 989	47	60,00	34,00	3,00	0,00
159	2 002	25	155,00	62,00	8,00	0,00
160	1 995	32	140,00	23,00	2,00	1,00
161	2 004	34	16,00	15,30	0,70	0,00
162	1 996	31	24,00	19,00	5,00	1,00
163	2 005	22	110,00	40,00	5,00	0,00
164	1 979	47	66,00	32,00	0,00	1,00
165	2 000	33	25,00	17,00	3,00	0,00
166	1 995	56	30,00	29,00	2,00	0,00
167	1 973	52	68,00	58,00	0,00	1,50
168	1 997	30	31,00	26,00	1,50	0,00
169	1 999	57	25,00	23,00	2,00	0,00
170	1 990	40	19,00	16,00	3,00	0,00
171	2 000	29	25,00	18,00	7,00	0,00
172	1 998	42	30,00	28,50	1,50	1,00
173	1 998	42	17,00	14,78	0,00	2,00
174	2 003	32	32,00	30,00	2,00	2,00
175	1 994	34	70,00	9,00	0,00	0,00
176	2 000	29	62,00	21,00	1,00	0,00
177	1 997	49	17,00	15,00	0,00	0,00
178	1 989	40	115,00	36,00	0,00	1,00
179	1 989	39	82,00	39,00	2,00	0,00
180	1 996	32	24,00	24,00	0,00	0,00
182	1 998	38	24,00	23,00	1,00	0,00
183	2 002	30	27,00	24,50	2,50	0,50
184	1 984	43	18,00	15,00	3,00	0,50
185	1 995	38	56,50	44,00	8,00	2,00

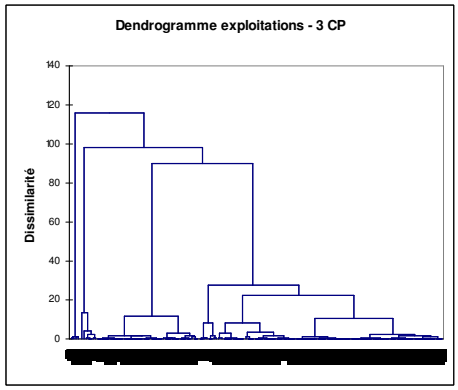
On pourrait effectuer une classification directement à partir de ces données. Cependant, si on considère la distance euclidienne comme critère de proximité (directement calculable quand on est en présence de données quantitatives), on voit bien que les variables de variance élevée auront une plus forte influence dans la distance entre exploitations que les variables de plus faible variance.

	an_instal	age_chef	sau_totale	sau_vign>3ans	sau_vign<3ans	salariés
Nbr. de valeurs utilisées	154	154	154	154	154	154
Minimum	1967,000	22,000	0,000	0,000	0,000	0,000
1er quartile	1985,000	33,000	23,000	15,000	1,000	0,000
Médiane	1994,000	40,000	30,000	23,000	2,900	0,000
3ème quartile	2000,000	47,000	55,000	31,000	5,000	1,000
Maximum	2005,000	69,000	220,000	137,000	58,000	15,000
Moyenne	1991,675	40,214	45,415	26,350	4,975	0,834
CV (écart-type/moyenne)	0,005	0,233	0,856	0,749	1,787	2,211
Ecart-type d'échantillon	9,333	9,357	38,769	19,685	8,861	1,839

Ainsi, une classification à partir de ces données directement tiendrait plus compte de la proximité entre les valeurs de sau_vign>3ans que des autres variables. Pour palier cet effet des différences entre les dispersions, on peut réduire toutes les variables avant classification. Cela revient au même que d'effectuer la classification avec toutes les composantes principales de l'ACP réduite de ces variables. Enfin, pour ne tenir compte que des informations les plus marquantes dans le fichier, et aussi utiliser une taille de fichier plus faible, on peut aussi effectuer la classification à partir des premières composantes principales de l'ACP, réduite dans notre cas.







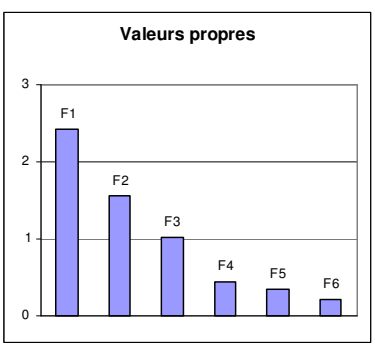
Statistiques des noeuds CAH - 6 variables non réduites :						Statistiques des noeuds CAH - 6 variables réduites :						Statistiques des noeuds CAH - 3 CP de l'ACP sur 6 variables :					
Noeud	Niveau	Poids	Objets	Fils gauche	Fils droit	Noeud	Niveau	Poids	Objets	Fils gauche	Fils droit	Noeud	Niveau	Poids	Objets	Fils gauche	Fils droit
307	179052,301	154	154	305	306	307	226,754	154	154	303	306	307	116,018	154	154	285	306
306	42538,714	138	138	302	304	306	175,291	147	147	305	302	306	98,395	150	150	302	305
305	21317,561	16	16	303	290	305	119,373	63	63	278	304	305	90,116	143	143	301	304
304	11965,838	34	34	268	299	304	39,878	59	59	296	301	304	27,540	101	101	298	303
303	8676,671	8	8	300	296	303	38,856	7	7	289	295	303	22,299	93	93	300	299
302	7836,990	104	104	294	301	302	37,027	84	84	299	297	302	13,244	7	7	25	297
301	6473,387	77	77	288	297	301	21,411	46	46	294	300	301	12,003	42	42	295	291
300	6175,167	3	3	25	293	300	19,175	35	35	273	298	300	10,437	61	61	281	292
299	5091,694	30	30	292	298	299	16,111	71	71	290	293	299	8,514	32	32	296	294
298	3818,959	21	21	283	291	298	15,514	31	31	284	292	298	8,218	8	8	278	286
297	3106,987	49	49	295	289	297	15,510	13	13	261	288	297	4,314	6	6	293	107
296	3017,183	5	5	285	281	296	12,635	13	13	283	287	296	3,479	21	21	284	290
295	1951,384	13	13	256	284	295	11,414	2	2	25	107	295	3,013	15	15	274	289
294	1840,776	27	27	279	259	294	10,454	11	11	272	277	294	2,840	11	11	268	279
293	1785,500	2	2	39	74	293	9,179	45	45	258	291	293	2,260	5	5	272	269
292	1475,557	9	9	273	282	292	7,463	23	23	274	285	292	2,131	38	38	266	288
291	1397,775	16	16	280	271	291	7,164	41	41	281	279	291	2,055	27	27	267	287
290	1312,866	8	8	287	278	290	6,520	26	26	275	263	290	1,994	17	17	277	275
289	1159,747	36	36	269	286	289	6,306	5	5	74	286	289	1,959	8	8	283	276
288	1152,640	28	28	267	264	288	6,027	11	11	237	276	288	1,822	22	22	282	271
287	998,151	5	5	266	274	287	5,490	9	9	268	282	287	1,691	16	16	258	263
286	921,873	29	29	272	242	286	5,458	4	4	41	280	286	1,550	4	4	228	265
285	874,750	2	2	29	107	285	4,443	13	13	239	265	285	1,452	4	4	273	280
284	868,944	9	9	249	276	284	4,290	8	8	112	270	284	1,056	4	4	50	245
283	677,506	5	5	270	261	283	4,191	4	4	251	266	283	0,947	3	3	214	115
282	654,008	5	5	258	265	282	4,140	4	4	33	267	282	0,940	12	12	242	250
281	607,167	3	3	55	263	281	3,945	16	16	262	252	281	0,901	23	23	254	264
280	562,340	5	5	277	112	280	3,328	3	3	259	102	280	0,896	2	2	5	82
279	561,940	21	21	237	275	279	3,278	25	25	264	254	279	0,864	9	9	209	261
278	502,500	3	3	133	262	278	3,272	4	4	250	271	278	0,790	4	4	248	262
277	490,688	4	4	253	98	277	2,960	3	3	50	227	277	0,769	8	8	239	243
276	488,396	7	7	254	224	276	2,458	8	8	260	230	276	0,761	5	5	256	259
275	454,629	17	17	243	252	275	2,337	9	9	269	199	275	0,607	9	9	260	200
274	436,500	2	2	115	130	274	2,214	10	10	241	249	274	0,570	7	7	270	98
273	425,794	4	4	9	245	273	2,195	4	4	172	235	273	0,567	2	2	11	45
272	422,583	25	25	248	257	272	2,160	8	8	253	242	272	0,557	2	2	41	74
271	399,574	11	11	247	255	271	2,007	2	2	5	82	271	0,529	10	10	249	252
270	391,500	2	2	41	50	270	1,910	7	7	234	244	270	0,487	6	6	244	230
269	384,619	7	7	91	233	269	1,891	7	7	243	224	269	0,449	3	3	241	102
268	354,313	4	4	244	251	268	1,881	5	5	223	257	268	0,410	2	2	129	133
267	329,501	12	12	250	260	267	1,875	3	3	130	236	267	0,395	11	11	202	246
266	295,813	3	3	37	228	266	1,833	2	2	39	115	266	0,389	16	16	257	253
265	271,667	3	3	197	149	265	1,647	5	5	231	195	265	0,384	2	2	47	83
264	262,593	16	16	239	240	264	1,532	15	15	256	229	264	0,378	14	14	222	255
263	257,500	2	2	33	148	263	1,479	17	17	226	255	263	0,376	8	8	219	236
262	257,500	2	2	102	129	262	1,474	12	12	176	247	262	0,351	2	2	12	81
261	249,998	3	3	92	213	261	1,449	2	2	129	133	261	0,277	5	5	72	240
260	241,402	6	6	10	234	260	1,429	5	5	204	245	260	0,267	5	5	67	247
259	227,917	6	6	226	241	259	1,372	2	2	29	55	259	0,258	2	2	39	137
258	213,125	2	2	6	65	258	1,320	4	4	42	233	258	0,248	8	8	232	237
257	205,090	14	14	230	220	257	1,310	3	3	65	207	257	0,230	7	7	210	223
256	197,025	4	4	181	229	256	1,282	9	9	238	211	256	0,227	3	3	112	238
255	190,557	7	7	227	246	255	1,234	14	14	209	240	255	0,216	10	10	173	233
254	185,813	4	4	225	221	254	1,177	10	10	194	246	254	0,208	9	9	167	251
253	182,500	3	3	57	199	253	1,159	6	6	248	203	253	0,202	9	9	190	234
252	163,406	9	9	217	236	252	1,110	4	4	169	232	252	0,194	7	7	225	207
251	155,625	2	2	5	82	251	1,002	2	2	6	37	251	0,190	7	7	203	221
250	155,151	6	6	214	194	250	0,951	2	2	11	45	250	0,180	3	3	108	204
249	141,000	2	2	49	113	249	0,946	5	5	218	228	249	0,176	3	3	9	213
248	140,802	11	11	183	235	248	0,930	4	4	220	225	248	0,165	2	2	42	80
247	135,635	4	4	222	238	247	0,790	9	9	202	214	247	0,153	4	4	51	224
246	124,863	5	5	67	218	246	0,742	7	7	206	213	246	0,149	6	6	220	111
245	124,708	3	3	232	146	245	0,742	2	2	40	92	245	0,142	3	3	92	205
244	123,500	2	2	11	45	244	0,737	3	3	212	147	244	0,139	3	3	54	211
243	122,635	8	8	231	216	243	0,729	2	2	8	10	243	0,138	5	5	206	231
242	120,896	4	4	111	207	242	0,722	2	2	89	98	242	0,134	9	9	235	218
241	93,667	3	3	48	210	241	0,721	5	5	197	222	241	0,132	2	2	29	55
240	88,729	9	9	200	205	240	0,657	9	9	205	221	240	0,130	4	4	208	229
239	86,347	7	7	175	219	239	0,601	8	8	201	215	239	0,122	3	3	196	46
238	81,500	2	2	34	72	238	0,583	4	4	108	182	238	0,120	2	2	3	134
237	80,854	4	4	32	211	237	0,552	3	3	34	189	237	0,119	4	4	62	217

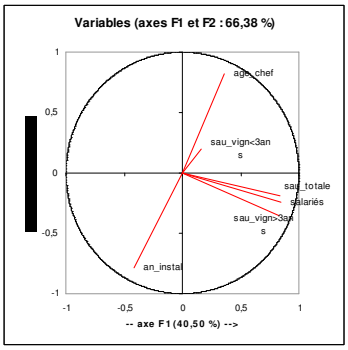
236	79,250	4	4	206	185	236	0,549	2	2	148	149	236	0,112	5	5	197	189
235	78,603	9	9	223	208	235	0,538	2	2	47	83	235	0,105	5	5	179	226
234	77,568	5	5	178	215	234	0,536	4	4	69	210	234	0,099	5	5	201	212
233	76,167	6	6	204	190	233	0,523	3	3	219	81	233	0,098	7	7	216	215
232	76,125	2	2	120	145	232	0,481	2	2	9	146	232	0,095	4	4	56	227
231	75,980	2	2	58	97	231	0,471	2	2	49	113	231	0,091	3	3	14	170
230	73,632	5	5	167	209	230	0,459	3	3	216	154	230	0,073	3	3	58	193
229	70,060	2	2	47	83	229	0,452	6	6	183	188	229	0,068	2	2	34	154
228	65,840	2	2	71	116	228	0,444	3	3	109	208	228	0,065	2	2	27	44
227	63,000	2	2	109	128	227	0,440	2	2	3	137	227	0,063	3	3	195	116
226	61,417	3	3	19	203	226	0,430	3	3	191	143	226	0,062	3	3	163	124
225	60,625	2	2	42	80	225	0,429	2	2	54	57	225	0,061	3	3	144	188
224	60,167	3	3	159	123	224	0,424	5	5	14	217	224	0,059	3	3	71	172
223	59,197	3	3	108	170	223	0,395	2	2	71	116	223	0,057	4	4	186	162
222	57,070	2	2	30	40	222	0,370	3	3	62	200	222	0,056	4	4	126	199
221	57,000	2	2	12	81	221	0,368	4	4	91	198	221	0,056	4	4	183	194
220	53,446	9	9	212	173	220	0,353	2	2	32	56	220	0,055	5	5	181	198
219	53,225	5	5	196	158	219	0,352	2	2	12	80	219	0,053	3	3	66	182
218	52,771	4	4	46	202	218	0,350	2	2	67	85	218	0,053	4	4	177	155
217	51,567	5	5	191	188	217	0,338	4	4	192	126	217	0,052	3	3	192	136
216	51,241	6	6	163	198	216	0,336	2	2	30	72	216	0,048	3	3	1	180
215	50,167	3	3	189	142	215	0,334	5	5	193	162	215	0,047	4	4	184	156
214	43,718	3	3	8	193	214	0,328	5	5	186	175	214	0,046	2	2	6	37
213	42,873	2	2	3	137	213	0,326	4	4	79	174	213	0,046	2	2	90	118
212	42,136	6	6	186	184	212	0,309	2	2	68	111	212	0,044	2	2	10	150
211	40,333	3	3	172	89	211	0,301	5	5	160	179	211	0,043	2	2	32	57
210	39,000	2	2	136	139	210	0,291	3	3	48	190	210	0,042	3	3	174	132
209	38,695	3	3	165	147	209	0,286	5	5	184	168	209	0,040	4	4	17	185
208	37,995	6	6	179	195	208	0,273	2	2	51	128	208	0,039	2	2	30	130
207	37,917	3	3	182	153	207	0,270	2	2	36	134	207	0,039	4	4	187	157
206	33,000	2	2	61	68	206	0,263	3	3	185	144	206	0,038	2	2	16	63
205	32,476	7	7	144	201	205	0,255	5	5	170	178	205	0,037	2	2	33	148
204	29,500	4	4	110	192	204	0,252	3	3	4	196	204	0,037	2	2	110	141
203	29,250	2	2	63	69	203	0,250	2	2	21	58	203	0,036	3	3	93	168
202	29,167	3	3	176	154	202	0,242	4	4	164	187	202	0,036	5	5	160	178
201	28,777	6	6	138	187	201	0,226	3	3	155	97	201	0,036	3	3	79	161
200	28,750	2	2	90	118	200	0,223	2	2	18	101	200	0,035	4	4	191	171
199	27,500	2	2	51	85	199	0,203	2	2	120	145	199	0,035	3	3	142	169
198	26,923	4	4	156	180	198	0,188	3	3	105	165	198	0,034	3	3	158	97
197	25,000	2	2	36	134	197	0,177	2	2	13	86	197	0,031	2	2	101	113
196	23,757	3	3	166	141	196	0,176	2	2	22	46	196	0,030	2	2	4	22
195	23,575	4	4	157	161	195	0,175	3	3	180	153	195	0,028	2	2	36	65
194	23,333	3	3	168	117	194	0,170	3	3	17	173	194	0,028	2	2	77	140
193	22,552	2	2	16	17	193	0,157	3	3	61	166	193	0,028	2	2	21	89
192	22,000	3	3	103	169	192	0,156	3	3	24	181	192	0,027	2	2	18	86
191	21,333	3	3	18	177	191	0,156	2	2	96	142	191	0,026	2	2	19	109
190	20,000	2	2	38	127	190	0,151	2	2	136	139	190	0,024	4	4	175	166
189	18,500	2	2	56	86	189	0,150	2	2	19	63	189	0,021	3	3	13	176
188	18,500	2	2	78	126	188	0,131	4	4	23	171	188	0,021	2	2	31	146
187	18,175	5	5	99	171	187	0,127	2	2	124	127	187	0,021	2	2	7	104
186	16,648	4	4	162	174	186	0,124	3	3	110	161	186	0,019	2	2	106	131
185	15,500	2	2	77	101	185	0,124	2	2	31	99	185	0,018	3	3	8	159
184	14,500	2	2	93	119	184	0,121	3	3	70	167	184	0,016	2	2	2	94
183	14,000	2	2	59	76	183	0,120	2	2	59	76	183	0,015	2	2	49	69
182	13,250	2	2	53	66	182	0,117	3	3	35	177	182	0,015	2	2	53	147
181	12,750	2	2	27	44	181	0,115	2	2	16	78	181	0,013	2	2	64	75
180	12,500	2	2	73	100	180	0,114	2	2	53	66	180	0,013	2	2	78	145
179	12,250	2	2	28	35	179	0,113	3	3	84	156	179	0,013	2	2	38	103
178	11,545	2	2	14	24	178	0,112	3	3	157	125	178	0,011	3	3	164	100
177	10,000	2	2	4	22	177	0,108	2	2	7	132	177	0,011	2	2	121	123
176	10,000	2	2	13	62	176	0,106	3	3	158	123	176	0,011	2	2	139	153
175	9,750	2	2	7	104	175	0,097	2	2	104	141	175	0,011	2	2	26	99
174	9,705	2	2	95	96	174	0,089	3	3	26	163	174	0,011	2	2	23	84
173	9,667	3	3	1	160	173	0,088	2	2	52	117	173	0,010	3	3	165	91
172	8,500	2	2	21	54	172	0,084	2	2	27	44	172	0,010	2	2	85	128
171	8,481	4	4	26	164	171	0,080	3	3	138	159	171	0,010	2	2	40	149
170	8,090	2	2	106	131	170	0,076	2	2	1	93	170	0,009	2	2	24	120
169	8,000	2	2	52	79	169	0,067	2	2	90	118	169	0,008	2	2	96	143
168	8,000	2	2	114	124	168	0,066	2	2	77	140	168	0,008	2	2	95	119
167	7,500	2	2	70	140	167	0,066	2	2	2	94	167	0,008	2	2	70	125
166	7,070	2	2	23	132	166	0,063	2	2	64	75	166	0,007	2	2	35	138
165	6,964	2	2	125	143	165	0,047	2	2	60	151	165	0,007	2	2	59	76
164	6,750	3	3	155	152	164	0,044	2	2	38	103	164	0,007	2	2	61	68
163	6,000	2	2	64	75	163	0,043	2	2	87	152	163	0,007	2	2	28	114
162	5,790	2	2	2	94	162	0,041	2	2	73	100	162	0,005	2	2	43	135
161	5,500	2	2	43	135	161	0,037	2	2	28	114	161	0,005	2	2	88	105
160	5,000	2	2	60	151	160	0,037	2	2	106	131	160	0,004	2	2	48	73
159	4,500	2	2	88	150	159	0,032	2	2	88	150	159	0,004	2	2	52	117
158	4,500	2	2	121	122	158	0,030	2	2	121	122	158	0,004	2	2	15	20
157	3,971	2	2	84	105	157	0,028	2	2	95	119	157	0,002	2	2	87	152
156	3,945	2	2	15	20	156	0,018	2	2	43	135	156	0,001	2	2	60	151
155	1,750	2	2	31	87	155	0,012	2	2	15	20	155	0,000	2	2	122	127

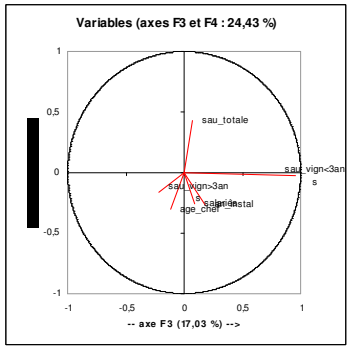
Les tableaux ci-dessus tracent toutes les étapes des trois classifications mentionnées.

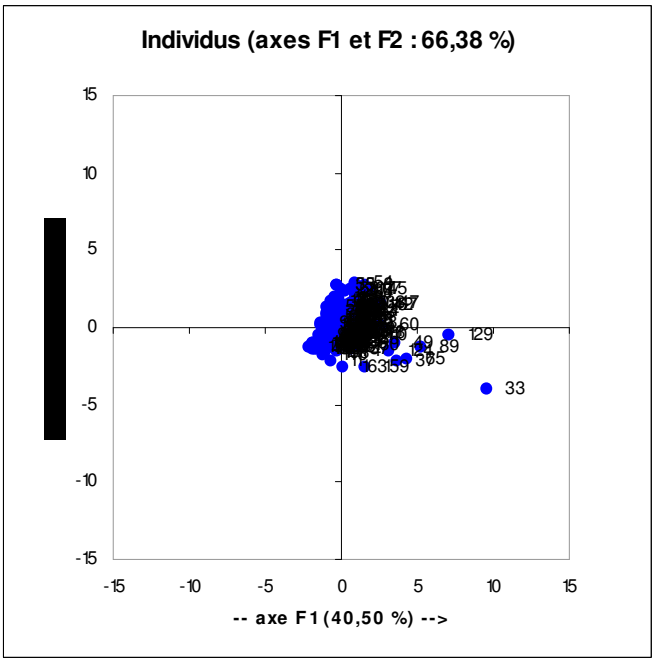
Exercice. Examiner les résultats de ces tableaux. A partir de ces tableaux, repérer les individus qui sont réunis en premier. Analyser les différences entre les premières réunions de chacun des 3 cas.

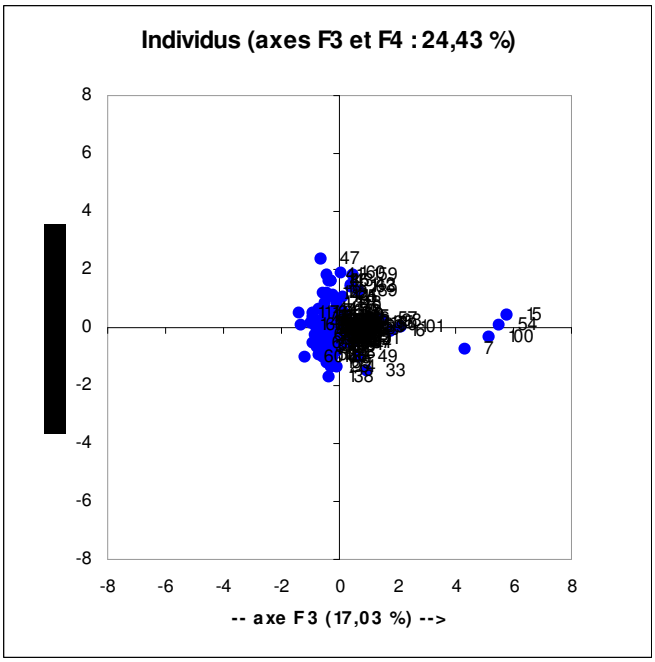
Pour information, les principaux résultats de l'ACP sont les suivants :





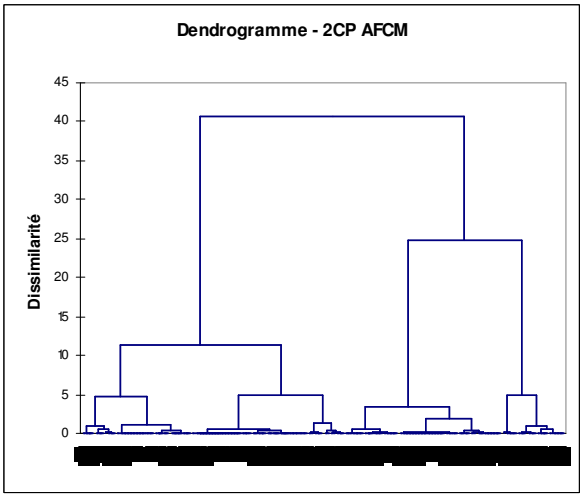






V.4.3. Application à la classification de données qualitatives

Reprenons les données sur les exploitations, et considérons les variables qualitatives analysées par l'AFCM. La classification peut se faire directement avec les composantes principales de l'AFCM. C'est la méthode que nous utiliserons le plus souvent. Pour travailler directement avec les variables qualitatives d'origine, il faut soit transformer le tableau en tableau disjonctif complet, soit disposer d'un logiciel qui accepte les données de type qualitatif directement, ce qui n'est pas le cas de beaucoup de logiciels.



V.5. Les méthodes non hiérarchiques

Ce sont des méthodes nécessitant beaucoup moins de calculs que les méthodes hiérarchiques. En conséquence, les tailles limites des fichiers pouvant être traités sont beaucoup plus élevées pour les méthodes non hiérarchiques que pour les méthodes hiérarchiques.

Le principe est de fixer a priori le nombre de classes k. On fixe aussi une première partition, puis, par itération, on crée une nouvelle partition qui paraît être meilleure que la précédente. On arrête les itérations quand on pense qu'on ne peut pas obtenir bien "meilleur". Présentons la méthode des centres mobiles. L'ensemble des étapes est comme suit :

1. choisir k = nombre de classes à créer;
2. choisir un ensemble de k points dans l'espace des points à classer (souvent, ce sont des points choisis dans l'ensemble des points); ces k points sont appelés "centres initiaux";
3. affecter chaque point au "centre" le plus près; on a ainsi formé k classes;
4. calculer les centres des classes obtenues en 3.; on a ainsi un nouvel ensemble de k centres;
5. tant que les centres de deux itérations successives sont distants de plus de ε, et tant que le nombre d'itérations est inférieur à nmax, revenir en 3., et incrémenter le compteur d'itérations de 1.

Pour mettre en pratique cette méthode, il faut fixer a priori le nombre de classes, la valeur de ε, et le nombre d'itérations maximum nmax. Les valeurs de ε et de nmax n'ont que peu d'influence sur le résultat, car souvent les centres ne bougent presque pas voire pas du tout à partir d'un certain nombre d'itérations, et le nombre maximum d'itérations est beaucoup plus grand que le nombre d'itérations nécessaire pour atteindre la convergence. Ces valeurs sont plutôt des sécurités pour les cas extrêmes.
C'est le choix du nombre de classes qui est le plus crucial, car il déterminera le nombre de profils émergents de l'ensemble des individus. Nous discutons de ce choix dans le paragraphe qui suit.

Remarque. Le résultat d'une telle méthode dépend du choix des "centres initiaux". Ce choix est souvent semi-aléatoire par défaut dans les logiciels, de telle manière que l'exécution de la méthode plusieurs fois de suite, avec les mêmes paramètres et les mêmes données peut conduire à des partitions différentes. Il est même conseillé d'essayer plusieurs fois pour retenir la partition d'inertie intra la plus petite.

D'autres noms paraissent dans la littérature (et dans les logiciels) pour nommer la méthode des centres mobiles : nuées dynamiques, k-means. Ces autres dénominations ont été données lors de généralisations de la méthode des centres mobiles en remplaçant les "points centres" par des ensembles de points centraux, ou nuées.
Une autre amélioration de la méthode des centres mobiles est de calculer le nouveau centre des classes après chaque réaffectation de point et non après la réaffectation de tous les points.

V.6. Choix d'une meilleure partition

V.6.1. Choix du nombre de classes

Le résultat de la classification hiérarchique propose un ensemble de n partitions, allant de 1 à n classes. Se pose alors le problème du choix du nombre de classes k. Le problème se pose différemment dans le cas non hiérarchique puisqu'il faut le faire a priori. Mais ce choix a priori n'est pas simple non plus.
On se fixera comme contrainte de choisir un nombre de classes minimum, pour une inertie intra minimum. Or, dans un ensemble de partitions hiérarchisées, l'inertie intra est une fonction décroissante du nombre de classes. Il faut donc faire un compromis entre une inertie intra pas trop grande pour un nombre de classes pas trop grand non plus. D'autres éléments peuvent entrer en jeu, comme par exemple l'exigence du spécialiste des données qui veut absolument créer tel nombre de profils d'individus.
En classification hiérarchique, l'analyse de l'arbre, et éventuellement de l'historique des itérations, permet de faire ce compromis. On choisira un nombre de classes k tel que le passage de k à k+1 classes ne diminue plus "significativement" l'inertie intra.

Exercice. Faire un choix du nombre de classes à partir des 4 dendrogrammes obtenus précédemment.

Quand on ne dispose que de résultats de méthodes non hiérarchiques, on peut essayer plusieurs nombres de classes, et voir comment évolue l'inertie intra en fonction du nombre de classes. On raisonnera alors comme avec la méthode hiérarchique.

Certains logiciels proposent des troncatures automatiques (SPAD, XLSTAT).

V.6.2. Choix d'une partition parmi plusieurs partitions à même nombre de classes

Quand le choix du nombre de classes est fait, il reste à comparer plusieurs partitions dont le nombre de classes est le même. Nous allons exposer deux méthodes de choix.

La méthode comparant les effectifs des classes. Cette méthode est surtout utilisée quand les sorties logiciel sont trop pauvres et qu'on n'a pas de possibilités de comparaison des inerties intra. Elle sert aussi à départager deux partitions qui auraient la même inertie intra.
Avec les variables indiquant les deux partitions, on dresse un tableau de contingence, et on retient la partition dont les écarts entre les effectifs des classes sont les plus faibles.

Exercice. Effectuer ce choix de partition entre les partitions obtenues par les méthodes hiérarchique et non hiérarchique sur les exploitations, avec les données quantitatives.

Effectifs observés (ClasseCAH / ClasseND) :					
	ClasseND-1	ClasseND-2	ClasseND-3	ClasseND-4	Total
ClasseCAH-1	87	18	0	0	105
ClasseCAH-2	0	37	1	0	38
ClasseCAH-3	0	0	4	0	4
ClasseCAH-4	0	0	0	7	7
Total	87	55	5	7	154

La méthode basée sur la comparaison des inerties intra. Le choix se portera sur la partition d'inertie intra minimum, bien entendu. On ne peut comparer que des partitions à même nombre de classes, vue la diminution systématique de l'inertie intra quand on augmente le nombre de classes.

Exercice. Effectuer ce choix de partition entre les partitions précédentes.

Décomposition de la variance pour la classification optimale (CAH, 4classes) :		Décomposition de la variance pour la classification optimale (ND, 4 classes):	
Intra-classe	1,731	Intra-classe	1,623
Inter-classes	3,306	Inter-classes	3,414
Totale	5,037	Totale	5,037

V.7. Description des classes d'une partition

La description se fait d'abord par comparaison des statistiques par classes aux statistiques sur les variables de départ. Cette comparaison passe par les méthodes d'analyse bivariée comparant la variable de partitionnement et les variables de départ. La variable de partitionnement est assimilée à une variable qualitative. Le choix des analyses dépend de la nature des variables de départ.

V.7.1 Description des classes d'une partition : cas de données initiales quantitatives

L'analyse des moyennes. On est en présence de p variables quantitatives et d'une variable qualitative. La première étude statistique sera le calcul de la moyenne des variables de départ pour chaque classe de la partition.

Exemple sur les exploitations. On effectue ce calcul des moyennes pour la partition en 4 classes obtenue par la méthode non hiérarchique.

moyennes	an_instal	age_chef	sau_totale	sau_vign>3ans	sau_vign<3ans	salariés	effectif
classeND-1	1997,828	34,414	33,853	22,634	3,410	0,394	87
classeND-2	1982,527	48,764	49,140	25,996	3,327	0,645	55
classeND-3	1988,600	47,000	54,200	6,300	46,800	1,350	5
classeND-4	1989,286	40,286	153,571	89,643	7,500	7,429	7
Total	*1991,675*	*40,214*	*45,415*	*26,350*	*4,975*	*0,834*	*154*

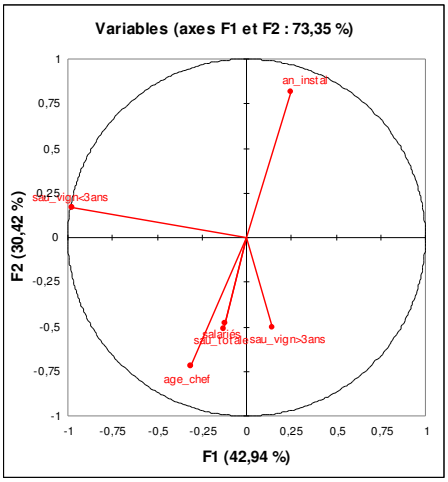
On peut simplifier la lecture de ce tableau en remplaçant les valeurs des moyennes par des statistiques d'ordre.

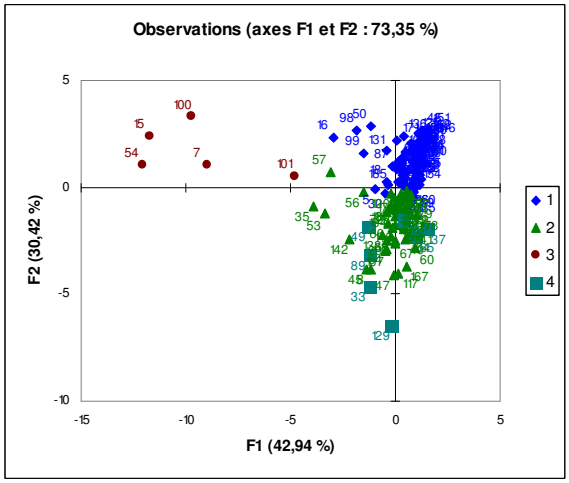
moyennes	an_instal	age_chef	sau_totale	sau_vign>3ans	sau_vign<3ans	salariés
classeND-1	++	--	--	-	-	--
classeND-2	--	++	-	+	--	-
classeND-3	-	+	+	--	++	+
classeND-4	+	-	++	++	+	++

Ainsi, la description des classes est facilitée.
Classe 1. Celle des exploitations installées le plus récemment, chefs d'exploitation les plus jeunes, nombre de salariés les plus faibles, les plus petites sau totales, et presque les plus petites sau en vignes, que ce soit de moins ou de plus de 3 ans.
Classe 2. Les plus anciennes exploitations, chefs les plus âgés, le moins de sau en vignes < 3 ans, peu de salariés, peu de sau totale, mais sau de vignes > 3 ans dans la moyenne.
Classe 3. Exploitations assez anciennes, chefs assez âgés, sau totale presque les plus élevées, mais sau en vignes > 3 ans les plus faibles alors que sau de vignes < 3 ans les plus élevées, nombre de salariés assez élevé.
Classe 4. Exploitations assez récentes, chef assez jeune (âge près de la moyenne), sau totale et sau en vignes > 3 ans les plus élevées, sau en vignes < 3 ans assez élevées, et nombre de salariés maximum.

L'Analyse Discriminante. Une autre analyse qu'on peut faire est l'analyse discriminante (AD). C'est une analyse qui peut être comprise comme une ACP. La différence est que, tandis que l'ACP calcule de nouvelles variables quantitatives, combinaisons linéaires des variables initiales, de variance maximum, et orthogonales entre elles, l'AD calcule aussi de nouvelles variables quantitatives, combinaisons linéaires des variables initiales, mais de variance inter maximum, la variance inter étant calculée avec pour variable qualitative la variable qui définit les classes.

Exemple sur les exploitations.





Exercice. Décrire les classes à l'aide de ces graphiques, qu'on utilisera comme ceux de l'ACP. Vérifier l'adéquation avec la description précédente basée sur les moyennes.

V.7.2 Description des classes d'une partition : cas de données initiales qualitatives

On utilisera les tableaux de contingence croisant les variables initiales et la variable qualitative issue de la partition.

Exemple des exploitations.

Effectifs observés (région / ClasseND-quali) :					
	ClasseND-quali-1	ClasseND-quali-2	ClasseND-quali-3	ClasseND-quali-4	Total
BRAM	1	13	0	0	14
CARC	11	9	11	1	32
MC	32	0	7	24	63
NH	3	0	36	6	45
Total	47	22	54	31	154

Effectifs théoriques (région / ClasseND-quali) :					
	ClasseND-quali-1	ClasseND-quali-2	ClasseND-quali-3	ClasseND-quali-4	Total
BRAM	4,273	2,000	4,909	2,818	14
CARC	9,766	4,571	11,221	6,442	32
MC	19,227	9,000	22,091	12,682	63
NH	13,734	6,429	15,779	9,058	45
Total	47	22	54	31	154

Significativité par case (région / ClasseND-quali) :				
	ClasseND-quali-1	ClasseND-quali-2	ClasseND-quali-3	ClasseND-quali-4
BRAM	<	>	<	<
CARC	>	>	<	<
MC	>	<	<	>
NH	<	<	>	<
p-value	< 0,0001			

Effectifs observés (statut / ClasseND-quali) :					
	ClasseND-quali-1	ClasseND-quali-2	ClasseND-quali-3	ClasseND-quali-4	Total
EARL	12	8	0	2	22
EI	29	10	38	14	91
GAEC	5	4	13	2	24
SCEA	1	0	3	13	17
Total	47	22	54	31	154

Effectifs théoriques (statut / ClasseND-quali) :					
	ClasseND-quali-1	ClasseND-quali-2	ClasseND-quali-3	ClasseND-quali-4	Total
EARL	6,714	3,143	7,714	4,429	22
EI	27,773	13,000	31,909	18,318	91
GAEC	7,325	3,429	8,416	4,831	24
SCEA	5,188	2,429	5,961	3,422	17
Total	47	22	54	31	154

Significativité par case (statut / ClasseND-quali) :				
	ClasseND-quali-1	ClasseND-quali-2	ClasseND-quali-3	ClasseND-quali-4
EARL	>	>	<	<
EI	>	<	>	<
GAEC	<	>	>	<
SCEA	<	<	<	>
p-value	< 0,0001			

Effectifs observés (adhérent / ClasseND-quali) :					
	ClasseND-quali-1	ClasseND-quali-2	ClasseND-quali-3	ClasseND-quali-4	Total
A	41	22	36	10	109
P	6	0	18	21	45
Total	47	22	54	31	154

Effectifs théoriques (adhérent / ClasseND-quali) :					
	ClasseND-quali-1	ClasseND-quali-2	ClasseND-quali-3	ClasseND-quali-4	Total
A	33,266	15,571	38,221	21,942	109
P	13,734	6,429	15,779	9,058	45
Total	47	22	54	31	154

Significativité par case (adhérent / ClasseND-quali) :				
	ClasseND-quali-1	ClasseND-quali-2	ClasseND-quali-3	ClasseND-quali-4
A	>	>	<	<
P	<	<	>	>
p-value	< 0,0001			

Effectifs observés (valorisation / ClasseND-quali) :					
	ClasseND-quali-1	ClasseND-quali-2	ClasseND-quali-3	ClasseND-quali-4	Total
Mixte	7	2	0	0	9
coop	35	19	47	15	116
part	5	1	7	16	29
Total	47	22	54	31	154

Effectifs théoriques (valorisation / ClasseND-quali) :					
	ClasseND-quali-1	ClasseND-quali-2	ClasseND-quali-3	ClasseND-quali-4	Total
Mixte	2,747	1,286	3,156	1,812	9
coop	35,403	16,571	40,675	23,351	116
part	8,851	4,143	10,169	5,838	29
Total	47	22	54	31	154

Significativité par case (valorisation / ClasseND-quali) :				
	ClasseND-quali-1	ClasseND-quali-2	ClasseND-quali-3	ClasseND-quali-4
Mixte	>	>	<	<
coop	<	>	>	<
part	<	<	<	>
p-value	< 0,0001			

Effectifs observés (vin / ClasseND-quali) :					
	ClasseND-quali-1	ClasseND-quali-2	ClasseND-quali-3	ClasseND-quali-4	Total
AO-VDQS	0	10	0	0	10
AOC	43	2	2	29	76
VDP	4	10	52	2	68
Total	47	22	54	31	154

Effectifs théoriques (vin / ClasseND-quali) :					
	ClasseND-quali-1	ClasseND-quali-2	ClasseND-quali-3	ClasseND-quali-4	Total
AO-VDQS	3,052	1,429	3,506	2,013	10
AOC	23,195	10,857	26,649	15,299	76
VDP	20,753	9,714	23,844	13,688	68
Total	47	22	54	31	154

Significativité par case (vin / ClasseND-quali) :				
	ClasseND-quali-1	ClasseND-quali-2	ClasseND-quali-3	ClasseND-quali-4
AO-VDQS	<	>	<	<
AOC	>	<	<	>
VDP	<	>	>	<
p-value	< 0,0001			

Exercice. Décrire les classes à l'aide de ces tableaux.