

4 Analyse des Correspondances Multiples

4.1 Introduction

L'analyse factorielle des correspondances, vue dans le paragraphe précédent, s'applique à des situations où les individus statistiques sont décrits par *deux* variables nominales. Mais il est fréquent que l'on dispose d'individus décrits par *plusieurs* (deux ou plus) variables nominales ou ordinales. C'est notamment le cas lorsque nos données sont les résultats d'une enquête basée sur des questions fermées à choix multiples. Une extension de l'AFC à ces situations a donc été proposée. Elle est généralement appelée Analyse des Correspondances Multiples ou ACM.

Nous nous plaçons donc dans la situation où nous disposons de N individus statistiques, décrits par q variables nominales ou ordinales X_1, X_2, \dots, X_q . L'ACM vise à mettre en évidence :

- les relations entre les modalités des différentes variables ;
- éventuellement, les relations entre individus statistiques ;
- les relations entre les variables, telles qu'elles apparaissent à partir des relations entre modalités.

4.2 Exemple

4.2.1 Enoncé

L'exemple qui suit est extrait de [Crucianu]. Il s'agit d'une partie des données issues de l'enquête "Les étudiants et la ville" effectuée en 2001 par des étudiants de sociologie sous la direction de S. Denèfle à l'Université François Rabelais de Tours.

L'analyse porte sur cinq questions en rapport avec le logement étudiant. L'ensemble des individus statistiques est ici un échantillon de 383 étudiants. Les questions sont les suivantes :

Question	N°	Réponses possibles	Poids (%)	Abréviation
Habitez-vous (variable "mode d'occupation")	1	seul	48,30%	Seul
	2	colocataires	13,84%	Coloc
	3	en couple	13,05%	Couple
	4	avec les parents	23,50%	Parents
	5	non réponse	1,31%	NR1
Quel type d'habitation occupez-vous ? (variable "type d'habitation")	6	cité universitaire	10,70%	Cité
	7	studio	28,20%	Studio
	8	appartement	30,29%	Appart
	9	chambre chez un particulier	5,22%	Chambre
	10	autre	19,84%	Autre
	11	non réponse	5,74%	NR2
Si vous vivez en dehors du foyer familial, depuis combien de temps ? (variable "ancienneté")	12	moins de 1 an	20,89%	< 1 an
	13	1 à 3 ans	24,80%	1-3 ans
	14	plus de 3 ans	28,72%	> 3 ans
	15	non applicable	24,80%	NA
	16	non réponse	0,78%	NR3
A quelle distance approximative de la Fac vivez-vous ? (variable "éloignement")	17	moins de 1 km	26,89%	< 1 km
	18	1 à 5 km	49,87%	1 à 5 km
	19	plus de 5 km	20,89%	> 5 km
	20	non réponse	2,35%	NR4

Quelle est la superficie de votre logement ? (variable "superficie")	21	moins de 10 m ²	9,14%	< 10 m ²
	22	10 à 20 m ²	17,75%	10 à 20 m ²
	23	20 à 30 m ²	24,80%	20 à 30 m ²
	24	plus de 30 m ²	39,16%	> 30 m ²
	25	non réponse	9,14%	NR5

4.2.2 Différentes représentations des données recueillies

Les données recueillies peuvent être représentées, de façon classique, à l'aide d'un tableau protocole ou d'un tableau d'effectifs. Cependant, deux autres représentations sont également utilisées : le tableau disjonctif complet (TDC) et le tableau de Burt (TdB).

4.2.2.1 Tableau protocole et tableau d'effectifs

Les données recueillies peuvent être représentées, de façon classique, à l'aide d'un tableau protocole ou d'un tableau d'effectifs :

Tableau protocole

Sujet	Mode d'occupation	Type d'habitation	Ancienneté	Eloignement	Superficie
S1	seul	cité	< 1 an	< 1 km	< 10 m2
S2	coloc	appart	1 à 3 ans	1 à 5 km	20 à 30 m2
...	...				

Tableau d'effectifs

Mode d'occupation	Type d'habitation	Ancienneté	Eloignement	Superficie	Effectif
seul	cité	< 1 an	< 1 km	< 10 m2	7
seul	cité	< 1 an	< 1 km	10 à 20 m2	2
...					

4.2.2.2 Tableau disjonctif complet (TDC)

Le tableau disjonctif complet comporte une colonne pour chaque modalité des variables étudiées, et une ligne pour chaque individu statistique. Les cellules du tableau contiennent 1 ou 0 selon que l'individu considéré présente la modalité correspondante ou non :

	Mode d'occupation					Type habitation						Ancienneté					Eloignement				Superficie				
	Seu 1	Col oc	Cou ple	Par ents	NR 1	Cité	Stu dio	Ap part	Cha mbr e	Aut re	NR 2	<= 1 an	1-3 ans	> 3 ans	NA	NR 3	- de 1k m	1 à 5 km	+ de 5k m	NR 4	- de 10 m2	10 à 20 m2	20 à 30 m2	+ de 30 m2	NR 5
i1	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0
i2	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0
i3	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0
i4	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1
i5	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
...																							

4.2.2.3 Tableau de Burt (TdB)

Le tableau de Burt comporte une ligne et une colonne pour chaque modalité des variables étudiées. Chaque cellule du tableau indique le nombre d'individus statistiques qui possèdent en même temps la

modalité ligne et la modalité colonne correspondantes. Pour l'exemple étudié, le tableau de Burt est le suivant :

	Seu l	Col oc	Co upl e	Par ents	NR 1	Cit é	Stu dio	Ap part	Cha mbr e	Aut re	NR 2	<= 1 an	1-3 ans	> 3 ans	NA	NR 3	- de 1k m	1 à 5 km	+ de 5 km	NR 4	- de 10 m2	10 à 20 m2	20 à 30 m2	+ de 30 m2	NR 5
Seul	185	0	0	0	0	34	90	40	13	3	5	61	60	59	5	0	70	101	11	3	32	61	71	21	0
Colo	0	53	0	0	0	5	6	32	2	3	5	13	18	21	1	0	13	33	6	1	1	4	8	40	0
Coup	0	0	50	0	0	2	10	34	0	3	1	5	14	28	3	0	15	23	12	0	2	2	14	32	0
Par	0	0	0	90	0	0	1	9	5	67	8	1	2	0	86	1	4	32	51	3	0	1	1	55	33
NR1	0	0	0	0	5	0	1	1	0	0	3	0	1	2	0	2	1	2	0	2	0	0	1	2	2
Cité	34	5	2	0	0	41	0	0	0	0	0	17	13	9	2	0	15	23	3	0	27	9	1	4	0
Stud	90	6	10	1	1	0	108	0	0	0	0	29	33	45	1	0	41	61	6	0	1	33	57	17	0
App	40	32	34	9	1	0	0	116	0	0	0	23	35	47	11	0	37	62	17	0	1	10	29	74	2
Cha	13	2	0	5	0	0	0	0	20	0	0	6	6	3	5	0	6	10	4	0	4	7	5	4	0
Autr	3	3	3	67	0	0	0	0	0	76	0	2	4	4	65	1	2	29	42	3	0	1	1	50	24
NR2	5	5	1	8	3	0	0	0	0	0	22	3	4	2	11	2	2	6	8	6	2	8	2	1	9
- de1	61	13	5	1	0	17	29	23	6	2	3	80	0	0	0	0	30	44	5	1	14	26	24	16	0
1-3	60	18	14	2	1	13	33	35	6	4	4	0	95	0	0	0	25	60	10	0	11	22	28	32	2
+de3	59	21	28	0	2	9	45	47	3	4	2	0	0	110	0	0	43	53	14	0	10	14	41	45	0
NA	5	1	3	86	0	2	1	11	5	65	11	0	0	0	95	0	5	33	51	6	0	6	2	56	31
NR3	0	0	0	1	2	0	0	0	0	1	2	0	0	0	0	3	0	1	0	2	0	0	0	1	2
- 1k	70	13	15	4	1	15	41	37	6	2	2	30	25	43	5	0	103	0	0	0	12	26	38	26	1
1 à 5	101	33	23	32	2	23	61	62	10	29	6	44	60	53	33	1	0	191	0	0	20	35	47	82	7
+ 5k	11	6	12	51	0	3	6	17	4	42	8	5	10	14	51	0	0	0	80	0	2	5	9	41	23
NR4	3	1	0	3	2	0	0	0	0	3	6	1	0	0	6	2	0	0	0	9	1	2	1	1	4
- 10	32	1	2	0	0	27	1	1	4	0	2	14	11	10	0	0	12	20	2	1	35	0	0	0	0
10-20	61	4	2	1	0	9	33	10	7	1	8	26	22	14	6	0	26	35	5	2	0	68	0	0	0
20-30	71	8	14	1	1	1	57	29	5	1	2	24	28	41	2	0	38	47	9	1	0	0	95	0	0
30+	21	40	32	55	2	4	17	74	4	50	1	16	32	45	56	1	26	82	41	1	0	0	0	150	0
NR5	0	0	0	33	2	0	0	2	0	24	9	0	2	0	31	2	1	7	23	4	0	0	0	0	35

Lecture de ce tableau :

- parmi les 383 étudiants interrogés, 185 logent seuls ;
- parmi les 383 étudiants interrogés, 34 logent seuls, en cité universitaire ;
- etc.

Le tableau de Burt possède de nombreuses propriétés remarquables :

- Le tableau est symétrique : $n_{ij} = n_{ji}$;
- Les encadrés situés le long de la diagonale principale (du haut à gauche vers le bas à droite) donnent les effectifs correspondant à chaque modalité ;
- Les autres encadrés sont les tableaux de contingence correspondant aux variables prises deux à deux ;
- La ligne j (ou la colonne j) du tableau est la somme des lignes du TDC correspondant aux individus qui possèdent la modalité j ;
- La somme des nombres situés sur une même ligne est égale au terme diagonal de la ligne multiplié par le nombre de variables ; propriété identique pour les colonnes ;
- La somme des nombres situés dans un encadré est égal à l'effectif total ;
- La somme de tous les nombres du tableau est égale à l'effectif total multiplié par le carré du nombre de variables.

Le tableau de Burt peut être vu comme une juxtaposition de tableaux de contingence. Il peut être obtenu facilement à partir du tableau disjonctif complet. En revanche, il n'existe pas de moyen simple permettant de recomposer le tableau disjonctif complet à partir du tableau de Burt.

4.2.3 L'analyse des correspondances multiples proprement dite

L'analyse peut être menée à partir du tableau disjonctif complet ou du tableau de Burt. Les deux méthodes conduisent à des résultats analogues (mais pas identiques). C'est essentiellement l'étude à partir du tableau de Burt que nous développerons, car elle seule est disponible dans Statistica.

D'un point de vue mathématique, le traitement opéré sur les données du tableau de Burt est identique à celui opéré sur un tableau de contingence lors d'une AFC. On obtiendra donc, comme lors d'une AFC :

- Des axes factoriels associés à des valeurs propres ;
- Pour chaque ligne (ou colonne) du tableau de Burt, des coordonnées, des contributions à la formation des axes et des qualités de représentation.

Cependant, la méthode diffère de l'AFC par divers aspects, et nous devons adapter notre grille d'interprétation.

Le tableau de Burt étant symétrique, les profils lignes et les profils colonnes sont identiques. Il en est de même des coordonnées des individus-lignes et des individus-colonnes. Nous nous intéresserons donc uniquement aux individus-lignes (par exemple).

La méthode ne fournit pas de coefficient analogue au Φ^2 . En effet, l'inertie totale du nuage des modalités est déterminée uniquement par le nombre total de modalités p et le nombre de variables q . Plus

précisément, l'inertie totale du nuage des modalités vaut : $I = \sum_{j=1}^p I(X_j) = \frac{p-q}{q}$. Cette valeur ne fournit

donc aucune information sur la présence ou l'absence de dépendances entre les variables.

Ainsi, dans notre exemple :

$$I = \frac{25-5}{5} = 4.$$

L'inertie totale peut être exprimée comme la somme des inerties de chacune des variables. Mais l'inertie

de la variable X_j est donnée par : $I(X_j) = \frac{p_j-1}{q}$, où p_j est le nombre de modalités de la variable X_j . Par

exemple, pour la première variable :

$$I(X_1) = \frac{5-1}{5} = 0,8$$

Autrement dit, l'influence d'une variable dépend seulement du nombre de ses modalités. Pour éviter que certaines variables prennent une importance excessive, ou au contraire soient peu présentes dans l'analyse, il faut donc éviter des différences trop marquées entre les nombres de modalités des variables à analyser.

L'inertie due à une modalité est donnée par : $I(X_j = a_k) = \frac{n - n_k}{nq}$ où n_k est l'effectif de la modalité a_k de la

variable X_j . Ainsi, plus une modalité est rare, plus son influence globale est élevée. Ainsi, dans l'exemple que nous traitons, il serait intéressant de reprendre l'étude en supprimant les non-réponses aux 4 premières questions, ou en les regroupant avec des réponses "proches".

Les profils-lignes (qui sont ici identiques aux profils-colonnes) ne sont pas directement interprétables. Le nuage des modalités a cependant une propriété intéressante : le centre de gravité des différentes modalités d'une même variable est l'origine des axes.

4.2.3.1 Valeurs propres

L'analyse du tableau de Burt produit au plus $p-q$ valeurs propres non nulles. La décroissance de ces valeurs propres est beaucoup moins rapide que dans le cas de l'AFC, et il est difficile d'énoncer un critère relatif au nombre d'axes factoriels à conserver. De fait, il arrive fréquemment qu'on se limite à l'étude du premier plan factoriel.

Nombre de Dims.	Valeurs Propres et Inertie de toutes les Dimensions (Etudiants-ville.sta)				
	Inertie Totale = 4,0000				
	ValSing.	ValProp.	%age Inertie	%age Cumulé	Chi ²

1	0,8505	0,7233	18,08	18,08	2287,23
2	0,6567	0,4313	10,78	28,86	1363,94
3	0,6213	0,3860	9,65	38,51	1220,72
4	0,5713	0,3264	8,16	46,67	1032,18
5	0,5023	0,2523	6,31	52,98	797,87
6	0,4659	0,2171	5,43	58,41	686,46
7	0,4473	0,2001	5,00	63,41	632,69
8	0,4442	0,1974	4,93	68,34	624,11
9	0,4293	0,1843	4,61	72,95	582,70
10	0,4111	0,1690	4,23	77,18	534,56
11	0,4086	0,1670	4,17	81,35	528,09
12	0,3799	0,1443	3,61	84,96	456,46
13	0,3519	0,1238	3,10	88,06	391,59
14	0,3293	0,1084	2,71	90,77	342,94
15	0,3237	0,1048	2,62	93,39	331,34
16	0,2926	0,0856	2,14	95,53	270,79
17	0,2719	0,0739	1,85	97,38	233,74
18	0,2328	0,0542	1,35	98,73	171,32
19	0,1948	0,0380	0,95	99,68	120,03
20	0,1134	0,0129	0,32	100,00	40,69

4.2.3.2 Résultats relatifs aux modalités

NomLigne	Coordonnées Colonne et Contributions à l'Inertie (Etudiants-ville.sta)									
	Inertie Totale = 4,0000									
	Ligne Numéro	Coord. Dim.1	Coord. Dim.2	Masse	Qualité	Inertie Relative	Inertie Dim.1	Cosinus2 Dim.1	Inertie Dim.2	Cosinus2 Dim.2
MODE:Seul	1	-0,6890	0,3383	0,0966	0,5505	0,0258	0,0634	0,4436	0,0256	0,1069
MODE:Coloc	2	-0,2283	-0,5691	0,0277	0,0604	0,0431	0,0020	0,0084	0,0208	0,0520
MODE:Couple	3	-0,2247	-0,9016	0,0261	0,1296	0,0435	0,0018	0,0076	0,0492	0,1221
MODE:Parents	4	1,6354	-0,1130	0,0470	0,8255	0,0383	0,1738	0,8215	0,0014	0,0039
MODE:NR1	5	0,7226	4,5647	0,0026	0,2825	0,0493	0,0019	0,0069	0,1261	0,2756
TYPE:Cité	6	-0,7458	0,6987	0,0214	0,1252	0,0446	0,0165	0,0667	0,0242	0,0585
TYPE:Studio	7	-0,7181	0,1428	0,0564	0,2105	0,0359	0,0402	0,2025	0,0027	0,0080
TYPE:Appart	8	-0,2417	-0,6967	0,0606	0,2362	0,0349	0,0049	0,0254	0,0682	0,2109
TYPE:Chambre	9	-0,2307	0,2778	0,0104	0,0072	0,0474	0,0008	0,0029	0,0019	0,0043
TYPE:Autre	10	1,5929	-0,2506	0,0397	0,6437	0,0401	0,1392	0,6281	0,0058	0,0156
TYPE:NR2	11	0,8965	2,2834	0,0115	0,3667	0,0471	0,0128	0,0490	0,1389	0,3177
ANCIENN:< 1 an	12	-0,6530	0,3547	0,0418	0,1458	0,0396	0,0246	0,1126	0,0122	0,0332
ANCIENN:1-3 ans	13	-0,4761	-0,0368	0,0496	0,0752	0,0376	0,0155	0,0748	0,0002	0,0004
ANCIENN:> 3ans	14	-0,4892	-0,3545	0,0574	0,1471	0,0356	0,0190	0,0964	0,0167	0,0506
ANCIENN:NA	15	1,5386	-0,0532	0,0496	0,7818	0,0376	0,1624	0,7809	0,0003	0,0009
ANCIENN:NR3	16	1,7019	6,3875	0,0016	0,3450	0,0496	0,0063	0,0229	0,1482	0,3221
ELOIGN:< 1km	17	-0,6525	0,0493	0,0538	0,1575	0,0366	0,0317	0,1566	0,0003	0,0009
ELOIGN:1 à 5 km	18	-0,2123	-0,0983	0,0997	0,0545	0,0251	0,0062	0,0448	0,0022	0,0096
ELOIGN:> 5 km	19	1,1946	-0,2628	0,0418	0,3950	0,0396	0,0824	0,3768	0,0067	0,0182
ELOIGN:NR4	20	1,3552	3,8589	0,0047	0,4025	0,0488	0,0119	0,0442	0,1623	0,3583
SUPERF:< 10 m2	21	-0,7705	0,8750	0,0183	0,1367	0,0454	0,0150	0,0597	0,0324	0,0770
SUPERF:10 à 20 m2	22	-0,6083	0,6080	0,0355	0,1597	0,0411	0,0182	0,0799	0,0304	0,0798
SUPERF:20 à 30 m2	23	-0,6704	-0,0253	0,0496	0,1485	0,0376	0,0308	0,1483	0,0001	0,0002
SUPERF:> 30 m2	24	0,4160	-0,6849	0,0783	0,4134	0,0304	0,0187	0,1114	0,0852	0,3019
SUPERF:NR5	25	1,9890	0,9474	0,0183	0,4881	0,0454	0,1000	0,3979	0,0380	0,0903

On retrouve dans ce tableau l'inertie relative de chaque variable, comme somme des inerties relatives des modalités qui la compose. Par exemple, pour la première variable :

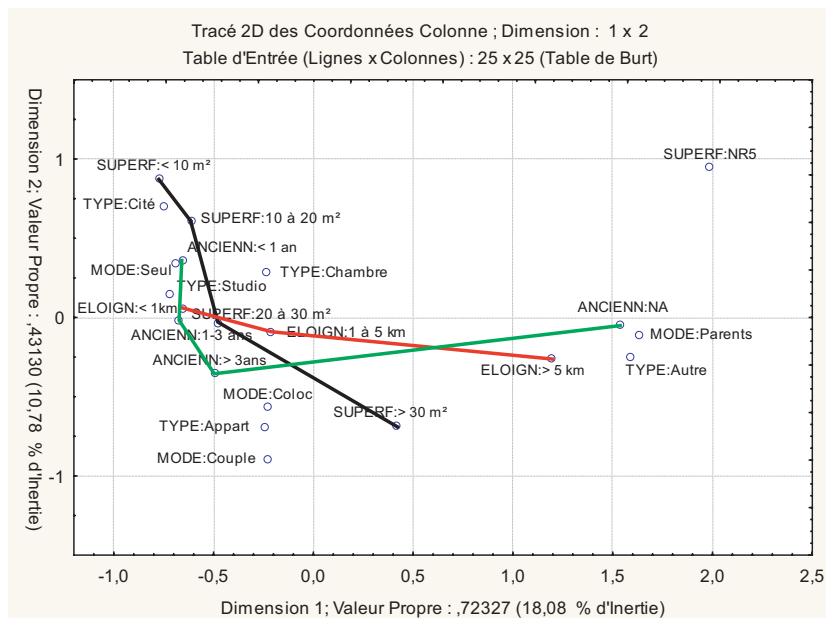
$$\frac{I(X_1)}{I} = \frac{0,8}{4} = 0,2 = 0,0258 + 0,0431 + 0,0435 + 0,0383 + 0,0493$$

L'interprétation utilisera essentiellement les modalités qui ont les meilleures qualités de représentation selon chacun des deux axes factoriels (colonnes Cosinus²) ou dans le premier plan factoriel (colonne "Qualité"). Mais, il faudra retenir les modalités jusqu'à un seuil assez bas, 0,25 par exemple.

Enfin, l'inertie relative par rapport à chaque axe permettra de retenir les modalités qui ont le plus fortement contribué à la formation de cet axe. On pourra par exemple, retenir les modalités dont l'inertie relative par rapport à un axe dépasse 0,04.

4.2.3.3 Résultats graphiques

Le graphique suivant représente les modalités sur le premier plan factoriel. Il a été obtenu sous Statistica. La partie visible des axes a été ajustée afin de faire disparaître la plupart des modalités "non réponse", très excentrées et de peu d'intérêt. Il a ensuite été retravaillé et mis en forme à l'aide d'Adobe Illustrator.



4.2.3.4 Quelques règles d'interprétation supplémentaires

On cherchera d'une part à interpréter les oppositions entre modalités (ou entre groupes d'individus, si l'étude porte sur le TDC), et d'autre part à interpréter les proximités entre modalités.

Ainsi, dans l'exemple fourni, on note de fortes oppositions entre d'un côté : "avec les parents", "plus de 5 km", "autre" et le l'autre côté : "seul", "moins de 1 km" et les autres types d'habitation. Cette opposition résume assez bien le premier axe factoriel.

Le deuxième axe marque une opposition entre "cité universitaire", "moins de 10 m²" d'une part et "appartement", "en couple" d'autre part.

L'interprétation des proximités entre les modalités devra tenir compte de la remarque suivante :

- Si deux modalités *d'une même variable* sont proches, cela signifie que les individus qui possèdent l'une des modalités et ceux qui possèdent l'autre sont globalement similaires *du point de vue des autres variables* ;
- Si deux modalités *de deux variables différentes* sont proches, cela peut signifier que ce sont globalement les mêmes individus qui possèdent l'une et l'autre.

4.3 L'ACM avec Statistica

4.3.1 Présentation des données étudiées

Référence : Les données présentées ici sont accessibles sur le site personnel de Gilles Hunault, à l'adresse :

Ce dossier contient des données relatives à une enquête réalisée dans des supermarchés angevins et parisiens entre 1996 et 1998 dans le but de connaître l'avis de consommateurs quant aux produits biologiques et aux produits diététiques.

La structure des données est la suivante :

- 1 - Matricule anonyme de la personne interrogée
- 2 - Connaissez-vous les produits biologiques ?
 - 0 - non réponse
 - 1 - oui
 - 2 - non
- 3 - Y a t-il une différence entre produit biologique et produit diététique ?
 - 0 - non réponse
 - 1 - oui
 - 2 - non
- 4 - Avez- vous déjà consommé des produits biologiques ?
 - 1 - non jamais
 - 2 - oui une seule fois
 - 3 - oui rarement
 - 4 - oui de temps en temps
 - 5 - oui plusieurs fois par mois
 - 6 - oui plusieurs fois par semaine
 - 7 - ne se prononce pas
- 5 - Parmi les marques suivantes lesquelles connaissez-vous ?
 - 0 - non réponse
 - 1 - bio vivre
 - 2 - bjorg
 - 3 - carrefour bio
 - 4 - la vie
 - 5 - vrai
 - 6 - prosain
 - 7 - favrichon
- 6 - Avez-vous déjà consommé des produits " La Vie " ?
 - 0 - non réponse
 - 1 - oui une fois
 - 2 - oui occasionnellement
 - 3 - oui régulièrement
 - 4 - non jamais
- 7 - Sexe de la personne
 - 1 - homme
 - 2 - femme
- 8 - Classe d'age
 - 1 - moins de 25 ans
 - 2 - entre 25 et 35 ans
 - 3 - entre 35 et 45 ans
 - 4 - entre 45 et 55 ans
 - 5 - entre 55 et 65 ans
 - 6 - plus de 65 ans
- 9 -Etat-civil
 - 0 - autre
 - 1 - marie

- 2 - célibataire
- 3 - divorcé
- 4 - en concubinage
- 5 - veuf
- 10 - Nombre d'enfants
 - 1 - 0 enfant
 - 2 - 1 enfants
 - 3 - 2 enfants
 - 4 - 3 enfant
 - 5 - plus de 3 enfants
- 11 - Situation professionnelle
 - 0 - non-réponse
 - 1 - agriculteur
 - 2 - artisan
 - 3 - cadre supérieur
 - 4 - cadre moyen
 - 5 - employé
 - 6 - ouvrier
 - 7 - retraité
 - 8 - autre
- 12 - Classe de revenus mensuels
 - 0 - non réponse
 - 1 - moins de 5 kF
 - 2 - entre 5 et 10 kF
 - 3 - entre 10 et 15 kF
 - 4 - entre 15 et 20 kF
 - 5 - plus de 20 kF
 - 6 - ne se prononce pas

L'échantillon interrogé comporte 419 observations. Sur le serveur de TD, deux feuilles de données Statistica rassemblent ces données : dans la première, PBIO.sta, les modalités sont représentées par leur code numérique ; dans la deuxième, PBIO-2.sta, des étiquettes de texte, représentant des abréviations des libellés de réponses, ont été introduits.

4.3.2 Exploration préalable des données

Avant de réaliser des analyses multivariées sur ces données, explorez-les en utilisant les menus Statistiques - Statistiques Élémentaires - Statistiques Descriptives et Statistiques - Statistiques Élémentaires - Tableaux et tris croisés. Vous pouvez ainsi obtenir des tableaux de recensement univariés tels que :

Table de fréquences : DIFF (PBIO-2.sta)				
	Effectif	Effectifs Cumulés	%age	%age Cumulé
NR	3	3	0,71599	0,7160
oui	328	331	78,28162	78,9976
non	88	419	21,00239	100,0000
VM	0	419	0,00000	100,0000

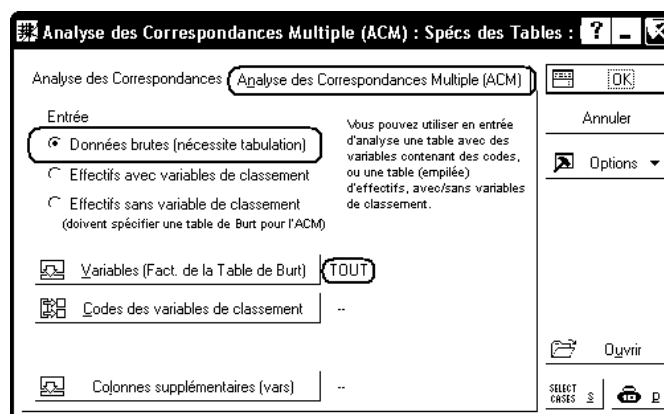
ou des tableaux de contingence tels que :

Table de Fréquences - Synthèse (PBIO-2.sta)								
Table :CONSOM(7) x AGE(6)								
	CONSOM	AGE	AGE	AGE	AGE	AGE	AGE	Totaux

		- de 25 ans	entre 25 et 35 ans	entre 35 et 45 ans	entre 45 et 55 ans	entre 55 et 65 ans	+ de 65 ans	Ligne
	jamais	15	39	19	21	9	8	111
	une fois	3	5	0	3	3	1	15
	rarement	16	25	17	19	4	8	89
	quelquefois	13	28	16	38	12	10	117
	souvent	2	4	3	10	2	3	24
	très souvent	0	11	13	8	9	8	49
	nr	3	4	2	2	2	1	14
	Ts Grpes	52	116	70	101	41	39	419

4.3.3 ACM sur les données étudiées

On peut essayer de réaliser une ACM avec, comme données actives, l'ensemble des données fournies. Pour cela, utilisez le menu Statistiques - Techniques exploratoires multivariées - Analyse des correspondances et sélectionnez ensuite l'item "Analyse des Correspondances Multiples". On voit que l'analyse peut être menée soit à partir d'un tableau protocole, soit à partir d'un tableau d'effectifs, soit à partir d'un tableau de Burt. Dans notre cas, les données se présentent sous forme d'un tableau protocole.

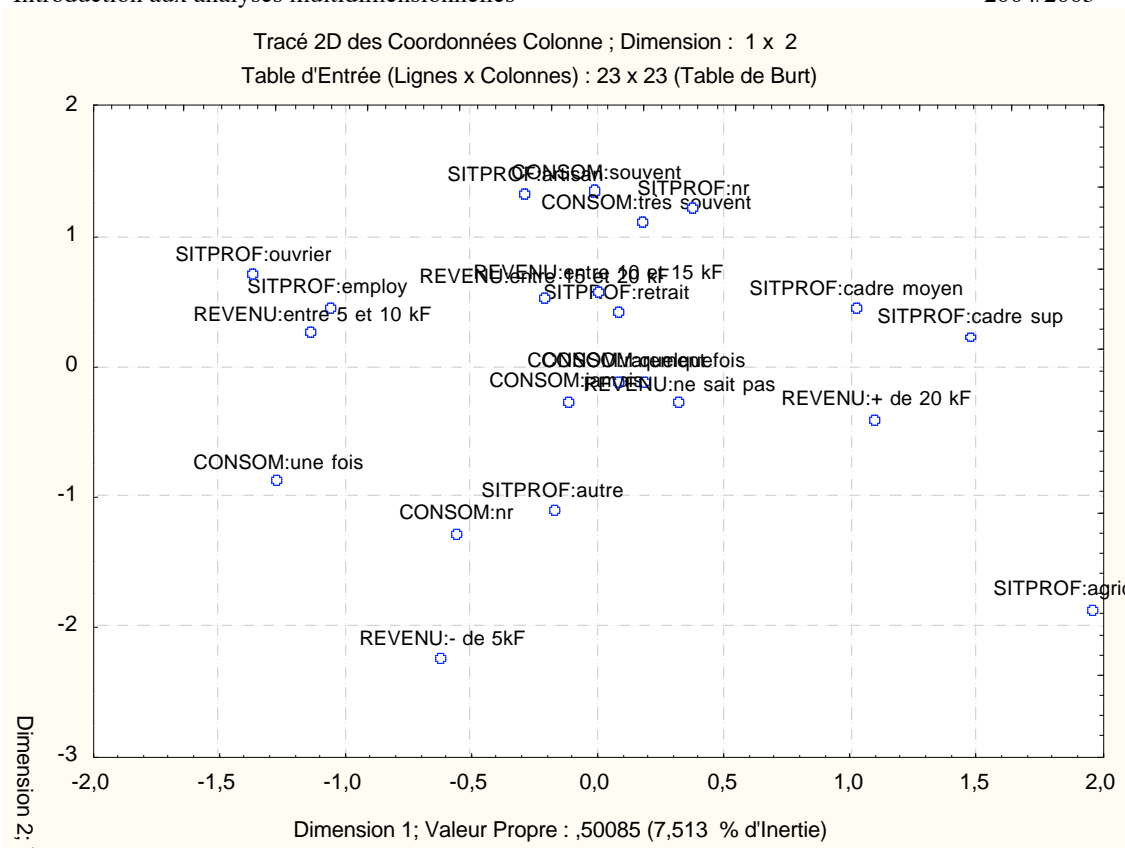


Cependant, les modalités sont alors trop nombreuses et les résultats sont difficilement interprétables. Par ailleurs, certains problèmes cités plus haut se rencontrent sur ces données :

- Certaines modalités ont des effectifs très faibles, et leur importance dans l'étude est surestimée;
- Le nombre de modalités de chaque variable varie de 3 à 9. Or, on sait qu'il vaut mieux utiliser des variables comportant à peu près le même nombre de modalités.

C'est pourquoi il est ici préférable de choisir 3 ou 4 variables, pour lesquelles nous pensons que l'étude conjointe présente un intérêt. Ce choix s'imposerait de lui-même si nous avions fixé un but bien déterminé à notre étude, ce qui n'est pas le cas ici. L'examen des tables de fréquences nous permet cependant de choisir des variables où les effectifs sont assez bien répartis entre les modalités, telles que, par exemple : CONSOM, SITPROF et REVENU.

En recadrant les échelles des axes (bouton droit sur l'axe, puis le menu local "Echelle...") de façon à éliminer certaines parties sans intérêt du dessin, on obtient le graphique suivant :



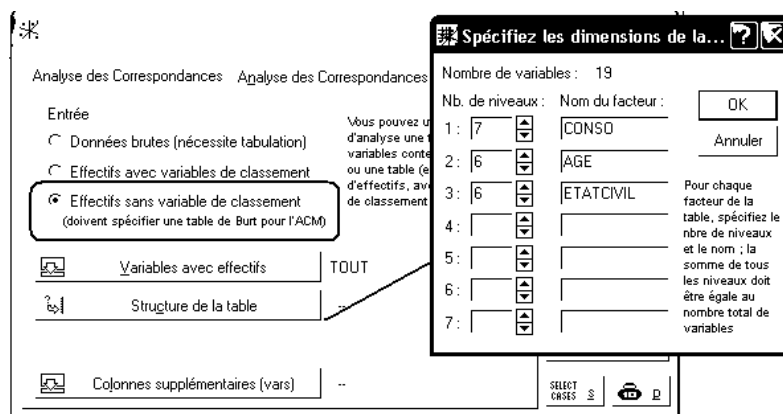
L'interprétation utilisera également les tableaux des coordonnées, qualités de représentation et contributions :

NomLigne	Coordonnées Colonne et Contributions à l'Inertie (PBIO-2.sta)									
	Inertie Totale = 6,6667									
	Ligne	Coord.	Coord.	Masse	Qualité	Inertie	Inertie	Cosinus2	Inertie	Cosinus2
	Numéro	Dim.1	Dim.2			Relative	Dim.1	Dim.1	Dim.2	Dim.2
CONSOM:jamais	1	-0,1147	-0,2745	0,0883	0,0319	0,0368	0,0023	0,0047	0,0146	0,0271
CONSOM:une fois	2	-1,2597	-0,8938	0,0119	0,0886	0,0482	0,0378	0,0589	0,0210	0,0297
CONSOM:rarement	3	0,0862	-0,1216	0,0708	0,0060	0,0394	0,0011	0,0020	0,0023	0,0040
CONSOM:quelquefois	4	0,1921	-0,1167	0,0931	0,0196	0,0360	0,0069	0,0143	0,0028	0,0053
CONSOM:souvent	5	-0,0048	1,3379	0,0191	0,1088	0,0471	0,0000	0,0000	0,0751	0,1088
CONSOM:très souvent	6	0,1899	1,1087	0,0390	0,1676	0,0442	0,0028	0,0048	0,1054	0,1628
CONSOM:nr	7	-0,5510	-1,2917	0,0111	0,0682	0,0483	0,0068	0,0105	0,0409	0,0577
SITPROF:nr	8	0,3839	1,2134	0,0008	0,0039	0,0499	0,0002	0,0004	0,0026	0,0035
SITPROF:agriculteur	9	1,9600	-1,8723	0,0008	0,0176	0,0499	0,0061	0,0092	0,0061	0,0084
SITPROF:artisan	10	-0,2851	1,3171	0,0072	0,0399	0,0489	0,0012	0,0018	0,0273	0,0381
SITPROF:cadre sup	11	1,4783	0,2187	0,0270	0,1972	0,0459	0,1180	0,1930	0,0028	0,0042
SITPROF:cadre moyen	12	1,0310	0,4370	0,0676	0,3191	0,0399	0,1435	0,2705	0,0284	0,0486
SITPROF:employé	13	-1,0596	0,4535	0,0835	0,4442	0,0375	0,1873	0,3755	0,0378	0,0688
SITPROF:ouvrier	14	-1,3569	0,7003	0,0064	0,0454	0,0490	0,0234	0,0358	0,0069	0,0095
SITPROF:retraité	15	0,0832	0,4072	0,0453	0,0272	0,0432	0,0006	0,0011	0,0165	0,0261
SITPROF:autre	16	-0,1707	-1,1109	0,0947	0,5011	0,0358	0,0055	0,0116	0,2569	0,4895
REVENU:nr	17	2,9320	4,2712	0,0008	0,0642	0,0499	0,0137	0,0206	0,0319	0,0436
REVENU:- de 5kF	18	-0,6142	-2,2576	0,0175	0,3034	0,0474	0,0132	0,0209	0,1961	0,2824
REVENU:entre 5 et 10 kF	19	-1,1327	0,2443	0,0788	0,4154	0,0382	0,2018	0,3969	0,0103	0,0185
REVENU:entre 10 et 15 kF	20	0,0009	0,5590	0,0628	0,0726	0,0406	0,0000	0,0000	0,0432	0,0726
REVENU:entre 15 et 20 kF	21	-0,2119	0,5275	0,0525	0,0604	0,0421	0,0047	0,0084	0,0321	0,0520
REVENU:+ de 20 kF	22	1,0996	-0,4079	0,0899	0,5080	0,0365	0,2170	0,4465	0,0329	0,0614
REVENU:ne sait pas	23	0,3174	-0,2993	0,0310	0,0195	0,0453	0,0062	0,0103	0,0061	0,0092

4.3.4 ACM à partir d'un tableau de Burt

Statistica permet également de réaliser une ACM à partir d'un tableau de Burt. Par exemple, le fichier de données PBIO-Burt.sta contient le tableau de Burt construit à partir des variables Consommation, Age et Etat-Civil. Chargez ce fichier et réalisez une ACM. La différence la plus importante par rapport à la situation précédente est la nécessité d'indiquer à Statistica les variables qui doivent être regroupées afin de

former un "facteur" de l'étude. Ici, "CONSOM" correspond aux 7 premières variables, "AGE" aux 6 suivantes et "ETAT CIVIL" aux 6 dernières :



Il faut noter que Statistica exige que le tableau fourni soit rigoureusement un tableau de Burt. Il n'est pas possible, par exemple, de supprimer les lignes correspondant aux non-réponses, car la cohérence du tableau n'est alors plus assurée.

4.3.5 Améliorer les graphiques fournis par Statistica

Les graphiques fournis par Statistica sont assez peu lisibles, notamment parce que les étiquettes des données se superposent. On peut, de diverses manières, améliorer cette situation.

- On s'efforcera de trouver des libellés courts (et sans caractères accentués) pour les variables et les modalités.
- Sous Statistica, on peut utiliser l'outil "zoom" pour éliminer certaines parties du graphique. Mais, il est souvent plus pratique de modifier l'échelle des axes (clic droit sur l'axe, puis le menu local "Echelle...").
- Enfin, on peut importer le graphique dans un autre logiciel, pour en modifier certains éléments. L'idéal est de disposer d'un outil de dessin vectoriel tel qu'Adobe Illustrator, mais la modification est aussi possible avec le module de dessin de Word.

Pour enregistrer le graphique dans un format exploitable par d'autres logiciels :

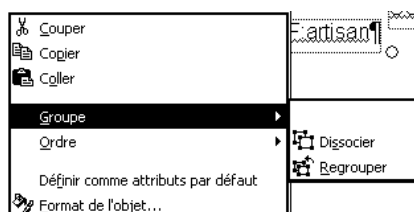
- Faites un clic droit sur le graphique (contenu dans un classeur, ou dans une fenêtre indépendante) ;
- Utilisez le menu "Enregistrez le graphique sous..." et choisissez le format d'enregistrement "Windows Metafile (wmf)" ou "Enhanced metafile (emf)".

Pour importer le graphique dans Word :

- Chargez Word, et ouvrez le document dans lequel le graphique doit s'insérer ;
- Placez le pointeur d'insertion à l'endroit voulu
- Utilisez le menu Insertion - Image - A partir d'un fichier...

Pour modifier le graphique dans Word :

- Double-cliquez sur le graphique pour l'ouvrir dans une fenêtre d'édition
- Utilisez les fonctions de zoom pour afficher la partie à modifier à une échelle suffisante
- Les étiquettes de texte sont des objets associés entre eux. Il faut donc sélectionner l'objet résultat de cette association, et utiliser (plusieurs fois) l'item de menu "Dissocier" avant de pouvoir faire les modifications voulues :



4.4 Comparaison entre ACM et AFC

Dans le cas où les individus statistiques étudiés sont décrits par deux variables nominales, on peut utiliser une AFC, aussi bien qu'une ACM pour explorer les données recueillies. La question se pose donc d'étudier les relations entre les résultats fournis par ces deux méthodes, dans cette situation.

Nous nous proposons de mener cette étude sur l'exemple "Nobel", déjà utilisé précédemment.

Pour réaliser l'AFC, nous prenons comme données de base le tableau protocole, le tableau d'effectifs ou le tableau de contingence :

PAYS	MEDE	PHYS	CHIM	LITT	SECO
USA	55	43	24	8	9
GB	19	20	21	6	2
RFA	11	14	24	7	0
FRAN	7	9	6	11	0

Pour l'ACM, nous pouvons partir du tableau d'effectifs ou du tableau de Burt :

	USA	GB	RFA	FRAN	MEDE	PHYS	CHIM	LITT	SECO
USA	139	0	0	0	55	43	24	8	9
GB	0	68	0	0	19	20	21	6	2
RFA	0	0	56	0	11	14	24	7	0
FRAN	0	0	0	33	7	9	6	11	0
MEDE	55	19	11	7	92	0	0	0	0
PHYS	43	20	14	9	0	86	0	0	0
CHIM	24	21	24	6	0	0	75	0	0
LITT	8	6	7	11	0	0	0	32	0
SECO	9	2	0	0	0	0	0	0	11

Remarquons les modalités de l'ACM sont formées par la réunion des individus-lignes et des individus colonnes de l'AFC.

L'inertie totale du nuage de points pour l'AFC est le coefficient $\Phi^2 = 0,15079$.

Pour l'ACM, l'inertie totale est : $I = \frac{9-2}{2} = 3,5$.

4.4.1 Valeurs propres obtenues par l'AFC et par l'ACM

Nombre de Dims.	Valeurs Propres et Inertie de toutes les Dimensions (Nobel-effectifs.sta)				
	Inertie Totale = ,15079 Chi ₂ = 44,633 dl = 12 p = ,00001				
	ValSing.	ValProp.	%age Inertie	%age Cumulé	Chi2
1	0,320054	0,102435	67,93261	67,9326	30,32064
2	0,219399	0,048136	31,92278	99,8554	14,24823
3	0,014767	0,000218	0,14461	100,0000	0,06455

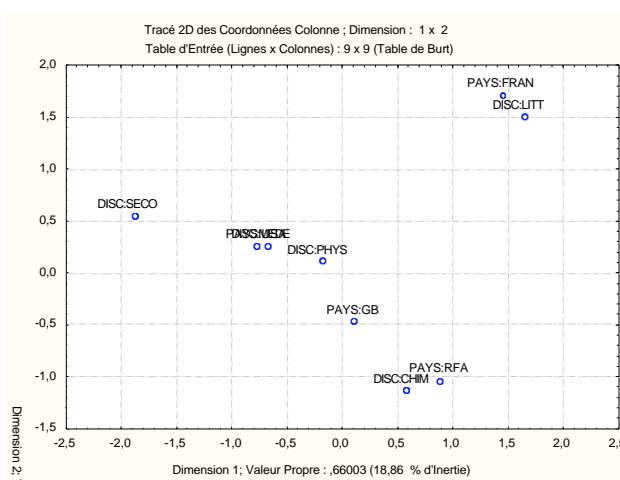
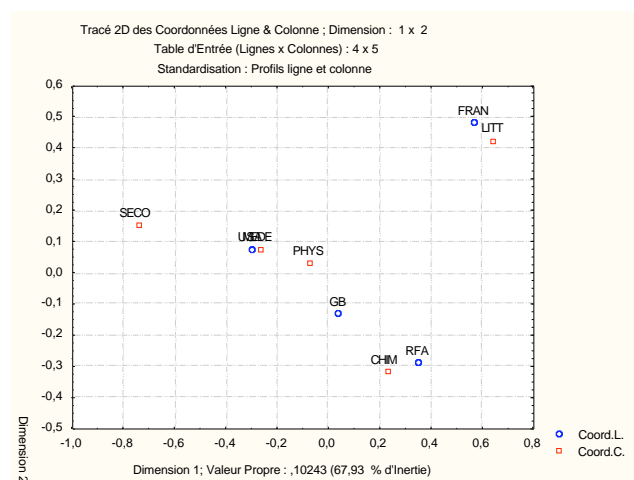
Nombre de Dims.	Valeurs Propres et Inertie de toutes les Dimensions (Nobel-Burt.sta)				
	Inertie Totale = 3,5000				

	ValSing.	ValProp.	%age Inertie	%age Cumulé	Chi2
1	0,812420	0,660027	18,85792	18,8579	407,5699
2	0,780833	0,609699	17,41999	36,2779	376,4924
3	0,712309	0,507383	14,49667	50,7746	313,3117
4	0,707107	0,500000	14,28571	65,0603	308,7524
5	0,701867	0,492617	14,07476	79,1350	304,1931
6	0,624740	0,390301	11,15144	90,2865	241,0124
7	0,583072	0,339973	9,71351	100,0000	209,9349

Appelons λ_1 , λ_2 et λ_3 les trois valeurs propres produites par l'AFC. L'ACM produit 7 valeurs propres qui sont reliées aux précédentes par les relations suivantes :

- Pour $i = 1, 2, 3$, on a : $\mu_i = \frac{1 + \sqrt{\lambda_i}}{2}$. Ces valeurs propres sont toutes supérieures à 1/2.
- La 4^e valeur propre vaut : $\mu_4 = \frac{1}{2}$. D'une manière générale, le nombre de valeurs propres égales à 1/2 est la différence entre le nombre d'individus lignes et le nombre d'individus colonnes.
- Pour $i = 5, 6, 7$, on a : $\mu_i = \frac{1 - \sqrt{\lambda_{8-i}}}{2}$. Ces valeurs propres sont toutes inférieures à 1/2.

4.4.1.1 Coordonnées des points-lignes et points-colonnes de l'AFC et des modalités pour l'ACM



On constate que la représentation graphique des modalités pour l'ACM est analogue à la représentation graphique conjointe des individus-lignes et individus-colonnes. La graduation des axes est cependant différente. En effet, les coordonnées des modalités sur le premier axe, par exemple, sont obtenues en

multipliant les coordonnées des individus lignes et des individus colonnes par $\sqrt{\frac{\mu_1}{\lambda_1}}$.

NomLigne	Coordonnées Ligne et Contributions à l'Inertie (Nobel-effectifs.sta)									
	Standardisation : Profils ligne et colonne									
	Ligne	Coord.	Coord.	Masse	Qualité	Inertie	Inertie	Cosinus2	Inertie	Cosinus2
	Numéro	Dim.1	Dim.2			Relative	Dim.1	Dim.1	Dim.2	Dim.2
USA	1	-0,298689	0,070541	0,469595	0,999641	0,293442	0,408993	0,946831	0,048544	0,052810
GB	2	0,042918	-0,134652	0,229730	0,968713	0,031412	0,004131	0,089335	0,086532	0,879378
RFA	3	0,350168	-0,292888	0,189189	0,998635	0,261831	0,226466	0,587570	0,337155	0,411064
FRAN	4	0,575454	0,477359	0,111486	0,999999	0,413315	0,360410	0,592372	0,527768	0,407627

Nom Col.	Coordonnées Colonne et Contributions à l'Inertie (Nobel-effectifs.sta)									
	Standardisation : Profils ligne et colonne									
	Colonne	Coord.	Coord.	Masse	Qualité	Inertie	Inertie	Cosinus2	Inertie	Cosinus2
	Numéro	Dim.1	Dim.2			Relative	Dim.1	Dim.1	Dim.2	Dim.2
MEDE	1	-0,262607	0,071397	0,310811	0,997798	0,152991	0,209248	0,929120	0,032914	0,068678
PHYS	2	-0,069169	0,028409	0,290541	0,915700	0,011765	0,013570	0,783527	0,004871	0,132173
CHIM	3	0,232856	-0,322084	0,253378	0,999783	0,265486	0,134121	0,343189	0,546057	0,656593
LITT	4	0,649223	0,421200	0,108108	0,999914	0,429420	0,444835	0,703713	0,398444	0,296200
SECO	5	-0,739184	0,151474	0,037162	0,999839	0,140337	0,198226	0,959545	0,017714	0,040294

NomLigne	Coordonnées Colonne et Contributions à l'Inertie (Nobel-Burt.sta)									
	Inertie Totale = 3,5000									
	Ligne	Coord.	Coord.	Masse	Qualité	Inertie	Inertie	Cosinus2	Inertie	Cosinus2
	Numéro	Dim.1	Dim.2			Relative	Dim.1	Dim.1	Dim.2	Dim.2
PAYS:USA	1	-0,75819	0,25105	0,234797	0,564745	0,075772	0,204496	0,508943	0,024272	0,055802
PAYS:GB	2	0,10894	-0,47922	0,114865	0,072033	0,110039	0,002065	0,003540	0,043266	0,068493
PAYS:RFA	3	0,88886	-1,04238	0,094595	0,437879	0,115830	0,113233	0,184351	0,168578	0,253528
PAYS:FRAN	4	1,46072	1,69890	0,055743	0,629884	0,126931	0,180205	0,267729	0,263884	0,362156
DISC:MEDE	5	-0,66660	0,25410	0,155405	0,229512	0,098456	0,104624	0,200394	0,016457	0,029118
DISC:PHYS	6	-0,17558	0,10111	0,145270	0,016811	0,101351	0,006785	0,012624	0,002436	0,004186
DISC:CHIM	7	0,59108	-1,14628	0,126689	0,564482	0,106660	0,067061	0,118566	0,273029	0,445916
DISC:LITT	8	1,64798	1,49904	0,054054	0,601569	0,127413	0,222418	0,329192	0,199222	0,272377
DISC:SECO	9	-1,87633	0,53909	0,018581	0,147101	0,137548	0,099113	0,135884	0,008857	0,011217

On peut enfin remarquer que, si l'on réalise une AFC en indiquant Nobel-Burt.sta comme tableau de contingence, Statistica produit des graphiques analogues aux précédents, mais indique encore des valeurs propres et une inertie totale différentes. Les valeurs propres sont alors les carrés des valeurs propres indiquées pour l'ACM.