

2.3 Analyse Factorielle des Correspondances

Bibliographie :

Escofier, Pagès : Analyses factorielles simples et multiples

Lebart, Morineau, Piron, Statistique exploratoire multidimensionnelle

G. Saporta. Probabilité, Analyse des données et statistique. Editions Technip , 1990

2.3.1 Introduction

L'analyse factorielle des correspondances (AFC), ou analyse des correspondances simples, est une méthode exploratoire d'analyse des tableaux de contingence. Elle a été développée essentiellement par J.-P. Benzecri durant la période 1970-1990.

Soient deux variables nominales X et Y, comportant respectivement p et q modalités. On a observé les valeurs de ces variables sur une population et on dispose d'un tableau de contingence à p lignes et q colonnes donnant les effectifs conjoints c'est-à-dire les effectifs observés pour chaque combinaison d'une modalité i de X et d'une modalité j de Y.

Les valeurs de ce tableau seront notées n_{ij} , l'effectif total sera noté N.

L'AFC vise à analyser ce tableau en apportant des réponses à des questions telles que :

- Y a-t-il des lignes du tableau (modalités de X) qui se "ressemblent", c'est-à-dire telles que les distributions des modalités de Y soient analogues ?
- Y a-t-il des lignes du tableau (modalités de X) qui s'opposent, c'est-à-dire telles que les distributions des modalités de Y soient très différentes ?
- Mêmes questions pour les colonnes du tableau.
- Y a-t-il des associations modalité de X - modalité de Y qui s'attirent (effectif conjoint particulièrement élevé) ou qui se repoussent (effectif conjoint particulièrement faible) ?

La méthode se fixe également comme but de construire des représentations graphiques mettant en évidence ces propriétés des données.

2.3.2 Traitement classique d'un tableau de contingence : test du khi-2 sur un exemple

	Droit	Sciences	Médecine	IUT
Exp. agri.	80	99	65	58
Patron	168	137	208	62
Cadre sup.	470	400	876	79
Employé	145	133	135	54
Ouvrier	166	193	127	129

Effectifs et fréquences marginaux

	Droit	Sciences	Médecine	IUT	Effectifs marginaux lignes	Fréquence
Exp. agri.	80	99	65	58	302	0,0798
Patron	168	137	208	62	575	0,1520
Cadre sup.	470	400	876	79	1825	0,4823
Employé	145	133	135	54	467	0,1234
Ouvrier	166	193	127	129	615	0,1625
Effectifs marginaux colonnes	1029	962	1411	382	3784	

Fréquence	0,2719	0,2542	0,3729	0,1010		
-----------	--------	--------	--------	--------	--	--

Fréquences théoriques dans l'hypothèse d'indépendance

$$\begin{bmatrix} 0,0798 \\ 0,1520 \\ 0,4823 \\ 0,1234 \\ 0,1625 \end{bmatrix} \times \begin{bmatrix} 0,2719 & 0,2542 & 0,3729 & 0,1010 \end{bmatrix} = \begin{bmatrix} 0,0217 & 0,0203 & 0,0298 & 0,081 \\ 0,0413 & 0,0386 & 0,0567 & 0,0153 \\ 0,1312 & 0,1226 & 0,1798 & 0,0487 \\ 0,0336 & 0,0314 & 0,0460 & 0,0125 \\ 0,0442 & 0,0413 & 0,0606 & 0,0164 \end{bmatrix}$$

Effectifs théoriques dans le cas d'indépendance

0,0217	0,0203	0,0298	0,0081	82,12	76,78	112,61	30,49
0,0413	0,0386	0,0567	0,0153	156,36	146,18	214,41	58,05
0,1312	0,1226	0,1798	0,0487	496,28	463,97	680,52	184,24
0,0336	0,0314	0,0460	0,0125	126,99	118,72	174,14	47,14
0,0442	0,0413	0,0606	0,0164	167,24	156,35	229,32	62,09

x 3784 =

Effectifs observés O

	Droit	Sciences	Médecine	IUT
Exp. agri.	80	99	65	58
Patron	168	137	208	62
Cadre sup.	470	400	876	79
Employé	145	133	135	54
Ouvrier	166	193	127	129

Effectifs théoriques T

	Droit	Sciences	Médecine	IUT
Exp. agri.	82,12	76,78	112,61	30,49
Patron	156,36	146,18	214,41	58,05
Cadre sup.	496,28	463,97	680,52	184,24
Employé	126,99	118,72	174,14	47,14
Ouvrier	167,24	156,35	229,32	62,09

Ecart à l'indépendance : E = O - T

	Droit	Sciences	Médecine	IUT
Exp. agri.	-2,12	22,22	-47,61	27,51
Patron	11,64	-9,18	-6,41	3,95
Cadre sup.	-26,28	-63,97	195,48	-105,24
Employé	18,01	14,28	-39,14	6,86
Ouvrier	-1,24	36,65	-102,32	66,91

Contributions au khi-2 : $(O - T)^2/T$

	Droit	Sciences	Médecine	IUT
Exp. agri.	0,05	6,43	20,13	24,83
Patron	0,87	0,58	0,19	0,27
Cadre sup.	1,39	8,82	56,15	60,11

Employé	2,55	1,72	8,80	1,00
Ouvrier	0,01	8,59	45,66	72,12

D'où : $\chi^2 = 320,2$.

Pour réaliser un test du χ^2 (ce qui suppose que les données observées constituent un échantillon tiré au hasard dans une population), on pose les hypothèses :

H_0 : Les variables X et Y sont indépendantes

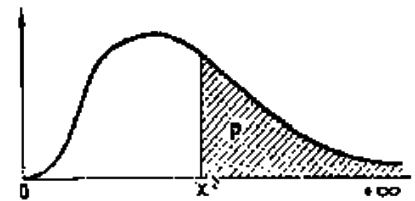
H_1 : Les variables X et Y sont dépendantes

Sous l'hypothèse H_0 , la distance entre les deux tableaux suit une loi du χ^2 à 12 degrés de liberté. Ce dernier nombre est défini par la formule :

$$ddl = (\text{Nb Modalités lignes} - 1)(\text{Nb Modalités colonnes} - 1) = 12$$

On choisit un seuil (5% par exemple) et on lit dans une table la valeur critique correspondante :

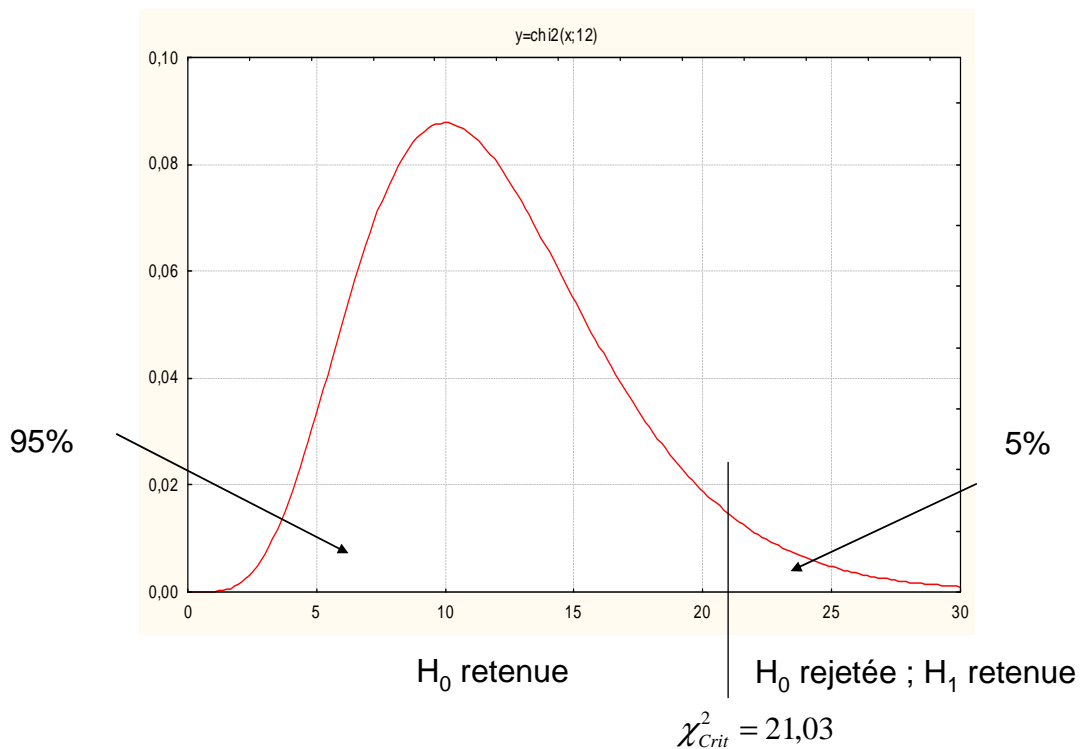
DISTRIBUTION DE χ^2 (Loi de K. Pearson)
Valeur de χ^2 ayant la probabilité P d'être dépassée.



v	0,9	0,8	0,7	0,5	0,3	0,2	0,1	0,05	0,02	0,01
1	0,0158	0,0642	0,148	0,455	1,074	1,642	2,706	3,841	5,412	6,635
2	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	7,824	9,210
3	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	9,837	11,345
4	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	11,668	13,277
5	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	13,388	15,086
6	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	15,033	16,812
7	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	16,622	18,475
8	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	18,168	20,090
9	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	19,679	21,666
10	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	21,161	23,209
11	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	22,618	24,725
12	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	24,054	26,217
13	7,041	8,634	9,926	12,340	15,119	16,985	19,812	22,362	25,471	27,688
14	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	26,873	29,141
15	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	28,259	30,578

On formule ensuite la règle de décision :

Loi du khi-2



Dans notre exemple, le χ^2 observé est très supérieur au χ^2 critique. On retient donc l'hypothèse H_1 : il existe un lien entre les deux variables étudiées.

2.3.3 Analyse factorielle des correspondances proprement dite

Notations :

Soit un tableau de contingence comportant p lignes et q colonnes.

- L'élément du tableau situé à l'intersection de la ligne i et de la colonne j est noté n_{ij} .
- La somme des éléments d'une ligne est notée $n_{i\bullet}$.
- La somme des éléments d'une colonne est notée $n_{\bullet j}$.

Distance (du Phi-2) entre deux profils lignes :

$$d_{ii'}^2 = \sum_{j=1}^q \frac{n_{\bullet j}}{n_{i\bullet}} \left(\frac{n_{ij}}{n_{i\bullet}} - \frac{n_{i'j}}{n_{i'\bullet}} \right)^2$$

Exemple : distance entre les lignes 1 et 2

	Droit	Sciences	Médecine	IUT	Effectifs marginaux lignes
Exp. agri.	80	99	65	58	302
Patron	168	137	208	62	575
Cadre sup.	470	400	876	79	1825
Employé	145	133	135	54	467
Ouvrier	166	193	127	129	615

Effectifs marginaux colonnes	1029	962	1411	382	3784
---------------------------------	------	-----	------	-----	------

$$d_{12}^2 = \frac{3784}{1029} \left(\frac{80}{302} - \frac{168}{575} \right)^2 + \frac{3784}{962} \left(\frac{99}{302} - \frac{137}{575} \right)^2 + \frac{3784}{1411} \left(\frac{65}{302} - \frac{208}{575} \right)^2 + \frac{3784}{382} \left(\frac{58}{302} - \frac{62}{575} \right)^2$$

Distance (du Phi-2) entre deux profils colonnes :

$$d_{jj'}^2 = \sum_{i=1}^p \frac{n}{n_{i\bullet}} \left(\frac{n_{ij}}{n_{\bullet j}} - \frac{n_{ij'}}{n_{\bullet j'}} \right)^2$$

Exemple : distance entre les colonnes 1 et 2

$$d_{12}^2 = \frac{3784}{302} \left(\frac{80}{1029} - \frac{99}{962} \right)^2 + \frac{3784}{575} \left(\frac{168}{1029} - \frac{137}{962} \right)^2 + \frac{3784}{1825} \left(\frac{470}{1029} - \frac{400}{962} \right)^2 + \frac{3784}{467} \left(\frac{145}{1029} - \frac{133}{962} \right)^2 + \frac{3784}{615} \left(\frac{166}{1029} - \frac{193}{962} \right)^2$$

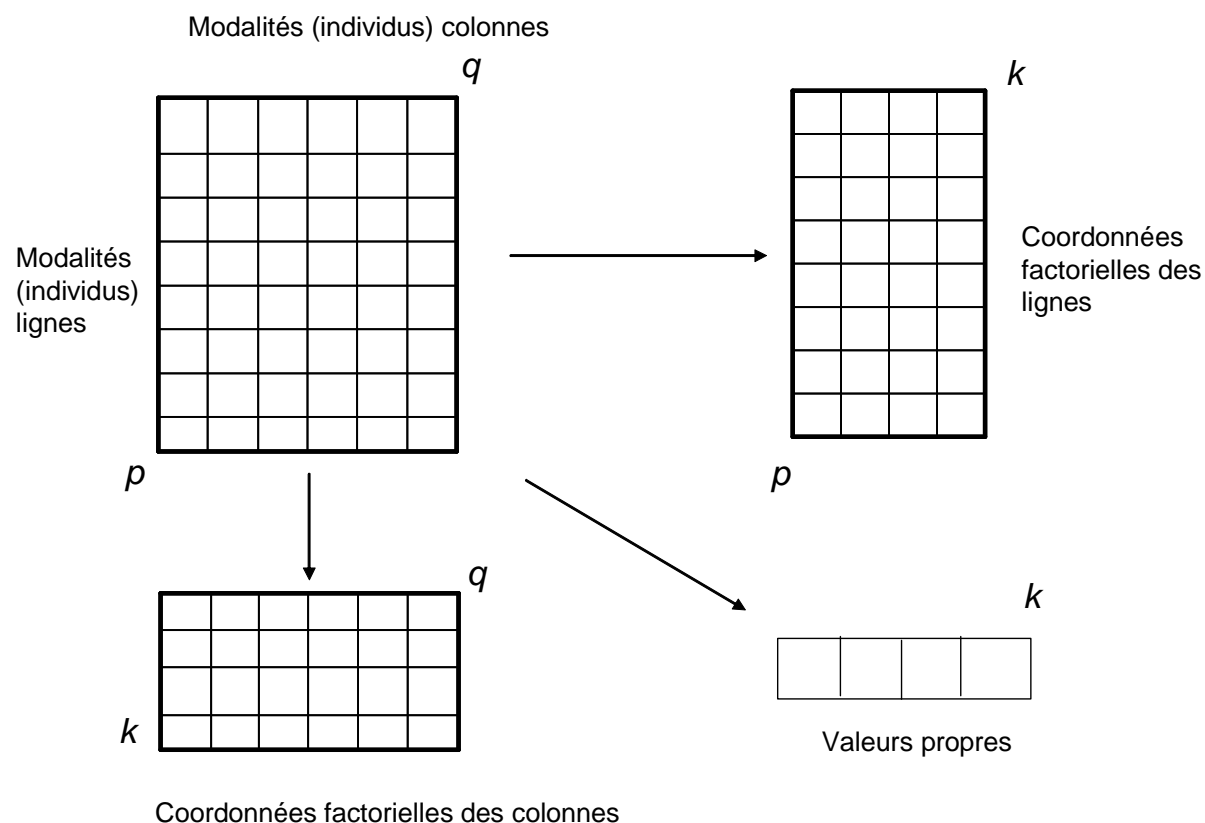
Propriété d'équivalence distributionnelle :

- Si on regroupe deux modalités lignes, les distances entre les profils-colonnes, ou entre les autres profils-lignes restent inchangées.
- Si on regroupe deux modalités colonnes, les distances entre les profils-lignes, ou entre les autres profils-colonnes restent inchangées.

L'analyse des correspondances détermine une représentation "optimale" de la distance du Phi-2 entre les individus lignes, et de même, une représentation optimale de la distance du Phi-2 entre les individus colonnes. Elle permet également de représenter les individus lignes et les individus colonnes sur une même carte factorielle.

Principaux résultats de l'AFC :

Principaux résultats d'une AFC



Valeurs propres

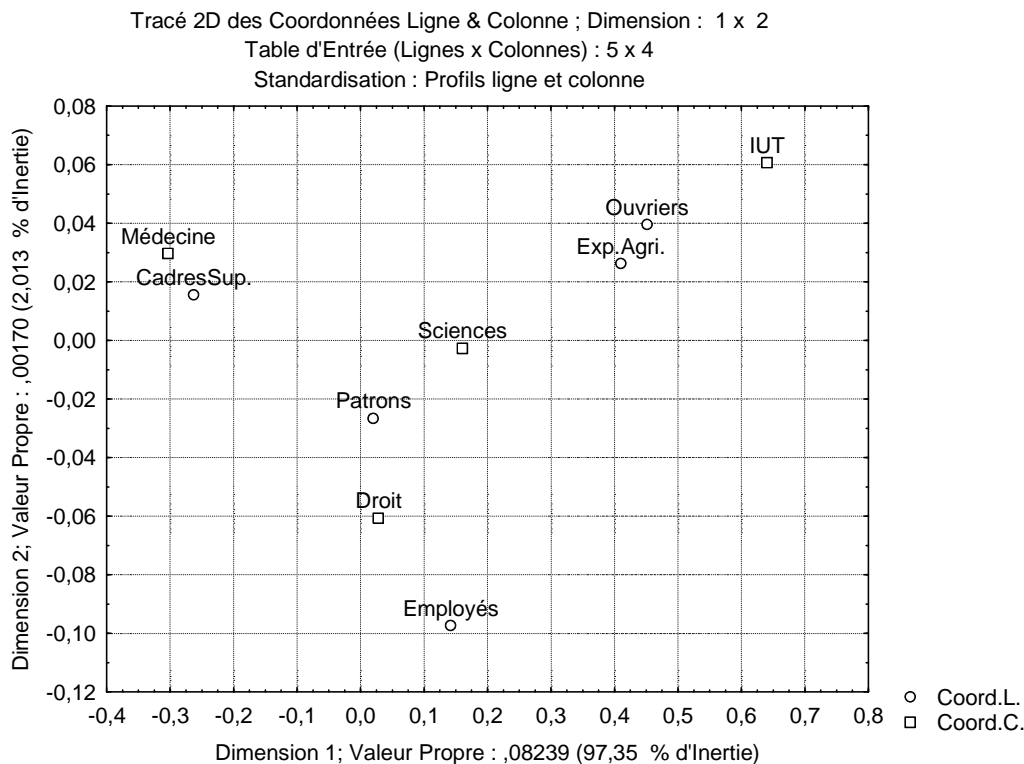
	ValProp.	%age inertie	%age cumulé	Chi ²
1	0,082	97,35	97,35	311,78
2	0,002	2,01	99,36	6,45
3	0,001	0,64	100,00	2,04

Résultats relatifs aux lignes

	Coord. Dim.1	Coord. Dim.2	Masse	Qualité	Inertie Relative	Inertie Dim.1	Cosinus ² Dim.1	Inertie Dim.2	Cosinus ² Dim.2
Exp. Agri.	0,410	0,026	0,080	0,991	0,161	0,163	0,987	0,032	0,004
Patrons	0,020	-0,027	0,152	0,336	0,006	0,001	0,123	0,063	0,213
Cadres Sup.	-0,263	0,016	0,482	0,999	0,395	0,404	0,996	0,069	0,004
Employés	0,142	-0,097	0,123	0,985	0,044	0,030	0,670	0,686	0,315
Ouvriers	0,451	0,040	0,163	1,000	0,395	0,402	0,992	0,150	0,008

Résultats relatifs aux colonnes

	Coord. Dim.1	Coord. Dim.2	Masse	Qualité	Inertie Relative	Inertie Dim.1	Cosinus ² Dim.1	Inertie Dim.2	Cosinus ² Dim.2
Droit	0,028	-0,061	0,272	0,942	0,015	0,003	0,165	0,588	0,777
Sciences	0,160	-0,003	0,254	0,948	0,082	0,079	0,948	0,001	0,000
Médecine	-0,303	0,030	0,373	1,000	0,409	0,416	0,990	0,193	0,009
IUT	0,640	0,061	0,101	0,998	0,494	0,502	0,989	0,219	0,009



2.3.4 Analyse factorielle des correspondances avec Statistica

2.3.4.1 Traitement des données avec Statistica

Source : Site Eurostat de l'Union Européenne.
<http://epp.eurostat.ec.europa.eu/portal/>

Ouvrez le classeur Regions-2001.stw

La feuille "Regions-Milliers-2001" rapporte des données relatives à la structure de la population : elle indique, pour chacune des 22 régions françaises (en lignes) le nombre d'habitants (en milliers) par âge (en colonnes) :

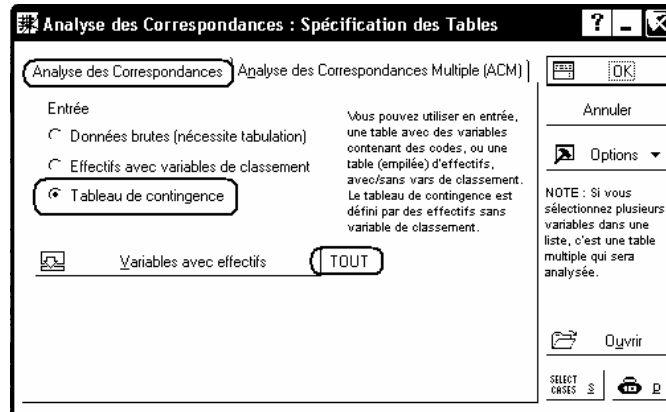
HF00 signifie Hommes et Femmes de 0 à 4 ans,

HF05 signifie Hommes et Femmes de 5 à 9 ans, ...

HF80 signifie Hommes et Femmes de plus de 80 ans

	1 HF00	2 HF05	3 HF10	4 HF15
ILEF	744	724	703	706
CHAM	82	86	93	95
PICA	120	128	138	134
HNOR	114	120	129	131

Pour effectuer l'AFC, nous utilisons le menu Statistiques - Techniques exploratoires multivariées - Analyse des correspondances.



La fenêtre de dialogue permet d'indiquer la manière dont se présentent nos données. La situation la plus classique est celle d'un tableau de contingence : les modalités lignes sont indiquées dans une variable spécifiques, les modalités colonnes sont les autres variables du tableau, et la feuille de données contient les effectifs n_{ij} .

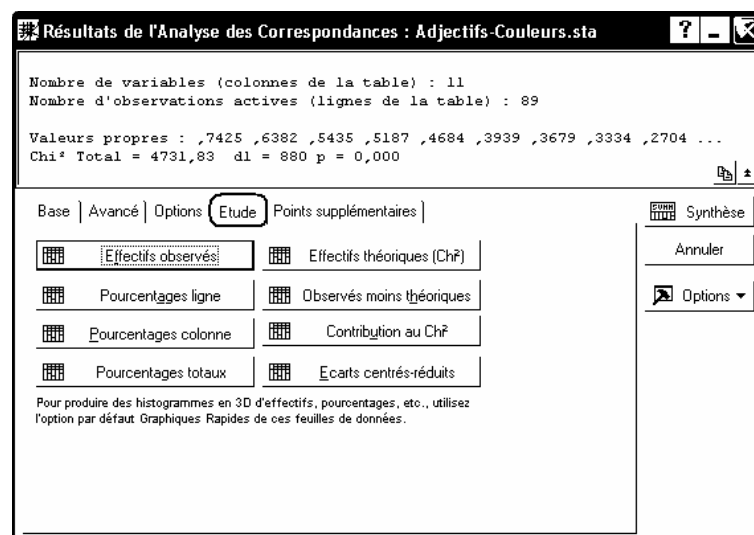
On indique également les variables qui participeront à l'analyse (ici toutes les variables). Notez que les zéros éventuels sont obligatoires, car une cellule laissée vide est interprétée comme une valeur manquante, et c'est alors l'ensemble de la ligne qui est éliminé de l'analyse.

N.B. Ne fermez pas l'analyse en cours pendant la suite des manipulations. Ainsi, vous n'aurez pas à indiquer de nouveau les options ci-dessus, vos résultats seront cohérents entre eux et se rassembleront dans un même classeur.

Statistiques descriptives

Les principaux résultats de statistiques descriptives pourront être obtenus à partir de l'onglet "Etude".

On peut ainsi obtenir les fréquences, les fréquences lignes, les fréquences colonnes et les profils moyens.



Par exemple, le tableau des fréquences et les profils ligne et colonne moyens sont :

Pourcentages Totaux (Regions-Milliers-2001 dans Regions-2001-Def.stw)																		
Table d'Entrée (Lignes x Colonnes) : 22 x 17																		
Inertie Totale = ,00882 Chi ² = 515,83 dl = 336 p = 0,0000																		
	HF00	HF05	HF10	HF15	HF20	HF25	HF30	HF35	HF40	HF45	HF50	HF55	HF60	HF65	HF70	HF75	HF80	Total
ILEF	1,27	1,24	1,20	1,21	1,29	1,59	1,56	1,49	1,37	1,37	1,26	0,88	0,74	0,67	0,57	0,46	0,54	18,71
CHAM	0,14	0,15	0,16	0,16	0,15	0,17	0,16	0,17	0,17	0,17	0,15	0,10	0,11	0,10	0,09	0,08	0,08	2,29
PICA	0,21	0,22	0,24	0,23	0,20	0,23	0,23	0,24	0,23	0,24	0,20	0,13	0,14	0,14	0,12	0,10	0,09	3,17
HNOR	0,19	0,21	0,22	0,22	0,19	0,22	0,22	0,22	0,23	0,22	0,19	0,13	0,13	0,13	0,11	0,09	0,10	3,04
CENT	0,24	0,26	0,27	0,27	0,24	0,29	0,29	0,30	0,30	0,31	0,27	0,19	0,21	0,21	0,19	0,16	0,18	4,17
BNOR	0,15	0,15	0,17	0,17	0,15	0,16	0,17	0,17	0,18	0,17	0,15	0,10	0,12	0,12	0,11	0,09	0,09	2,43
BOUR	0,15	0,16	0,17	0,18	0,16	0,18	0,19	0,19	0,20	0,20	0,18	0,13	0,14	0,15	0,13	0,11	0,12	2,75
NORD	0,46	0,48	0,52	0,54	0,49	0,50	0,49	0,49	0,49	0,48	0,41	0,26	0,29	0,29	0,26	0,21	0,18	6,82
LORR	0,23	0,25	0,27	0,28	0,26	0,28	0,29	0,30	0,29	0,29	0,24	0,18	0,19	0,19	0,17	0,12	0,12	3,95
ALSA	0,19	0,19	0,19	0,19	0,19	0,23	0,24	0,23	0,23	0,22	0,17	0,14	0,14	0,13	0,11	0,09	0,08	2,96
FCOM	0,12	0,12	0,13	0,14	0,12	0,14	0,14	0,14	0,14	0,14	0,12	0,09	0,09	0,09	0,08	0,06	0,07	1,91
PAYS	0,34	0,35	0,37	0,40	0,36	0,38	0,38	0,39	0,39	0,39	0,34	0,24	0,26	0,26	0,24	0,19	0,20	5,51
BRET	0,29	0,30	0,32	0,34	0,32	0,34	0,34	0,36	0,35	0,35	0,31	0,22	0,26	0,26	0,24	0,19	0,19	4,97
POIT	0,15	0,16	0,17	0,18	0,16	0,18	0,19	0,19	0,20	0,20	0,18	0,14	0,15	0,16	0,14	0,12	0,13	2,80
AQUI	0,26	0,28	0,30	0,31	0,30	0,33	0,34	0,36	0,36	0,37	0,33	0,24	0,25	0,26	0,25	0,21	0,22	4,97
MIDI	0,23	0,24	0,25	0,27	0,27	0,29	0,31	0,32	0,31	0,31	0,28	0,21	0,22	0,23	0,22	0,18	0,20	4,36
LIMO	0,05	0,06	0,06	0,07	0,07	0,08	0,08	0,08	0,09	0,09	0,08	0,06	0,07	0,08	0,07	0,06	0,07	1,21
RHON	0,61	0,63	0,64	0,66	0,62	0,70	0,73	0,72	0,69	0,68	0,63	0,47	0,43	0,43	0,38	0,31	0,32	9,65
AUVE	0,11	0,12	0,13	0,14	0,14	0,15	0,15	0,16	0,16	0,17	0,15	0,11	0,12	0,12	0,11	0,10	0,10	2,24
LANG	0,22	0,23	0,24	0,25	0,24	0,26	0,26	0,28	0,27	0,28	0,25	0,19	0,20	0,22	0,20	0,17	0,17	3,92
PROV	0,44	0,47	0,48	0,48	0,45	0,50	0,54	0,55	0,54	0,54	0,51	0,41	0,39	0,40	0,36	0,31	0,34	7,70
CORS	0,02	0,03	0,03	0,03	0,02	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,02	0,02	0,02	0,02	0,02	0,44
Total	6,09	6,29	6,55	6,74	6,36	7,20	7,31	7,37	7,22	7,22	6,43	4,66	4,66	4,66	4,17	3,43	3,63	100,00

Remarque :

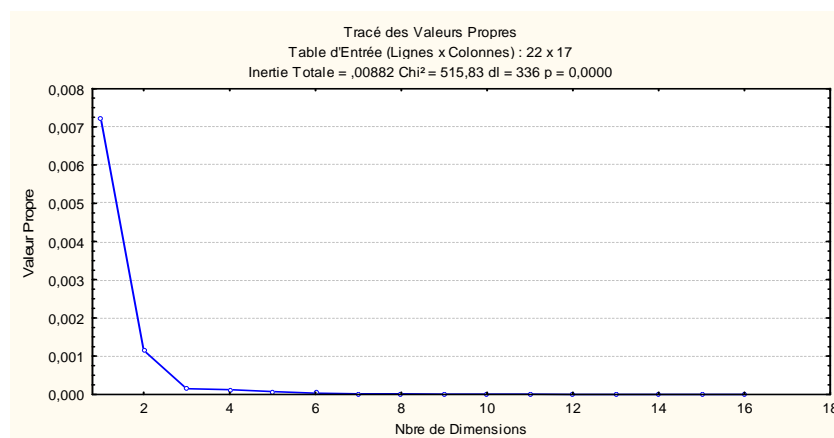
Statistica ne permet pas d'obtenir directement le tableau des taux de liaison, qui est pourtant un outil exploratoire intéressant. Mais on pourra utiliser les tableaux "Observés moins théoriques" et "Effectifs théoriques". On peut même recopier ces deux tableaux dans une feuille Excel et diviser chaque cellule du premier par la cellule correspondante du second pour obtenir le tableau des taux de liaison :

	HF00	HF05	HF10	HF15	HF20	HF25	HF30	HF35	HF40	HF45	HF50	HF55	HF60	HF65	HF70	HF75	HF80
ILEF	0,12	0,05	-0,02	-0,04	0,09	0,18	0,14	0,08	0,02	0,02	0,05	0,01	-0,15	-0,23	-0,27	-0,28	-0,21
CHAM	0,00	0,02	0,06	0,05	0,01	0,00	-0,02	-0,01	0,00	0,01	-0,00	-0,07	-0,01	-0,02	-0,02	-0,02	-0,04
PICA	0,06	0,10	0,14	0,07	-0,03	-0,01	0,01	0,01	0,02	0,04	-0,02	-0,10	-0,06	-0,09	-0,08	-0,12	-0,18
HNOR	0,05	0,07	0,11	0,09	-0,01	-0,01	-0,02	-0,00	0,04	0,02	-0,01	-0,08	-0,06	-0,07	-0,10	-0,10	-0,10
CENT	-0,04	-0,02	-0,01	-0,03	-0,09	-0,05	-0,05	-0,03	-0,01	0,01	0,01	-0,01	0,05	0,07	0,08	0,11	0,20
BNOR	0,00	0,01	0,05	0,05	-0,04	-0,07	-0,07	-0,04	0,00	-0,02	-0,05	-0,08	0,09	0,10	0,10	0,07	0,01
BOUR	-0,09	-0,06	-0,03	-0,03	-0,09	-0,09	-0,07	-0,05	-0,01	0,02	0,01	0,02	0,09	0,15	0,16	0,18	0,23
NORD	0,11	0,12	0,17	0,16	0,12	0,02	-0,03	-0,02	-0,01	-0,03	-0,07	-0,18	-0,10	-0,08	-0,08	-0,11	-0,28
LORR	-0,03	0,01	0,04	0,04	0,02	-0,02	-0,01	0,02	0,03	0,02	-0,04	-0,02	0,04	0,04	0,01	-0,08	-0,17
ALSA	0,03	0,04	0,01	-0,02	0,01	0,07	0,10	0,07	0,06	0,01	-0,09	0,04	-0,02	-0,06	-0,11	-0,14	-0,24
FCOM	-0,00	-0,00	0,04	0,06	-0,03	-0,02	-0,01	-0,03	0,00	-0,01	0,02	-0,00	0,02	0,02	-0,01	-0,03	-0,04
PAYS	0,02	0,01	0,03	0,09	0,03	-0,04	-0,05	-0,03	-0,01	-0,01	-0,05	-0,07	0,02	0,02	0,03	0,03	0,02
BRET	-0,04	-0,04	-0,02	0,02	0,00	-0,06	-0,08	-0,03	-0,01	-0,02	-0,03	-0,05	0,11	0,12	0,15	0,14	0,06
POIT	-0,14	-0,12	-0,07	-0,04	-0,08	-0,12	-0,09	-0,06	0,01	0,01	-0,01	0,04	0,14	0,19	0,23	0,25	0,31
AQUI	-0,13	-0,11	-0,09	-0,06	-0,07	-0,07	-0,06	-0,03	-0,00	0,02	0,02	0,04	0,10	0,14	0,19	0,22	0,24
MIDI	-0,12	-0,11	-0,12	-0,09	-0,03	-0,07	-0,03	-0,01	-0,01	-0,01	0,01	0,04	0,10	0,14	0,19	0,22	0,27
LIMO	-0,26	-0,22	-0,18	-0,14	-0,11	-0,12	-0,13	-0,08	-0,02	0,01	0,03	0,06	0,21	0,33	0,42	0,44	0,59
RHON	0,04	0,04	0,02	0,02	0,00	0,01	0,03	0,01	-0,01	-0,02	0,02	0,04	-0,04	-0,05	-0,06	-0,08	-0,08
AUVE	-0,17	-0,15	-0,12	-0,06	-0,05	-0,09	-0,07	-0,06	-0,00	0,04	0,06	0,08	0,13	0,18	0,23	0,25	0,22
LANG	-0,09	-0,07	-0,06	-0,05	-0,04	-0,09	-0,08	-0,05	-0,05	-0,02	0,00	0,07	0,09	0,18	0,20	0,24	0,22
PROV	-0,06	-0,03	-0,05	-0,07	-0,09	-0,10	-0,04	-0,03	-0,03	-0,03	0,02	0,13	0,08	0,11	0,13	0,17	0,21
CORS	-0,12	-0,08	-0,06	-0,09	-0,21	-0,09	-0,00	-0,01	0,01	0,01	0,02	0,24	0,15	0,16	0,11	0,12	0,17

Choix des valeurs propres

C'est ensuite l'onglet "Avancé" qui nous permettra d'afficher les valeurs propres, et donc de choisir le nombre d'axes à garder.

Valeurs Propres et Inertie de toutes les Dimensions) Table d'Entrée (Lignes x Colonnes) : 22 x 17 Inertie Totale = ,00882 Chi ² = 515,83 dl = 336 p = 0,0001					
Nombre de Dims.	ValSing.	ValProp.	%age Inertie	%age Cumulé	Chi ²
1	0,0850	0,0072	81,92	81,92	422,54
2	0,0340	0,0012	13,14	95,06	67,78
3	0,0123	0,0002	1,71	96,76	8,81
4	0,0113	0,0001	1,44	98,20	7,42
5	0,0086	0,0001	0,85	99,05	4,37
6	0,0058	0,0000	0,38	99,43	1,97
7	0,0042	0,0000	0,20	99,64	1,05
8	0,0035	0,0000	0,14	99,78	0,71
9	0,0024	0,0000	0,07	99,84	0,35
10	0,0023	0,0000	0,06	99,90	0,30
11	0,0018	0,0000	0,04	99,94	0,18
12	0,0015	0,0000	0,02	99,96	0,13
13	0,0012	0,0000	0,02	99,98	0,08
14	0,0010	0,0000	0,01	99,99	0,06
15	0,0008	0,0000	0,01	100,00	0,04
16	0,0006	0,0000	0,00	100,00	0,02

*Résultats relatifs aux individus-lignes et aux individus-colonnes.*

Pour les résultats qui suivent, on indique le nombre d'axes factoriels à conserver sous l'onglet "Base" ou sous l'onglet "Options". Ce dernier permet également de choisir plusieurs types d'échelles pour représenter lignes et colonnes. Le type de représentation le plus classique, qui fait jouer des rôles symétriques aux lignes et aux colonnes, correspond à la première option.

Résultats de l'Analyse des Correspondances : Adjectifs-Couleurs.sta

Nombre de variables (colonnes de la table) : 11
 Nombre d'observations actives (lignes de la table) : 89

Valeurs propres : ,7425 ,6382 ,5435 ,5187 ,4684 ,3939 ,3679 ,3334 ,2704 ...
 Chi² Total = 4731,83 dl = 880 p = 0,000

Base | Avancé | Options | Etude | Points supplémentaires

Nombre de dimensions
 Nombre de dimensions : 2
 Contribution cumulée à l'inertie :

Centrer/réduire des coordonnées
 Profils ligne & colonne
 Standardisation canonique
 Profils ligne (interpréter dist. ligne)
 Profils colonne (interpréter dist. col.)

Synthèse
 Annuler
 Options

On retourne ensuite sous l'onglet "Avancé" pour afficher les coordonnées des individus-lignes et des individus-colonnes. On notera que Statistica produit deux tableaux de résultats, et on passera de l'un à l'autre à l'aide des onglets du classeur.

Coordonnées Ligne et Contributions à l'Inertie (Regions-Milliers-2001 dans Regions-2001-Def.stw)											
Table d'Entrée (Lignes x Colonnes) : 22 x 17											
Standardisation : Profils ligne et colonne											
NomLigne	Ligne Numéro	Coord. Dim.1	Coord. Dim.2	Masse	Qualité	Inertie Relative	Inertie Dim.1	Cosinus² Dim.1	Inertie Dim.2	Cosinus² Dim.2	Cos² 1&2
ILEF	1	-0,1223	-0,0427	0,1871	0,9968	0,3574	0,3877	0,8885	0,2944	0,1082	0,9968
CHAM	2	-0,0129	0,0222	0,0229	0,7971	0,0022	0,0005	0,1997	0,0098	0,5974	0,7971
PICA	3	-0,0568	0,0396	0,0317	0,8540	0,0202	0,0142	0,5745	0,0430	0,2794	0,8540
HNOR	4	-0,0444	0,0356	0,0304	0,8471	0,0132	0,0083	0,5163	0,0332	0,3308	0,8471
CENT	5	0,0547	-0,0018	0,0417	0,8358	0,0170	0,0173	0,8349	0,0001	0,0009	0,8358
BNOR	6	0,0332	0,0420	0,0243	0,9231	0,0086	0,0037	0,3556	0,0369	0,5675	0,9231
BOUR	7	0,0900	0,0005	0,0275	0,9771	0,0258	0,0308	0,9771	0,0000	0,0000	0,9771
NORD	8	-0,0774	0,0788	0,0682	0,9772	0,0967	0,0566	0,4795	0,3662	0,4977	0,9772
LORR	9	-0,0151	0,0250	0,0395	0,4218	0,0090	0,0012	0,1126	0,0213	0,3092	0,4218
ALSA	10	-0,0618	-0,0090	0,0296	0,6520	0,0201	0,0157	0,6385	0,0021	0,0135	0,6520
FCOMTE	11	-0,0029	0,0160	0,0191	0,3961	0,0014	0,0000	0,0124	0,0042	0,3837	0,3961
PAYS	12	0,0088	0,0342	0,0551	0,8109	0,0096	0,0006	0,0498	0,0557	0,7611	0,8109
BRET	13	0,0569	0,0241	0,0497	0,8945	0,0241	0,0223	0,7586	0,0249	0,1359	0,8945
POIT	14	0,1223	-0,0047	0,0280	0,9891	0,0481	0,0580	0,9876	0,0005	0,0015	0,9891
AQUI	15	0,0985	-0,0184	0,0497	0,9830	0,0576	0,0668	0,9500	0,0145	0,0330	0,9830
MIDI	16	0,0994	-0,0293	0,0436	0,9659	0,0550	0,0597	0,8888	0,0323	0,0771	0,9659
LIMO	17	0,2133	-0,0393	0,0121	0,9774	0,0663	0,0765	0,9454	0,0162	0,0321	0,9774
RHON	18	-0,0312	-0,0007	0,0965	0,7453	0,0143	0,0130	0,7450	0,0000	0,0003	0,7453
AUVE	19	0,1155	-0,0238	0,0224	0,9426	0,0374	0,0413	0,9042	0,0110	0,0385	0,9426
LANG	20	0,1002	-0,0046	0,0392	0,9730	0,0460	0,0546	0,9709	0,0007	0,0021	0,9730
PROV	21	0,0791	-0,0202	0,0770	0,9122	0,0638	0,0667	0,8561	0,0272	0,0561	0,9122
CORS	22	0,0872	-0,0389	0,0044	0,7346	0,0063	0,0047	0,6129	0,0058	0,1216	0,7346

Coordonnées Colonne et Contributions à l'Inertie (Regions-Milliers-2001 dans Regions-2001-Def.stw)											
Table d'Entrée (Lignes x Colonnes) : 22 x 17											
Standardisation : Profils ligne et colonne											
Nom Col.	Colonne Numéro	Coord. Dim.1	Coord. Dim.2	Masse	Qualité	Inertie Relative	Inertie Dim.1	Cosinus ² Dim.1	Inertie Dim.2	Cosinus ² Dim.2	Cos ² 1&2
HF00	1	-0,0882	0,0106	0,0609	0,9564	0,0570	0,0656	0,9427	0,0059	0,0137	0,9564
HF05	2	-0,0637	0,0262	0,0629	0,9260	0,0365	0,0353	0,7916	0,0374	0,1343	0,9260
HF10	3	-0,0426	0,0602	0,0655	0,9515	0,0424	0,0164	0,3172	0,2047	0,6342	0,9515
HF15	4	-0,0262	0,0634	0,0674	0,9676	0,0372	0,0064	0,1407	0,2338	0,8269	0,9676
HF20	5	-0,0585	0,0094	0,0636	0,7551	0,0336	0,0302	0,7360	0,0049	0,0191	0,7551
HF25	6	-0,0881	-0,0335	0,0720	0,9645	0,0753	0,0774	0,8427	0,0698	0,1219	0,9645
HF30	7	-0,0661	-0,0373	0,0731	0,9848	0,0485	0,0443	0,7471	0,0878	0,2378	0,9848
HF35	8	-0,0384	-0,0207	0,0737	0,9260	0,0172	0,0150	0,7180	0,0271	0,2080	0,9260
HF40	9	-0,0121	-0,0029	0,0722	0,3271	0,0039	0,0015	0,3089	0,0005	0,0182	0,3271
HF45	10	-0,0041	-0,0067	0,0722	0,1446	0,0034	0,0002	0,0391	0,0028	0,1055	0,1446
HF50	11	-0,0018	-0,0329	0,0643	0,6808	0,0116	0,0000	0,0021	0,0601	0,6788	0,6808
HF55	12	0,0316	-0,0602	0,0466	0,7834	0,0312	0,0065	0,1696	0,1456	0,6138	0,7834
HF60	13	0,0982	0,0074	0,0466	0,9549	0,0537	0,0623	0,9495	0,0022	0,0054	0,9549
HF65	14	0,1405	0,0229	0,0466	0,9866	0,1086	0,1274	0,9612	0,0211	0,0255	0,9866
HF70	15	0,1672	0,0250	0,0417	0,9936	0,1361	0,1615	0,9719	0,0225	0,0217	0,9936
HF75	16	0,1851	0,0122	0,0343	0,9904	0,1351	0,1626	0,9861	0,0044	0,0043	0,9904
HF80	17	0,1932	-0,0471	0,0363	0,9643	0,1688	0,1876	0,9103	0,0694	0,0540	0,9643

On utilise ensuite les boutons du bloc "Tracé des coordonnées" pour obtenir des représentations graphiques des résultats de l'AFC.

Tracé des coordonnées

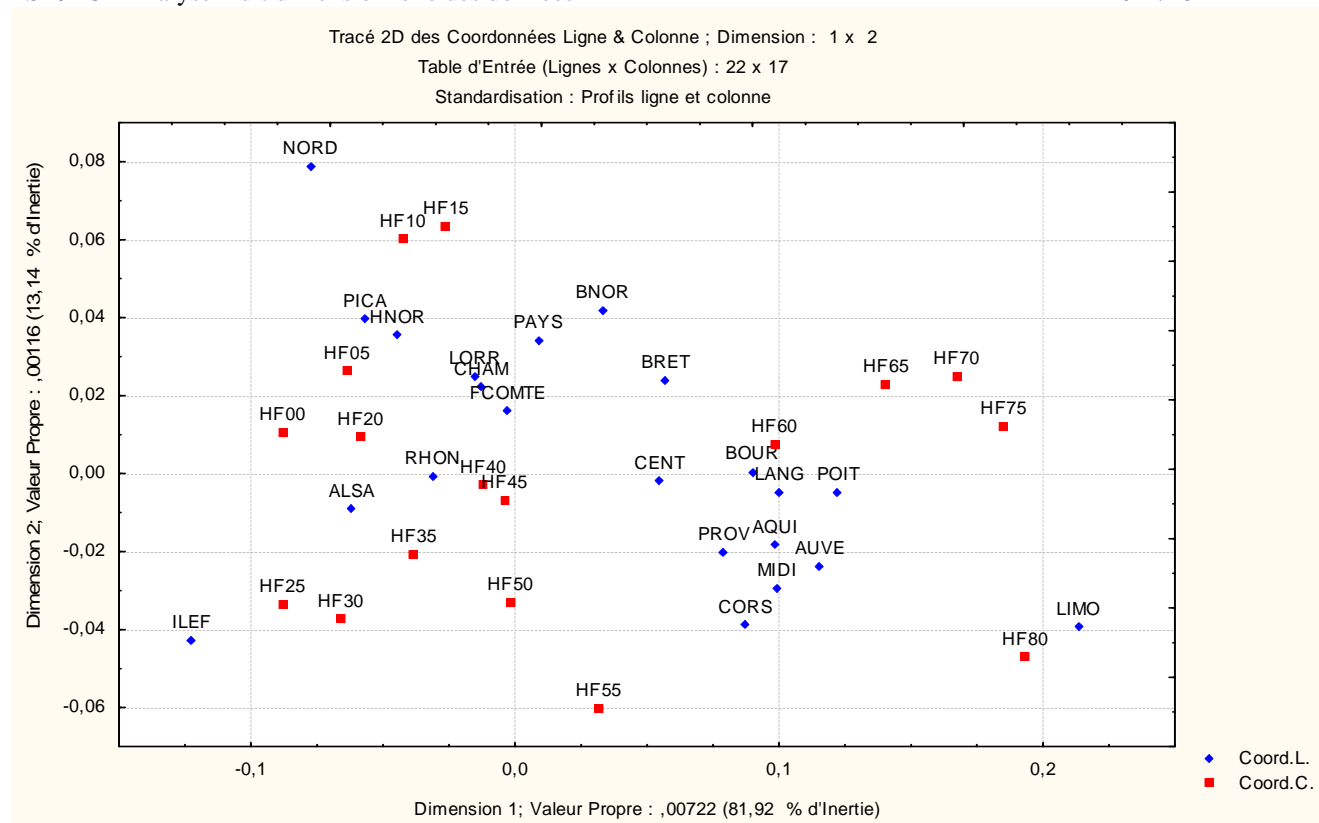
<input type="checkbox"/> Lignes, 1D	<input type="checkbox"/> 2D	<input type="checkbox"/>
<input type="checkbox"/> Colonne, 1D	<input type="checkbox"/> 2D	<input type="checkbox"/>
<input type="checkbox"/> Ligne & colonne, 1D	<input type="checkbox"/> 2D	<input type="checkbox"/>

Ne tracer que les dimensions sélectionnées

Tronquer les étiquettes à caractères

Utiliser des échelles X/Y/(Z) identiques

Les graphiques "par axe" pourront être obtenus à l'aide du bouton "Ligne & colonne, 1D". Le graphique dans un plan, superposant les résultats des lignes et des colonnes, pourra être obtenu à l'aide du bouton "2D" de la même ligne. En revanche, il n'est pas évident d'éliminer certaines étiquettes pour améliorer la lisibilité du graphique. La seule méthode paraît être de faire un clic droit sur une étiquette, de sélectionner l'item de menu "Propriétés..." puis d'éditer manuellement le tableau des étiquettes qui s'affiche.



2.3.5 Interprétation des résultats de l'AFC

On sait que la distance du khi-2 est sensible à l'importance de l'effectif observé. Sans surprise, même en exprimant les effectifs en milliers, nous obtenons ici un khi-2 de 515,83. En revanche, le coefficient Phi-2 est assez faible : 0,0088.

De même, on constate que les taux de liaison restent modérés, compris entre -0,28 et +0,59. Autrement dit le déficit d'une classe d'âge est au plus de 28% de l'effectif théorique que l'on obtiendrait si la structure par âge de la population française se retrouvait à l'identique dans toutes les régions, et l'excès d'une classe d'âge est d'au plus 59% de cet effectif théorique.

On sait que, dans une AFC, les valeurs propres sont toutes inférieures à 1, et que leur somme est égale au coefficient Phi-2. Ici, la décroissance des valeurs propres est très rapide, puisque la première représente plus de 80% de l'inertie. La deuxième, bien que très inférieure à la première, est supérieure à la moyenne $0,0088/16 = 0,00055$. Nous étudierons donc les deux premiers axes factoriels.

2.3.5.1 Interprétation du premier axe

Les individus lignes dont la contribution à l'inertie du premier axe est supérieure à la moyenne sont :

-	+
ILEF (39%)	LIMO (7,7%)
NORD (5,7%)	AQUI (6,7%)
	PROV (6,7%)
	MIDI (6%)

	POIT (5,8%) LANG (5,5%)
--	----------------------------

On voit que cet axe oppose des régions telles que l'Ile de France et le Nord Pas de Calais à un ensemble de régions "du sud" : Limousin, Aquitaine, Provence, etc. L'Ile de France représente à elle seule 39% de l'inertie de cet axe, et on peut s'étonner que cette région, malgré son poids démographique, soit représentée par un point aussi éloigné de l'origine des axes.

Pour les individus colonnes, les résultats sont :

-	+
HF25 (8%)	HF80 (19%)
HF00 (7%)	HF75 (16%)
	HF70 (16%)
	HF65 (13%)
	HF60 (6%)

Clairement, le premier axe oppose les classes d'âge élevées (partie positive de l'axe) aux autres classes, notamment la classe 25-29 ans et la classe 0-4 ans.

La synthèse des études menées sur les individus lignes et sur les individus colonnes en découle aussitôt : le premier axe oppose des régions où la population âgée est importante à des régions plus jeunes, ou dans lesquelles apparaît un déficit en personnes âgées (Ile de France et Nord Pas de Calais, mais aussi Alsace, Picardie, Haute Normandie, etc).

2.3.5.2 Etude du second axe factoriel

Les individus lignes dont la contribution à l'inertie du premier axe est supérieure à la moyenne sont :

-	+
ILEF (29,4%)	NORD (36,6%)

Les individus colonnes dont la contribution à l'inertie du premier axe est supérieure à la moyenne sont :

-	+
HF55 (14,6%)	HF15 (23,3%)
HF30 (8,8%)	HF10 (20,4%)
HF25 (7%)	
HF80 (7%)	
HF50 (6%)	

Le tableau des individus lignes semble montrer que cet axe oppose essentiellement deux régions "jeunes" : l'Ile de France et le Nord Pas de Calais. En fait, dans la partie négative de cet axe, on retrouve à la fois des régions "jeunes", telles que l'Ile de France et des régions "âgées" telles que le Limousin, pendant que la partie positive de l'axe rassemble des régions (Nord, mais aussi Picardie, Basse Normandie, Pays de la Loire, etc) où la population des adolescents (HF10, HF15) est bien représentée.

2.3.5.3 Quelques remarques sur les qualités de représentation

On voit que les âges correspondant aux adultes actifs (HF35, HF40, HF45) sont très peu intervenus dans l'étude. Les effectifs de ces classes d'âge diffèrent peu de l'indépendance : il y a peu de différences entre les régions du point de vue de la proportion de 35-49 ans dans la population. De faible inertie et donc intervenant peu dans la formation des premiers axes, ces individus colonnes peuvent être mal représentés (qualité de représentation égale à 0,14, par exemple, pour HF45 et à 0,32 pour HF40 : il faut donc s'abstenir d'interpréter, sans élément supplémentaire, leur proximité sur le graphique).

De même, la qualité de représentation de la Franche Comté (0,39) est assez faible, car cette région est peu importante numériquement et a un profil assez proche du profil moyen. Sur le schéma, elle apparaît proche de la Champagne, ce qui ne correspond pas vraiment à la réalité.

2.3.5.4 Synthèse

L'élément dominant que l'AFC fait apparaître est l'opposition entre d'une part les régions comportant beaucoup de personnes âgées (60 ans et plus), et par voie de conséquence, un déficit d'enfants et de jeunes adultes, et d'autre part, les régions comportant beaucoup de jeunes de moins de 35/40 ans et peu de personnes âgées. Une structure secondaire distingue, parmi les régions "jeunes" celles dont la population comporte de nombreux adultes (classes HF25, HF30 particulièrement nombreuses) à celles dont la population comporte beaucoup d'enfants (HF05, HF10, HF15).

On est ainsi tenté de définir quatre groupes de régions, sans pour autant pouvoir affecter objectivement chaque région à un groupe :

- Régions à population de personnes âgées importante : Limousin, Corse, Midi-Pyrénées, Auvergne, Provence, Aquitaine, Languedoc, Poitou, Bourgogne.
- Régions "intermédiaires" : Centre, Bretagne, Basse Normandie, Pays de la Loire et peut-être Lorraine, Champagne, Franche Comté
- Régions à forte population de jeunes adultes : Ile de France, Alsace et peut-être Rhône-Alpes.
- Régions à forte population de jeunes enfants : Nord Pas-de-Calais, Picardie, Haute-Normandie.

2.3.6 Structures possibles pour les données d'entrée

Source : Exemple fourni avec le logiciel Statistica.

Supposons que vous ayez collecté des données sur les habitudes de différents salariés d'une entreprise concernant la cigarette. Les données suivantes sont présentées dans l'ouvrage de Greenacre (1984, p. 55).

Ouvrez le classeur Smoking.stw et observez les 3 feuilles de données saisies.

2.3.6.1 Données structurées sous forme d'un tableau de contingence

Commençons, par exemple, par rendre active la feuille de données Smoking1.sta (tableau de contingence).

	Analyse des correspondances simple.			
	1 NON_FUM	2 OCCAS	3 MOYEN	4 GROS_FUM
CADRE_EXPER	4	2	3	2
CADRE_DEBUT	4	3	7	4
EMPLOY_EXPER	25	10	12	4
EMPLOY_DEBUT	18	24	33	13
SECRETAIRES	10	6	7	2

Réalisez une AFC sur ce tableau de données.

N.B. On remarquera que le test du khi-2 sur ce tableau ne démontre pas l'existence d'une dépendance significative entre les habitudes concernant la cigarette et l'emploi occupé. L'analyse factorielle des correspondances est donc d'un intérêt très limité ici.

2.3.6.2 Données structurées sous forme de tableau d'effectifs

Statistica nous permet également de réaliser l'AFC à partir d'un tableau d'effectifs (feuille de données Smoking2.sta).

Refaites l'AFC précédente, d'abord en utilisant Smoking2.sta comme feuille de données active.

2.3.6.3 Données structurées sous forme de tableau protocole

On peut aussi réaliser l'AFC à partir d'un tableau protocole (données non recensées - feuille de données Smoking3.sta).

Refaites l'AFC précédente, d'abord en utilisant Smoking3.sta comme feuille de données active.

2.3.7 Ajout de lignes ou de colonnes supplémentaires : application à la comparaison de tableaux de fréquence binaire

On dispose des données suivantes relatives aux élèves scolarisés en 1972/73, sortis du système éducatif en 1973 et ayant trouvé un emploi :

Hommes	Sans diplôme	BEPC	BEP/CAP	BAC général	BAC technique	DEUG/ENT	DUT/BTS/Santé	SUP	Total
Agriculteurs	15068	2701	5709	297	1242	0	322	0	25339
Ingénieurs	0	337	309	917	0	308	0	4383	6254
Techniciens	302	1697	2242	1969	1399	357	1943	381	10290
Ouvriers Qualifiés	10143	3702	30926	314	1861	0	0	337	47283
Ouvriers non qualifiés	59394	8087	17862	2887	1696	0	0	323	90249
Cadres supérieurs	596	298	892	1227	298	2362	318	6781	12772
Cadres Moyens	2142	2801	672	6495	924	2807	2301	4030	22172
Employés qualifiés	5445	7348	4719	4353	1280	614	982	0	24741
Employés non qualifiés	4879	4987	1514	3478	886	1326	0	661	17731
Total	97969	31958	64845	21937	9586	7774	5866	16896	256831

Femmes	Sans diplôme	BEPC	BEP/CAP	BAC général	BAC technique	DEUG/ENT	DUT/BTS/Santé	SUP	Total
Agriculteurs	5089	1212	1166	0	0	0	0	0	7467
Ingénieurs	0	0	0	316	0	0	304	1033	1653

Techniciens	281	0	320	320	283	0	683	0	1887
Ouvriers Qualifiés	7470	1859	4017	1752	657	0	285	0	16040
Ouvriers non qualifiés	29997	4334	4538	1882	0	0	0	0	40751
Cadres supérieurs	0	0	0	2236	595	911	569	6788	11099
Cadres Moyens	1577	1806	4549	17063	875	4152	15731	3991	49744
Employés qualifiés	21616	19915	32452	16137	5865	1256	3332	1286	101859
Employés non qualifiés	19849	7325	6484	5111	898	294	635	0	40596
Total	85879	36451	53526	44817	9173	6613	21539	13098	271096

Source : B. Escoffier, J. Pagès, Analyses factorielles simples et multiples, 3è édition - Dunod 1998.

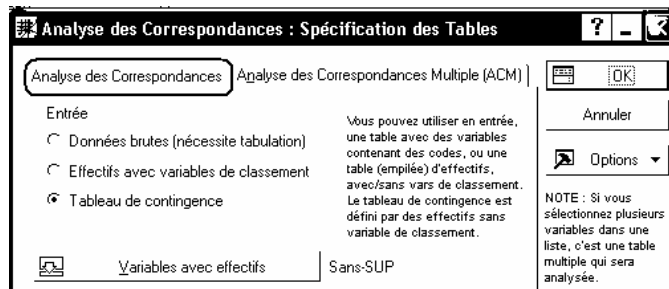
Ces tableaux croisent trois variables qualitatives : l'emploi, le diplôme et le sexe. Les buts de notre étude peuvent être multiples. D'une part, on peut s'intéresser à la liaison entre emploi et diplôme, indépendamment du sexe, et mettre ainsi en évidence une structure commune à ces deux tableaux. D'autre part, et de façon complémentaire, on peut s'intéresser aux écarts entre ces deux tableaux : les répartitions croisées des emplois et des diplômes sont-elles similaires selon le sexe, ou au contraire, sont-elles très différentes ?

2.3.7.1 Première analyse : les tableaux "par sexe" en éléments supplémentaires dans l'AFC de leur somme

Ouvrez le classeur Diplomes-emploi-1973.stw et observez la manière dont les données y ont été saisies. Ouvrez également le classeur Excel Diplomes-emplois-1973.xls.

On va réaliser une AFC sur le tableau "somme", en plaçant en éléments supplémentaires les tableaux relatifs aux données par sexe.

Réalisez une AFC sur les variables 1 à 8 du tableau de données Statistica :



Activez ensuite l'onglet "Points supplémentaires" et cliquez sur le bouton "Ajouter des points lignes". Plutôt que de saisir ces données supplémentaires à la main, copiez, puis collez dans la fenêtre la plage A11:I30 de la feuille "Donnees" du classeur Excel.

Points Ligne Supplémentaires (Diplomes-emplois-1973 dans Diplomes-emplois-1973.stu)

Saisissez les valeurs (effectifs) des nouveaux points supplémentaires puis cliquez sur OK.

Point	Nom du Pt Suppl	Sans	BEPC	BEP/CAP	BACG	BACT	DEUG	DUT	SUP
10	F-Agri	5089	1212	1166	0	0	0	0	0
11	F-Ingé	0	0	0	316	0	0	304	1033
12	F-Tech	281	0	320	320	283	0	683	0
13	F-Ouv Q	7470	1859	4017	1752	657	0	285	0
14	F-Ouv nor	29997	4334	4538	1882	0	0	0	0
15	F-Cadre E	0	0	0	2236	595	911	569	6788
16	F-Cadre Iv	1577	1806	4549	17063	875	4152	,E+4	3991
17	F-Empl Q	21616	19915	32452	16137	5865	1256	3332	1286
18	F-Empl nc	19849	7325	6484	5111	898	294	635	0
19	H-Tous Er	37969	31958	64845	21937	9586	7774	5866	,E+4
20	F-Tous Er	35879	36451	53526	44817	9173	6613	,E+4	,E+4
21	--								
--									

OK Annuler

Après avoir validé, cliquez de même sur "Ajouter des points colonnes" et collez la plage A10:J27 de la feuille Excel "Donnees transposees".

Points Colonne Supplémentaires (Diplomes-emplois-1973 dans Diplomes-emplois-1973.stu)

Saisissez les valeurs (effectifs) des nouveaux points supplémentaires puis cliquez sur OK.

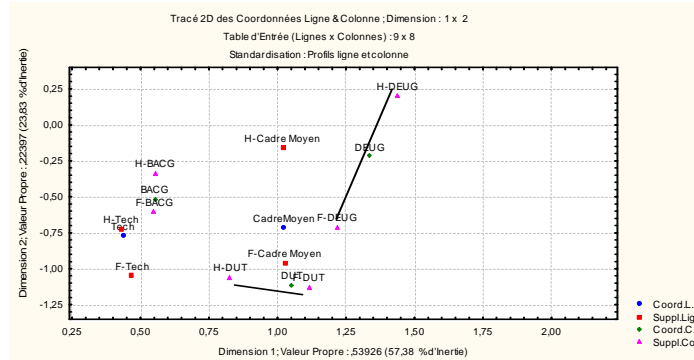
Point	Nom du Pt Suppl	Agri	Ingé	Tech	Ouv Q	Ouv non-Q	Cadre Sup	Cadre Moyen	Empl
1	H-Sans ,E+4	0	302	10143	59394	596	2142	54	
2	H-BEPC	2701	337	1697	3702	8087	298	2801	73
3	H-BEP/CAP	5709	309	2242	30926	17862	892	672	47
4	H-BACG	297	917	1969	314	2887	1227	6495	43
5	H-BACT	1242	0	1399	1861	1696	298	924	12
6	H-DEUG	0	308	357	0	0	2362	2807	6
7	H-DUT	322	0	1943	0	0	318	2301	9
8	H-SUP	0	4383	381	337	323	6781	4030	
9	F-Sans	5089	0	281	7470	29997	0	1577	216
10	F-BEPC	1212	0	0	1859	4334	0	1806	199
11	F-BEP/CAP	1166	0	320	4017	4538	0	4549	324
12	F-BACG	0	316	320	1752	1882	2236	17063	161

OK Annuler

Poursuivez ensuite l'exécution de l'ACP : valeurs propres, coordonnées lignes et colonnes, graphiques des points lignes, des points colonnes et graphique lignes et colonnes.

Pour interpréter les résultats trouvés, on commence par s'intéresser aux individus lignes et colonnes actifs. Ici, le premier axe classe les emplois et les diplômes en plaçant sur la partie gauche de l'axe "Sans diplôme" et les emplois peu qualifiés et sur la partie droite les diplômes "supérieurs" et les emplois d'ingénieurs et cadres supérieurs. Le second axe oppose les diplômes et emplois "moyens" (techniciens, cadres moyens, Bac, DEUG), qui occupent la partie négative de l'axe aux diplômes et emplois "extrêmes" (emplois non qualifiés, cadres supérieurs, sans diplôme, études supérieures) sur la partie positive de l'axe. Cette configuration est classique lorsque l'ACP s'applique à des variables ordinales, et porte le nom d'effet Guttman.

Pour étudier les points lignes et points colonnes supplémentaires, on compare leur position à celle du point correspondant du tableau "somme" :



Le point DEUG, par exemple, est situé à la moyenne pondérée des points H-DEUG et F-DEUG. Comme les effectifs masculins et féminins pour le DEUG sont sensiblement équivalents, ce point se trouve approximativement au milieu du segment (H-DEUG, F-DEUG).

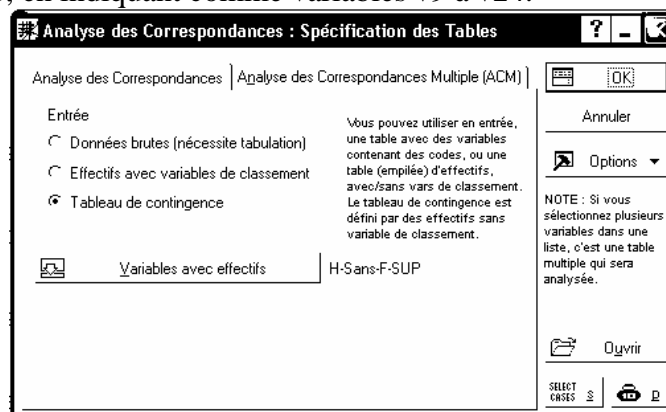
On constate que, sur le premier axe, pour tous les diplômes, les deux points représentant les hommes et les femmes sont presque confondus. En revanche, pour les points relatifs au DEUG par exemple, la différence des coordonnées sur le deuxième axe est très importante. D'une manière générale, on constate que, s'agissant des diplômes, les points relatifs aux femmes ont généralement une coordonnée sur l'axe 2 inférieure à celle du correspondant relatif aux hommes : les femmes occupent, plus que les hommes, les emplois "moyens". Inversement, les hommes sont plus nombreux à occuper des emplois "extrêmes".

Deux remarques méritent d'être faites

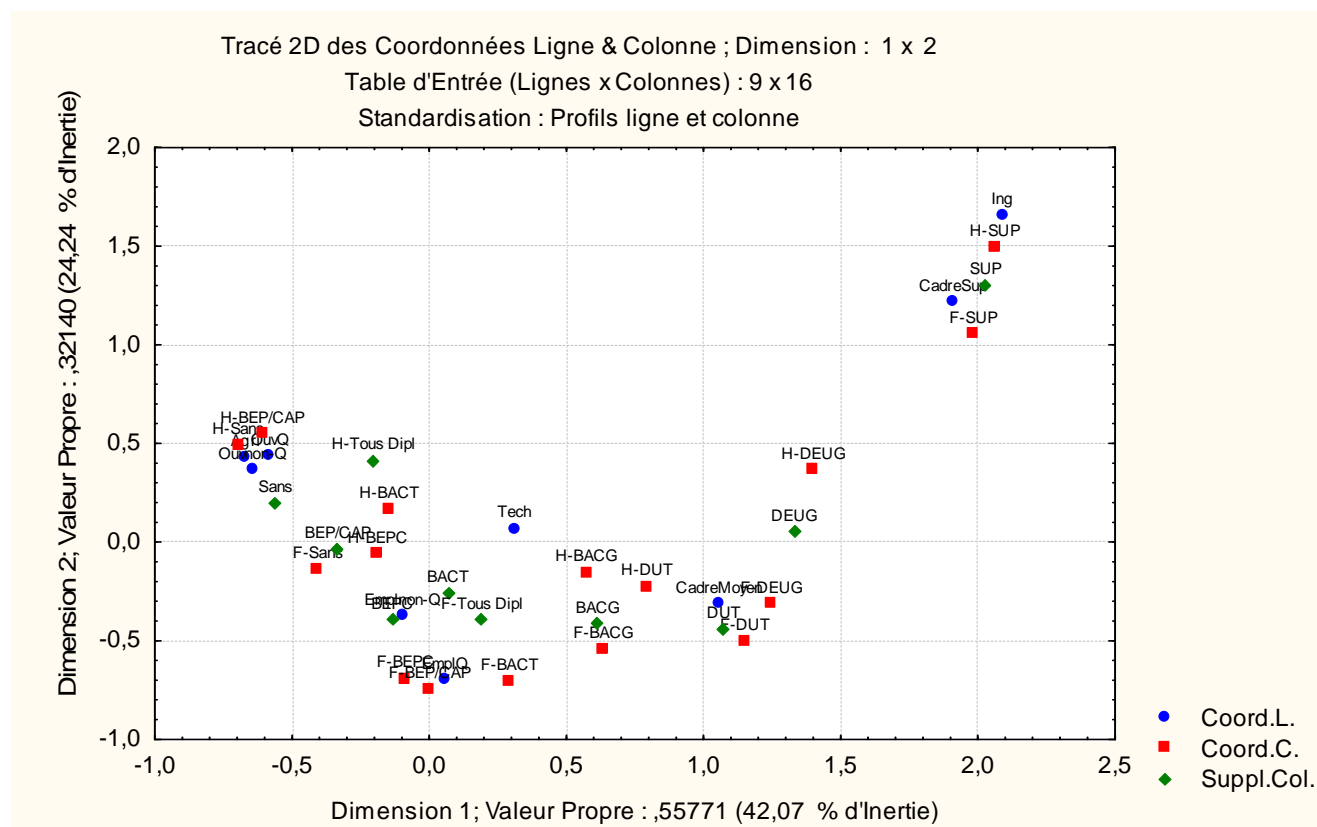
- Dans l'étude menée ici, l'inertie prise en compte ($\Phi^2 = 0,94$) est celle du tableau "somme". On ne tient donc pas compte de la dispersion des données due aux discriminations liées au sexe.
- Deux points supplémentaires correspondant aux deux sexes peuvent être représentés proches l'un de l'autre sur le graphique, alors qu'il existe une forte disparité entre hommes et femmes pour cette modalité, et nous disposons de peu de moyens pour le mettre en évidence. Ce type de situation se produit lorsque la dispersion "entre sexes" est orthogonale à la dispersion due aux autres deux autres facteurs.

2.3.7.2 Deuxième analyse : tableaux "par sexe" juxtaposés et tableau "somme" en éléments supplémentaires.

Réalisez une autre AFC, en indiquant comme variables v9 à v24.



Ajoutez comme points colonnes supplémentaires des données relatives au tableau somme et à la synthèse des emplois par sexe, c'est-à-dire les plages A2:J9 et A26:J27 de la feuille Excel "Donnees transposees".



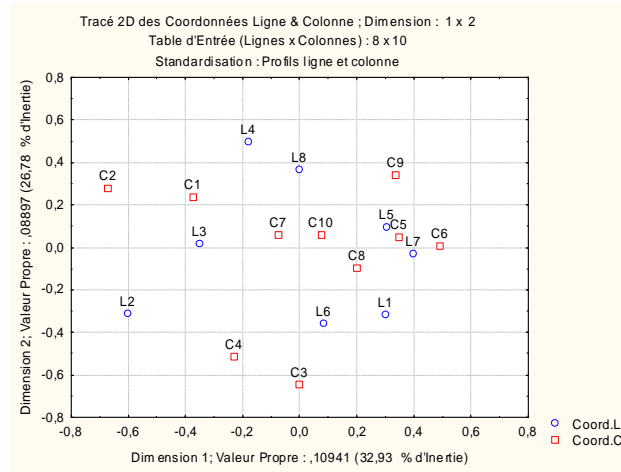
L'interprétation du graphique se fait comme précédemment. Cependant, l'inertie du nuage de points fait maintenant intervenir les points relatifs aux diplômes par sexe. On constate que le coefficient Phi-2 est plus élevé que dans l'étude précédente : 1,326 au lieu de 0,940. La différence entre les deux coefficients représente l'inertie "intra" (dispersion liée au sexe, pour chaque diplôme), qui représente ici presque 30% du total. Une étude plus poussée (dont les détails sortent du cadre de ce cours) permettrait de montrer que cette inertie intra est très faible sur le premier axe, mais représente presque la moitié de l'inertie du deuxième axe. Ainsi cette méthode permet, dans une certaine mesure, d'évaluer l'importance des écarts des colonnes homologues aux colonnes moyennes.

2.3.8 Quelques configurations remarquables dans les résultats produits par une AFC.

On pourra consulter le fichier [Configurations-Types.stw](#) qui rassemble quelques configurations classiques de nuages, générées à partir de données fictives.

2.3.8.1 Forme générale du nuage

L'inertie totale (le Phi-2) est un indicateur de la dispersion totale du nuage. La comparaison des inerties de chacun des axes (c'est-à-dire des valeurs propres associées aux axes) renseigne sur la forme du nuage de points. Si les premières valeurs propres sont proches les unes des autres, la dispersion est relativement homogène : il n'y a pas vraiment de direction privilégiée et le nuage de points est approximativement sphérique. Si au contraire, les valeurs propres sont nettement différentes, cela traduit un nuage de points fortement allongé selon une (ou plusieurs) direction.

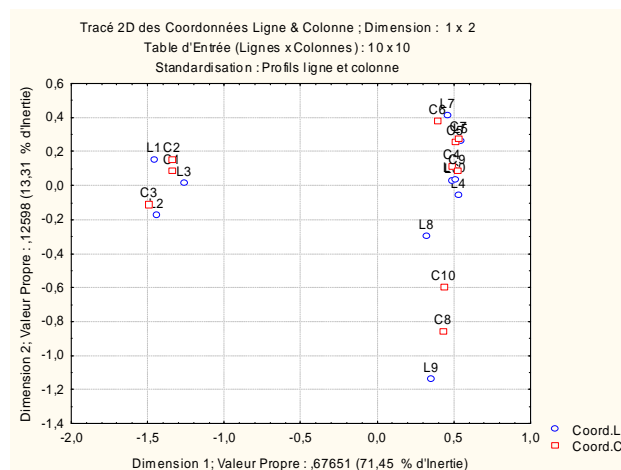


2.3.8.2 Deux paquets de points - Valeurs propres proches de 1

Les valeurs propres sont toutes inférieures à 1. Mais, une valeur propre proche de 1 indique une dichotomie des données, c'est-à-dire un tableau de contingence qui, après reclassement des modalités, aurait l'allure suivante :

	0
0	

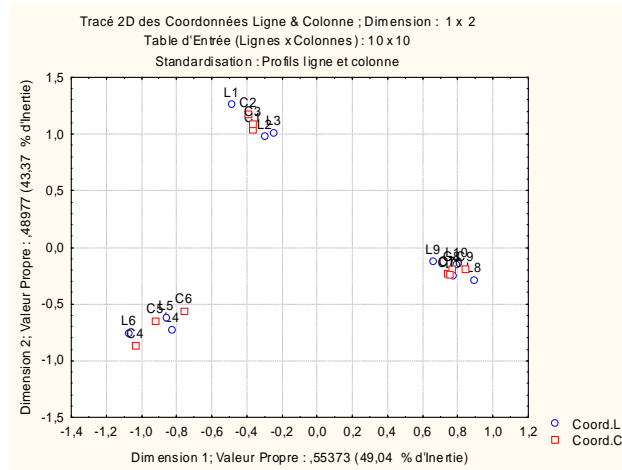
Le nuage est alors divisé en deux paquets de points. La feuille de données "Deux-paquets" fournit une illustration de cette situation.



2.3.8.3 Trois paquets de points

De même, l'existence de deux valeurs propres proches de 1 indique une partition des observations en 3 groupes. Si toutes les valeurs propres sont proches de 1, cela indique une correspondance entre chaque modalité ligne et une modalité colonne "associée". Avec une réorganisation convenable des modalités, les effectifs importants se trouvent alors le long de la diagonale.

La feuille de données "Trois-paquets" fournit une illustration de cette situation.

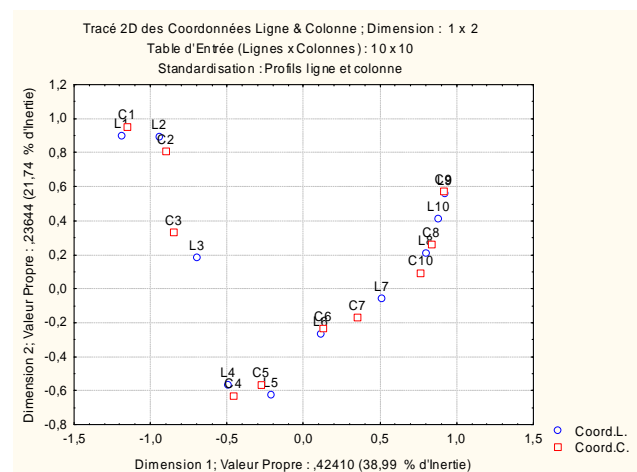
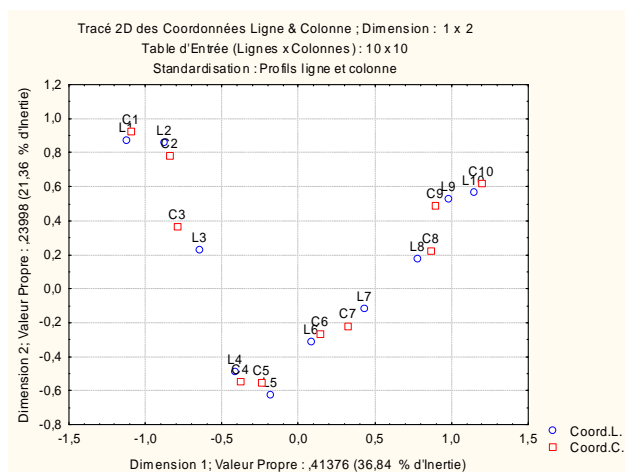


2.3.8.4 L'effet Guttman.

Un nuage de points de forme parabolique indique une redondance entre les deux variables étudiées : la connaissance de la ligne i donne pratiquement celle de la colonne j . Dans un tel cas, pratiquement toute l'information est contenue dans le premier facteur. Cette configuration se rencontre notamment lorsque les deux variables sont ordinales, et classent les sujets de la même façon. Dans ce cas, le premier axe oppose les valeurs extrêmes et classe les valeurs, tandis que le deuxième axe oppose les intermédiaires aux extrêmes.

La feuille de données "Effet-Guttman" fournit une illustration assez caractéristique de cette situation. Dans ce cas, on a intérêt à ne pas limiter l'étude au plan (1, 2). La configuration-type dans les trois plans de projection définis par les 3 premiers axes prend souvent les allures indiquées dans l'exemple.

Il pourra alors être intéressant d'examiner les accidents des courbes qui joignent les points, qui reflètent les particularités des situations étudiées. Voir par exemple la situation des modalités L10 et C10 dans l'exemple "Guttman-perturbé".



2.3.8.5 Nuage tétraédrique

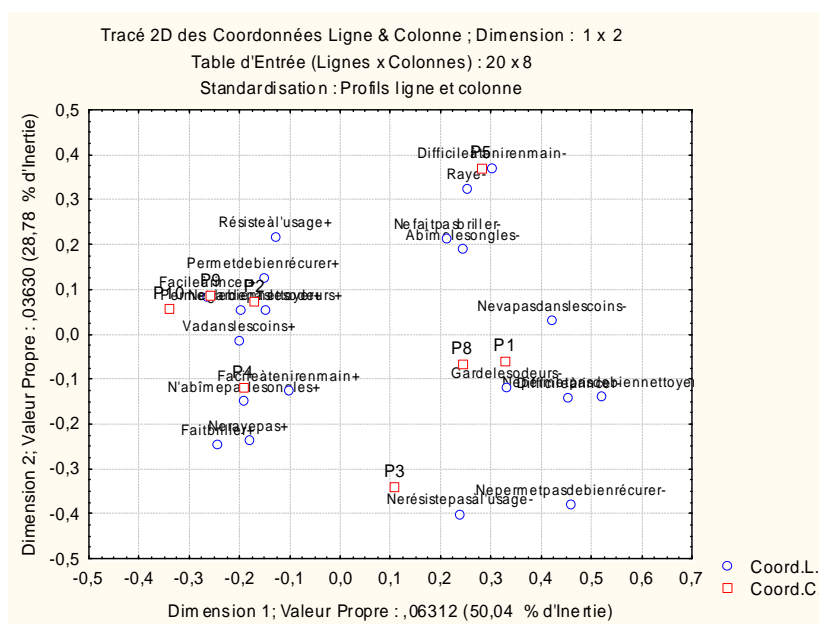
Le premier exemple ("Deux-paquets") est également caractéristique d'une forme classique de nuage : tétraédrique, ou en forme de "berlingot" comme on peut s'en rendre compte en construisant les projections du nuage sur les 3 premiers axes.

2.3.9 L'extension de la notion de tableau de contingence

En toute rigueur, l'analyse de correspondances ne s'applique qu'aux tableaux de contingence. Elle peut cependant être appliquée à des tableaux qui, a priori, ne sont pas des tableaux de contingence. Un critère essentiel pour décider si un tableau peut être assimilé à un tableau de contingence est le suivant : on doit pouvoir donner un sens à la somme des cases du tableau, qu'elle soit faite par ligne ou par colonne.

2.3.9.1 Tableaux juxtaposés

Considérons l'exemple fourni dans le classeur Echelles-Likert.stw. On obtient ainsi un point par produit et deux points par échelle bipolaire. On peut facilement montrer que le barycentre (pondéré) des deux points correspondant à une échelle donnée se trouve au centre de gravité du nuage. Si le point "+" se trouve plus près de l'origine que le point "-", cela signifie que l'intensité de la propriété positive est supérieure à celle de la propriété négative correspondante. Cet effet est connu sous le nom d'effet de levier.



Dans certains cas, on peut juxtaposer, par exemple, deux tableaux de contingence correspondant à des dates différentes, par exemple la ventilation de la population française par région et par CSP pour deux recensements différents. Il sera alors pertinent d'étudier comment chaque modalité s'est déplacée entre l'époque 1 et l'époque 2.

2.3.9.2 Juxtaposer plusieurs tableaux : vers l'ACM

Source : Hahn A., Eirmbter W. H., Jacob R., Le sida : savoir ordinaire et insécurité, traduction française de Herrmann M.

Il s'agit d'une enquête réalisée durant l'été 1990, auprès d'un échantillon représentatif des ménages de RFA.

Résumé du questionnaire :

Variable	Modalité	Codage
Sexe	masculin	m
	féminin	f

Confession	protestant	ev
	catholique	rk
	autre	an
	sans	ke
Liens avec l'église	forts	f1
	moyens	f2
	inexistants	f3
Catégorie Sociale	élèves/étud	s1
	classe sup.	s2
	cl. moy. sup.	s3
	cl. moyenne	s4
	cl. moy. inf.	s5
	cl. populaire	s6
	autres	s7
Taille du lieu de résidence	< 2	k1
	2 à < 5	k2
	5 à < 20	k3
	20 à < 50	k4
	50 à < 100	k5
	100 à < 500	k6
	> 500	k7
Classe d'âge	18 à < 30	a1
	30 à < 40	a2
	40 à < 50	a3
	50 à < 60	a4
	60 et plus	a5
Fidélité dans les rapports sexuels	très pour	t1
	plutôt pour	t2
	indécis	t3
	plutôt contre	t4
	très contre	t5
Plusieurs partenaires	oui	p1
	non	p2
Préférences politiques	CDU/CSU	cd
	SPD	sp
	FDP	fd
	Verts	gr
Nombre de situations jugées contaminantes	0	w0
	1	w1
	2	w2
	3	w3
	4	w4
	5	w5
	6	w6

7	w7
8	w8

Le sida est la conséquence d'une faute et d'une punition

très pour	c1
plutôt pour	c2
indécis	c3
plutôt contre	c4
très contre	c5

Dispositions d'évitement et d'expulsion des contaminés de la sphère personnelle

très pour	m1
plutôt pour	m2
indécis	m3
plutôt contre	m4
très contre	m5

Nombre de mesures obligatoires acceptées

0	z0
1	z1
2	z2
3	z3
4	z4
5	z5

Nombre de situations en public jugées dangereuses

0-1	o1
2	o2
3	o3
4	o4
5-6	o5

Le sida est un péril omniprésent

d'accord	g1
indécis	g2
pas d'accord	g3

Ouvrez le classeur Hahn.stw et observez la façon dont a été constitué le tableau de contingence : la variable "groupe" est croisée avec toutes les autres variables, et on juxtapose ainsi 14 tableaux de contingence portant sur des populations presque identiques (*presque*, car pour la plupart des questions, il y a quelques non-réponses).

Réalisez une analyse des correspondances sur ce tableau et retrouvez ainsi les résultats de l'auteur :

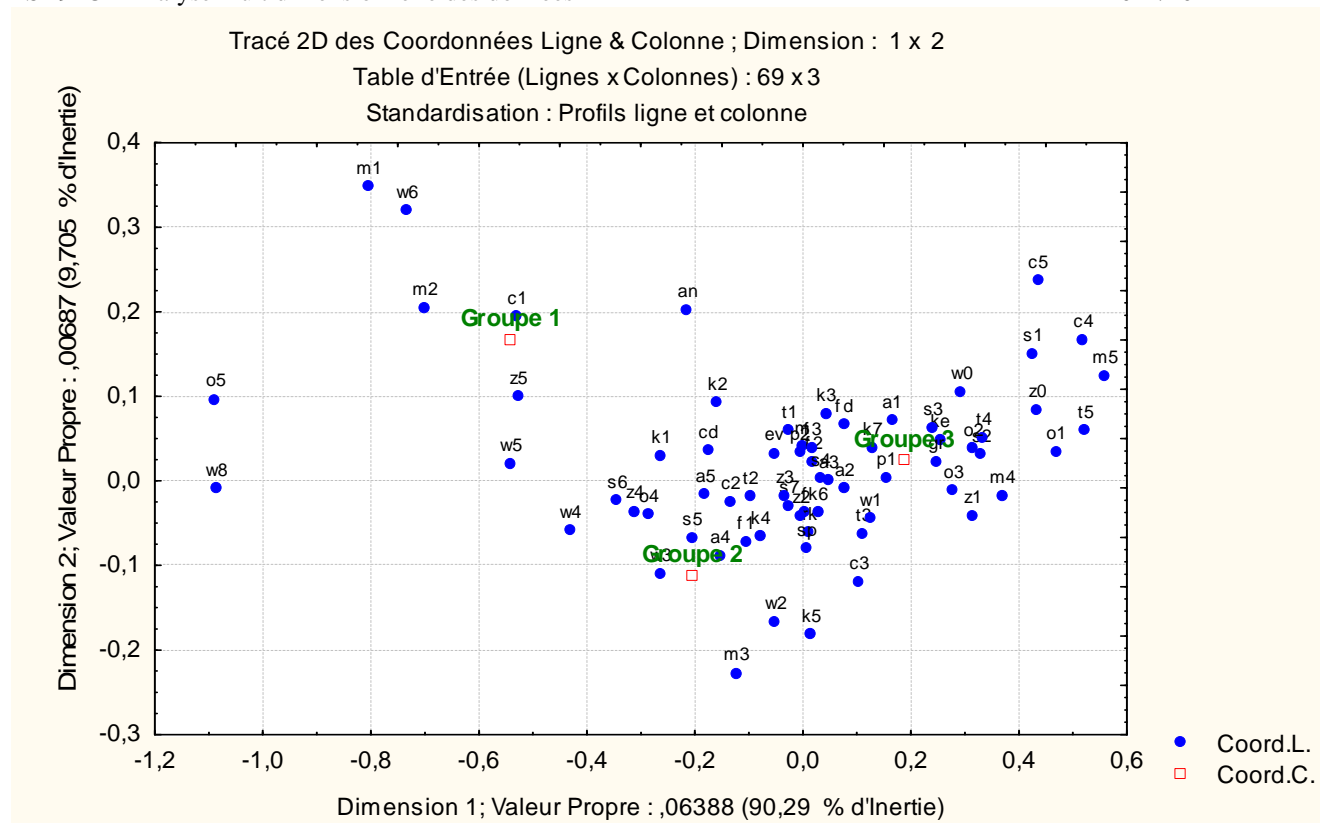
"L'analyse des correspondances confirme l'existence de deux syndromes nettement distincts, attribuables, avec la prudence qui s'impose, à deux catégories ou milieux, qu'à la suite de Schulze on pourrait appeler "milieu harmoniste" et "milieu autodéterministe".

Notre analyse utilise la dangerosité ressentie du sida comme la variable à décrire, les autres caractéristiques servant d'indices de cette appréciation. Etant donné les trois configurations de la variable à décrire, une solution bidimensionnelle serait théoriquement possible. Mais, puisque le premier axe d'inertie rend compte de 90,25% de la variation, nous négligerons ce deuxième axe.

Graphique et tableau numérique montrent que la vision du sida comme péril a été reportée sur l'ordonnée. On distingue nettement deux groupes, qui approuvent ou rejettent les termes de la question. Ceux qui ne se prononcent pas se situent entre les deux, mais sont enclins le cas échéant à considérer le sida comme une maladie omniprésente et très infectieuse.

À cela correspond la localisation des indicateurs de dispositions (perceptions, réactions) et des repères de morphologie sociale. Les enquêtés considérant le sida comme un péril le jugent très infectieux jusque dans la vie quotidienne (3 situations courantes ou plus jugées contaminantes par un taux supérieur à la moyenne). La maladie est ressentie comme conséquence et punition d'une faute morale; les dispositions d'exclusion se manifestent nettement, et les mesures obligatoires antisida - y compris la généralisation du test obligatoire - rencontrent un taux d'adhésion supérieur à la moyenne. Ceci vérifie nos hypothèses de départ : poussée à l'extrême, la conception du sida comme danger permanent de contamination fait considérer comme porteurs de virus potentiels non seulement les membres des principaux groupes à risque.(donc une minorité), mais tous les étrangers. Les mêmes enquêtés ressentent la sphère publique comme généralement inquiétante et hostile. Leurs opinions politiques plutôt conservatrices sont attestées par une préférence très nette pour les partis CDU/CSU. Ce groupe comprend une proportion importante de personnes âgées, de niveau social peu élevé, résidant plutôt dans des communes petites ou très petites.

A l'inverse, ceux pour qui le sida n'est pas un péril au sens indiqué ci-dessus, ont pour caractéristique commune de ne pas chercher un risque de contamination là où, en l'état actuel des connaissances, un tel risque n'existe pas. On n'envisage guère la maladie en termes de culpabilité, et on réclame rarement l'exclusion des contaminés ou l'adoption de mesures répressives. Or, ces personnes sont objectivement plus exposées à la contamination.: la fidélité sexuelle est jugée relativement moins importante, le changement de partenaire est relativement fréquent. Les considérations éthico-religieuses passent à l'arrière-plan, la proportion des personnes sans confession est relativement élevée. Politiquement, ce segment se situe majoritairement à gauche du centre, avec une préférence marquée pour les Verts. Morphologiquement, il s'agit d'une population plutôt jeune, étudiante, de niveau social élevé et majoritairement citadine."



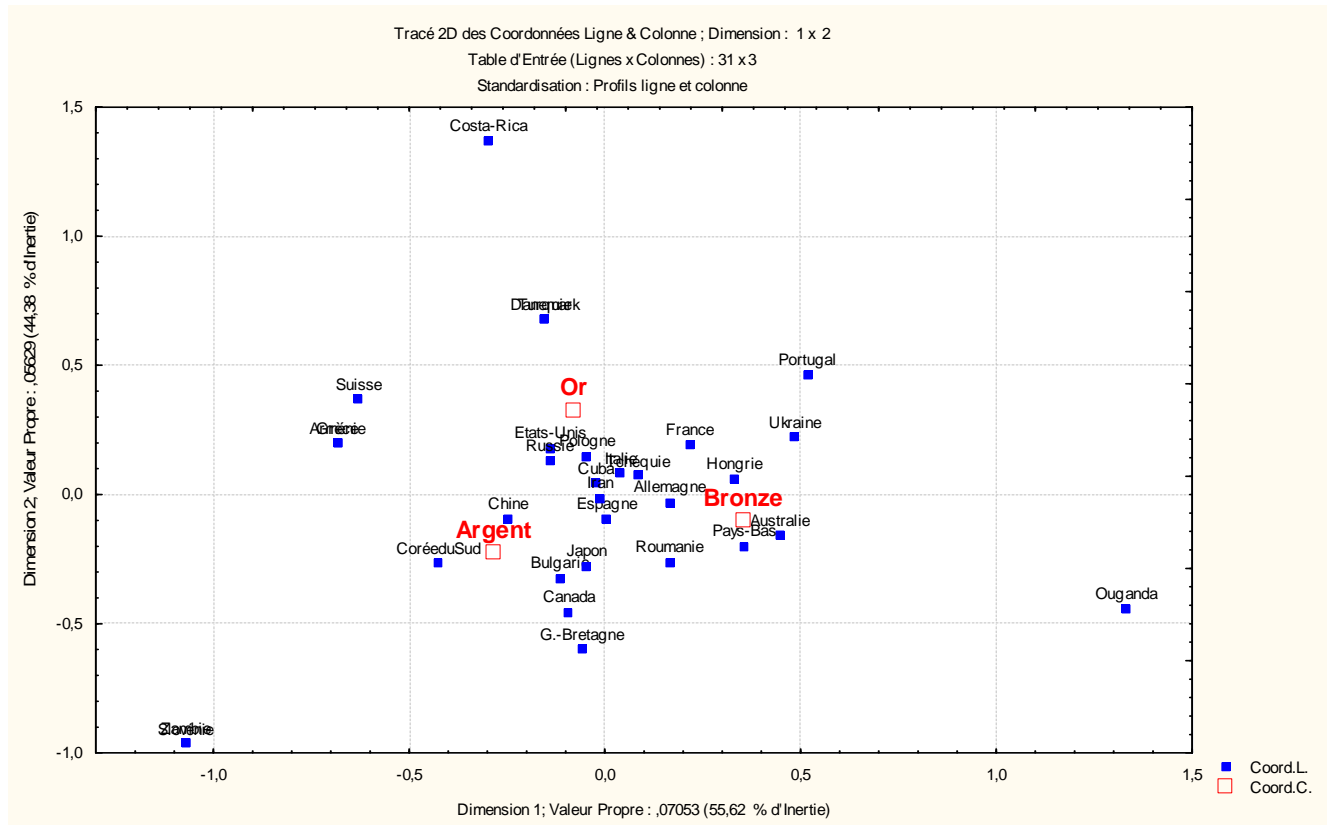
2.3.10 Conclusion

Au travers des exemples qui précèdent, on voit que l'AFC peut être utilisée dans des situations variées. En revanche, comme l'indique Philippe Cibois dans son article "les pièges de l'AFC", il existe des situations où il vaut mieux s'abstenir d'utiliser cette méthode :

- L'AFC mettra toujours en évidence des attractions - répulsions entre modalités lignes et modalités colonne. Mais, lorsqu'on travaille sur un échantillon et que le khi-2 du tableau de contingence n'est pas significatif, l'effet mis en évidence n'est que le fruit du hasard.
- L'AFC n'a d'intérêt que si notre étude porte sur les liaisons existant entre lignes et colonnes. Comme exemple, caricatural, d'un mauvais usage de l'AFC, P. Cibois fournit le tableau suivant qui donne, pour 30 pays, le nombre de médailles d'or, d'argent et de bronze obtenues aux Jeux Olympiques d'Atlanta :

	Or	Argent	Bronze		Or	Argent	Bronze
Etats-Unis	44	32	25	Grèce	4	4	0
Russie	26	21	16	Tchéquie	4	3	4
Allemagne	20	18	27	Suisse	4	3	0
Chine	16	22	12	Danemark	4	1	1
France	15	7	15	Turquie	4	1	1
Italie	13	10	12	Canada	3	11	8
Australie	9	9	23	Bulgarie	3	7	5
Cuba	9	8	8	Japon	3	6	5
Ukraine	9	2	12	G.-Bretagne	1	8	6
Corée du Sud	7	15	5	Iran	1	1	1
Pologne	7	5	5	Arménie	1	1	0
Hongrie	7	4	10	Portugal	1	0	1
Espagne	5	6	6	Costa-Rica	1	0	0
Roumanie	4	7	9	Slovénie	0	2	0
Pays-Bas	4	5	10	Zambie	0	1	0

Le pays le plus "attiré" par la modalité "Or" est le Costa-Rica, qui n'a obtenu qu'une seule médaille, mais en or, alors que des pays tels que Cuba et l'Iran, avec des palmarès très différents, sont représentés proches l'un de l'autre, au voisinage de l'origine. En effet, les résultats de l'AFC ne concernent pas le nombre de médailles obtenues par les différents pays, mais l'écart entre les proportions de médailles de bronze, argent, or obtenues par le pays et la distribution totale (environ 1/3 de médailles de chaque type). Mais cet écart constitue-t-il vraiment un sujet d'étude ?



2.4 Analyse des Correspondances Multiples

2.4.1 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. 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.

2.4.2 Forme des données d'entrée

Selon leur origine, les données sur lesquelles nous nous proposons de faire une ACM peuvent se présenter sous différentes formes.

Imaginons, par exemple, une mini-enquête dans laquelle nous avons posé trois questions à 10 sujets : le sexe (F ou H), le niveau de revenus (M : modeste, E : élevé) et leur préférence sur un sujet donné (3 modalités : A, B ou C). Les données peuvent se présenter sous l'une des formes décrites ci-dessous. Le classeur Mini-ACM.stw contient 5 feuilles de données correspondant à ces 5 formes.

2.4.2.1 Tableau protocole

	1	2	3
	Sexe	Revenu	Preference
s1	F	M	A
s2	F	M	A
s3	F	E	B
s4	F	E	C
s5	F	E	C
s6	H	E	C
s7	H	E	B
s8	H	M	B
s9	H	M	B
s10	H	M	A

2.4.2.2 Tableau d'effectifs

	1	2	3	4
	Sexe	Revenu	reference	Effectif
1	F	M	A	2
2	F	E	B	1
3	F	E	C	2
4	H	E	C	1
5	H	E	B	1
6	H	M	B	2
7	H	M	A	1

2.4.2.3 Tableau disjonctif complet

Le tableau disjonctif complet ou TDC 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 ou non.

	1	2	3	4	5	6	7
	Sexe:F	Sexe:H	Rev:M	Rev:E	Pref:A	Pref:B	Pref:C
s1	1	0	1	0	1	0	0
s2	1	0	1	0	1	0	0
s3	1	0	0	1	0	1	0
s4	1	0	0	1	0	0	1
s5	1	0	0	1	0	0	1
s6	0	1	0	1	0	0	1
s7	0	1	0	1	0	1	0
s8	0	1	1	0	0	1	0
s9	0	1	1	0	0	1	0
s10	0	1	1	0	1	0	0

2.4.2.4 Tableau disjonctif des patrons

En regroupant les lignes identiques du tableau disjonctif complet, on obtient le tableau disjonctif des patrons :

	1	2	3	4	5	6	7
	Sexe:F	Sexe:H	Rev:M	Rev:E	Pref:A	Pref:B	Pref:C
FMA	2	0	2	0	2	0	0
FEB	1	0	0	1	0	1	0
FEC	2	0	0	2	0	0	2
HEC	0	1	0	1	0	0	1
HEB	0	1	0	1	0	1	0
HMB	0	2	2	0	0	2	0
HMA	0	1	1	0	1	0	0

2.4.2.5 Tableau de Burt

L'ACM peut également être réalisée à partir d'une structuration particulière des données, appelée tableau de Burt (TdB). Ce dernier tableau 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. Le tableau de Burt apparaît ainsi comme une juxtaposition de tableaux de contingence des variables prises deux à deux.

Table Observée (Effectifs) (Protocole dans Classeur2)							
Table d'Entrée (Lignes x Colonnes) : 7 x 7 (Table de Burt)							
	F	H	M	E	A	B	C
Sexe:F	5	0	2	3	2	1	2
Sexe:H	0	5	3	2	1	3	1
Revenu:M	2	3	5	0	3	2	0
Revenu:E	3	2	0	5	0	2	3
Preference:A	2	1	3	0	3	0	0
Preference:B	1	3	2	2	0	4	0
Preference:C	2	1	0	3	0	0	3

On peut noter qu'il est possible, sans grand problème de passer de l'une des 4 premières structures de données à une autre. De même, le TdB peut être obtenu facilement à partir du tableau disjonctif complet. En revanche, il n'existe pas de moyen simple pour recomposer l'une des 4 premières structures de données à partir du tableau de Burt.

2.4.3 Quelques règles d'interprétation

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.

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.

Nous pouvons, comme en AFC, nous intéresser aux profils ligne et colonne, aux taux de liaison et au Φ^2 du tableau disjonctif complet, vu comme un tableau de contingence. Le nombre de lignes de ce tableau est égal au nombre d'individus statistiques étudiés. Cependant, nous avons vu que la métrique du Φ^2 , utilisée pour l'AFC, possède la propriété d'équivalence distributionnelle : si on regroupe deux lignes correspondant au même patron de réponses, on ne change rien aux autres profils lignes, ni aux autres profils colonnes. Autrement dit, on retrouvera les mêmes résultats en effectuant une AFC sur le tableau disjonctif des patrons.

Comme en AFC, on peut calculer des fréquences, des fréquences lignes, des fréquences colonnes et des profils lignes et profils colonnes moyens.

L'élément le plus facile à interpréter est le profil colonne moyen : ce sont les fréquences des différents patrons de réponses dans la population étudiée.

L'élément le plus facile à interpréter est le profil ligne moyen : ce sont les fréquences des différents patrons de réponses dans la population étudiée.

Le profil ligne moyen est obtenu en calculant, pour chaque modalité, le quotient de sa fréquence par le nombre Q de questions. En notant respectivement n_k et f_k l'effectif et la fréquence de la modalité k , on a :

$$f_k = \frac{n_k}{N} = \frac{\text{Nombre d'individus ayant choisi la modalité } k}{\text{Nombre total d'individus}}$$

et le k-ième élément du profil-ligne moyen est :

$$f_{\bullet k} = \frac{f_k}{Q} = \frac{n_k}{QN} = \frac{\text{Nombre d'individus ayant choisi la modalité } k}{\text{Nombre de questions} \times \text{Nombre total d'individus}}$$

N.B. Ici, f_k et $f_{\bullet k}$ désignent des quantités différentes : f_k est la fréquence de la modalité k dans la population étudiée; $f_{\bullet k}$ est définie comme pour l'AFC, fréquence ligne marginale de la k -ième colonne du tableau disjonctif des patrons.

2.4.3.1 Taux de liaison et Phi-2

Pour le tableau disjonctif complet, ou le tableau disjonctif des patrons, considérés comme des tableaux de contingence, le coefficient Phi-2 vaut :

$$\Phi^2 = \frac{K - Q}{Q} = \frac{\text{Nombre de modalités} - \text{Nombre de questions}}{\text{Nombre de questions}}$$

où K désigne le nombre de modalités et Q le nombre de questions

Dans notre exemple, on a : $K=7$, $Q=3$, et donc : $\Phi^2 = \frac{7}{3} - 1 = 1,33$.

Ce coefficient représente l'inertie totale du nuage de points des modalités colonnes. On montre que l'inertie absolue de chacune des questions est donnée par :

$$I(X_q) = \frac{K_q - 1}{Q}$$

où K_q représente le nombre de modalités de la question q .

L'inertie relative de chacune des questions est donnée par :

$$\text{Inr}(X_q) = \frac{K_q - 1}{K - Q} = \frac{\text{Nb de modalités de la question} - 1}{\text{Nb total de modalités} - \text{Nb de questions}}$$

Sur notre exemple, on a, pour l'inertie absolue :

$$I(\text{Sexe}) = I(\text{Revenu}) = \frac{2-1}{3} = 0,33$$

$$I(\text{Pref}) = \frac{3-1}{3} = 0,67$$

Quant aux inerties relatives :

$$\text{Inr}(\text{Sexe}) = \text{Inr}(\text{Revenu}) = \frac{2-1}{4} = 25\%$$

$$\text{Inr}(\text{Pref}) = \frac{3-1}{4} = 50\%$$

L'inertie d'une question est ainsi directement liée au nombre de ses modalités : on évitera donc d'utiliser la méthode lorsque les différentes questions présentent des nombres de modalités trop différents.

2.4.3.2 Distances entre profils lignes

En AFC, nous avons donné les formules permettant de calculer les distances entre deux profils lignes ou entre deux profils colonnes. La distance utilisée est la *métrique du Φ^2* . Ici, compte tenu de la structure particulière du tableau de contingence utilisé, les formules indiquées deviennent :

$$d_{\Phi^2}^2(L_i, L_{i'}) = \frac{1}{Q} \sum_k \frac{(\delta_{ik} - \delta_{i'k})^2}{f_k}$$

Notations utilisées : L_i et $L_{i'}$ désignent deux patrons, Q est le nombre de questions. δ_{ik} prend la valeur 1 si la modalité k fait partie du patron i , et la valeur 0 sinon. Enfin, f_k est la fréquence de la modalité k dans la population.

Cette formule montre que deux individus (ou deux patrons) sont d'autant plus éloignés que leurs réponses diffèrent pour un plus grand nombre de questions et pour des modalités rares. Cette formule peut encore être écrite sous la forme :

$$d_{\Phi^2}^2(\text{Patron } i, \text{Patron } i') = \frac{1}{\text{Nb de Questions}} \sum \frac{1}{\text{fréquence de la modalité } k}$$

où la somme est étendue à toutes les modalités faisant partie de l'un des deux patrons, sans faire partie des deux patrons.

Autrement dit, deux individus (ou deux patrons) sont d'autant plus éloignés que leurs réponses diffèrent pour un plus grand nombre de questions et pour des modalités rares.

Ainsi, sur notre exemple :

$$d_{\Phi^2}^2([\text{FMA}], [\text{HMA}]) = \frac{1}{3} \left(\frac{1}{0,5} + \frac{1}{0,5} \right) = 1,33$$

La distance d'un patron au profil ligne moyen est :

$$d_{\Phi^2}^2(O, L_i) = \left(\frac{1}{Q} \sum_k \frac{\delta_{ik}}{f_k} \right) - 1$$

Autrement dit, un patron sera d'autant plus loin de l'origine qu'il fait intervenir des modalités plus rares. On peut aussi écrire cette formule sous la forme :

$$d_{\Phi^2}^2(O, \text{Patron } i) = \left(\frac{1}{\text{Nombre de Questions}} \sum \frac{1}{\text{fréquence de la modalité } k} \right) - 1$$

où la somme est étendue à toutes les modalités faisant partie du patron i .

Par exemple :

$$d_{\Phi^2}^2(O, [\text{FMA}]) = \left(\frac{1}{3} \left(\frac{1}{0,5} + \frac{1}{0,5} + \frac{1}{0,3} \right) \right) - 1 = 1,44$$

La contribution (absolue) d'un patron à la variance du nuage est obtenue en multipliant la distance précédente par la fréquence du patron dans la population.

2.4.3.3 Distances entre profils colonnes

La distance entre les modalités k et k' est donnée par :

$$d_{\Phi^2}^2(M_k, M_{k'}) = \frac{1}{f_k} + \frac{1}{f_{k'}} - 2 \frac{f_{kk'}}{f_k f_{k'}} = \frac{n_k + n_{k'} - 2n_{kk'}}{n_k n_{k'} / n}$$

où $f_{kk'}$ est la fréquence de la combinaison de modalités k et k' , ou encore :

$$d_{\Phi^2}^2(M_k, M_{k'}) = \frac{\text{Effectif de } k + \text{Effectif de } k' - 2 \times \text{Effectif de la combinaison } k \& k'}{\text{Effectif de } k \times \text{Effectif de } k' / \text{Effectif total}}$$

Deux modalités sont d'autant plus éloignées qu'elles sont de fréquences faibles et rarement rencontrées simultanément.

Exemple :

$$d_{\Phi^2}^2(\text{Sexe : F, Revenu : M}) = \frac{1}{0,5} + \frac{1}{0,5} - 2 \frac{0,2}{0,5 \times 0,5} = \frac{5+5-2 \times 2}{5 \times 5 / 10} = 2,4$$

La distance d'une modalité au profil colonne moyen est donnée par :

$$d_{\Phi^2}^2(O, M_k) = \frac{1}{f_k} - 1 = \frac{n}{n_k} - 1 = \frac{\text{Effectif total}}{\text{Effectif de } k} - 1$$

Autrement dit, une modalité sera d'autant plus éloignée du profil moyen que sa fréquence est faible. Afin d'éviter que quelques modalités très rares ne prennent une importance excessive dans les résultats obtenus, il sera nécessaire de regrouper les modalités de fréquence trop faible (fréquence inférieure à 5% par exemple).

Exemple :

$$d_{\Phi^2}^2(O, \text{Pref : B}) = \frac{1}{0,4} - 1 = \frac{10}{4} - 1 = 1,5$$

La contribution absolue d'une modalité à la variance du nuage de points est :

$$Cta(M_k) = \frac{1 - f_k}{Q}$$

La contribution relative d'une modalité à la variance du nuage de points est :

$$Ctr(M_k) = \frac{1 - f_k}{K - Q}$$

Exemples :

$$Ctr([\text{Sexe : F}]) = \frac{1 - 0,5}{4} = 12,5\%$$

$$Ctr([\text{Pref : A}]) = \frac{1 - 0,3}{4} = 17,5\%$$

2.4.4 Résultats de l'ACM sur l'exemple

Le tableau des valeurs propres est donné par :

Valeurs Propres et Inertie de toutes les Dimensions (Protocole dans Mini-ACM-v7.stw)					
Table d'Entrée (Lignes x Colonnes) : 7 x 7 (Table de Burt)					
Inertie Totale = 1,3333					
Nombre de Dims.	ValSing.	ValProp.	%age Inertie	%age Cumulé	Chi ²
1	0,7764	0,6028	45,2128	45,2128	25,3794
2	0,6810	0,4637	34,7781	79,9909	19,5221
3	0,4505	0,2030	15,2219	95,2128	8,5446
4	0,2526	0,0638	4,7872	100,0000	2,6872

Taux d'inertie modifiés

La décroissance des valeurs propres est en général très lente. Pour déterminer le nombre d'axes factoriels à conserver, Benzécri a proposé de calculer des taux d'inertie modifiés en utilisant la méthode suivante.

La somme des valeurs propres est égale à l'inertie totale, c'est-à-dire $\frac{K-Q}{Q}$ et la moyenne des

valeurs propres est égale à $\lambda_m = \frac{1}{Q} = \frac{1}{\text{Nb de questions}}$. On ne conserve que les valeurs propres λ

supérieures à λ_m et on calcule pour chacune d'entre elles : $\lambda' = (\lambda - \lambda_m)^2$. Le taux d'inertie modifié

est alors calculé par : $\frac{\lambda'}{\sum \lambda'}$ et on conserve les valeurs propres dont le taux modifié est supérieur à la

moyenne (des taux modifiés). Pour l'exemple traité, l'application de cette méthode donne les résultats suivants :

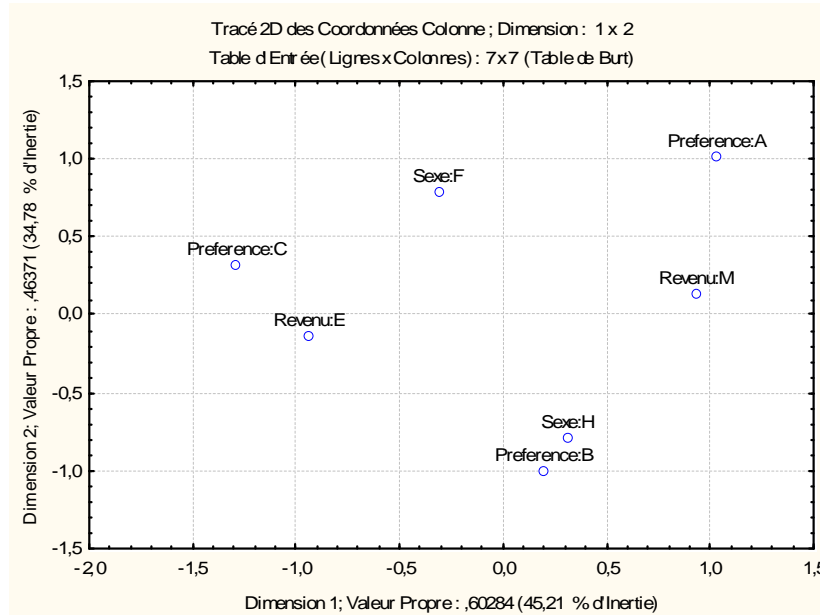
La moyenne des valeurs propres est : $\lambda_m = \frac{1}{3} = 0,33$, ce qui conduit à ne conserver que les 2 premières valeurs propres. La transformation précédente donne alors :

Nb de dim.	Val Prop.	$\lambda' = (\lambda - \lambda_m)^2$	Taux d'inertie modifié
1	0,6028	0,0726	81,04%
2	0,4637	0,0170	18,96%
3	0,2030		
4	0,0638		

Le taux d'inertie modifié moyen est de $100\%/2 = 50\%$. Seule la première valeur propre dépasse ce taux, mais une étude limitée seulement au premier axe principal présenterait peu d'intérêt. Nous étudierons donc les deux premiers.

Remarque : Selon Benzécri, les taux modifiés représentent l'écart du nuage de points par rapport au nuage parfaitement sphérique qui serait obtenu si aucun lien n'existait entre les modalités.

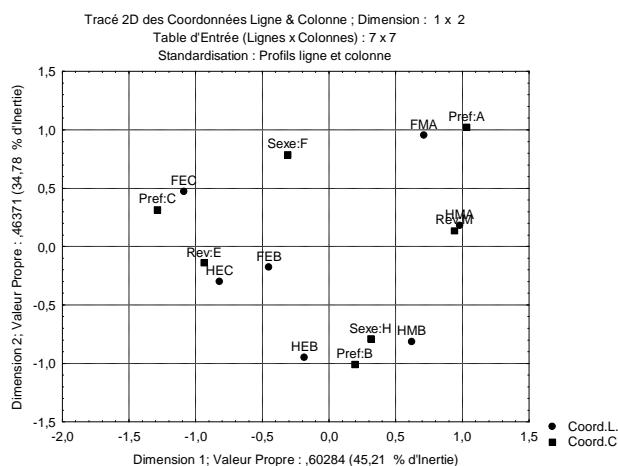
Coordonnées Colonne et Contributions à l'Inertie (Protocole dans Mini-ACM-v7.stw Table d'Entrée (Lignes x Colonnes) : 7 x 7 (Table de Burt) Inertie Totale = 1,3333											
NomLigne	Ligne Numéro	Coord. Dim.1	Coord. Dim.2	Masse	Qualité	Inertie relative	Inertie Dim.1	cosinus Dim.1	Inertie Dim.2	cosinus Dim.2	
Sexe:F	1	-0,311	0,788	0,167	0,718	0,125	0,027	0,097	0,223	0,621	
Sexe:H	2	0,311	-0,788	0,167	0,718	0,125	0,027	0,097	0,223	0,621	
Revenu:M	3	0,938	0,138	0,167	0,899	0,125	0,243	0,880	0,007	0,019	
Revenu:E	4	-0,938	-0,138	0,167	0,899	0,125	0,243	0,880	0,007	0,019	
Preference:A	5	1,032	1,024	0,100	0,906	0,175	0,177	0,456	0,226	0,450	
Preference:B	6	0,193	-1,007	0,133	0,701	0,150	0,008	0,025	0,292	0,677	
Preference:C	7	-1,288	0,319	0,100	0,755	0,175	0,275	0,711	0,022	0,044	



Bien que l'exemple ne comporte qu'un petit nombre d'observations, on remarque la proximité des modalités Préférence:B et Sexe:H, de même que l'opposition Préférence C, revenu E d'une part, Préférence A, Revenu M d'autre part selon le premier axe.

On note également que l'origine du repère est le milieu du segment joignant les deux modalités de la variable "Sexe", et aussi le milieu du segment joignant les deux modalités de la variable "Revenu". En effet, ces deux variables ont seulement deux modalités (d'où l'alignement de l'origine avec les modalités) et les deux modalités sont équiprobables (d'où la propriété du milieu).

La représentation du nuage de points représentant simultanément les modalités et les patrons de réponses est la suivante :



L'étude menée à partir du tableau de Burt mérite un commentaire particulier. En effet, dans un exposé théorique sur l'ACM, tels que ceux de [Crucianu] ou de [Rouanet, Le Roux], l'analyse du tableau de Burt est distinguée de celle du TDC ou du tableau disjonctif des patrons. Il est notamment indiqué que les valeurs propres produites par cette analyse sont les carrés des valeurs propres précédentes, et que le Phi-2 du tableau de Burt n'est pas celui du TDC. Cependant, les représentations graphiques produites (limitées aux seules modalités) peuvent être interprétées de façon analogue.

Lorsque l'on effectue une AFC en spécifiant le tableau de Burt comme tableau de contingence, on retrouve alors les résultats indiqués dans les exposés théoriques. Par exemple, le tableau des valeurs propres est alors donné par :

Valeurs Propres et Inertie de toutes les Dimensions (Tableau de Burt dans Mini-ACM-v7.stw)					
Table d'Entrée (Lignes x Colonnes) : 7 x 7					
Inertie Totale = ,62370 Chi ² = 56,133 dl = 36 p = ,01747					
Nombre de Dims.	ValSing.	ValProp.	%age Inertie	%age Cumulé	Chi ²
1	0,6028	0,3634	58,2668	58,2668	32,7071
2	0,4637	0,2150	34,4755	92,7423	19,3523
3	0,2030	0,0412	6,6045	99,3468	3,7073
4	0,0638	0,0041	0,6532	100,0000	0,3667
5	0,0000	0,0000	0,0000	100,0000	0,0000
6	0,0000	0,0000	0,0000	100,0000	0,0000

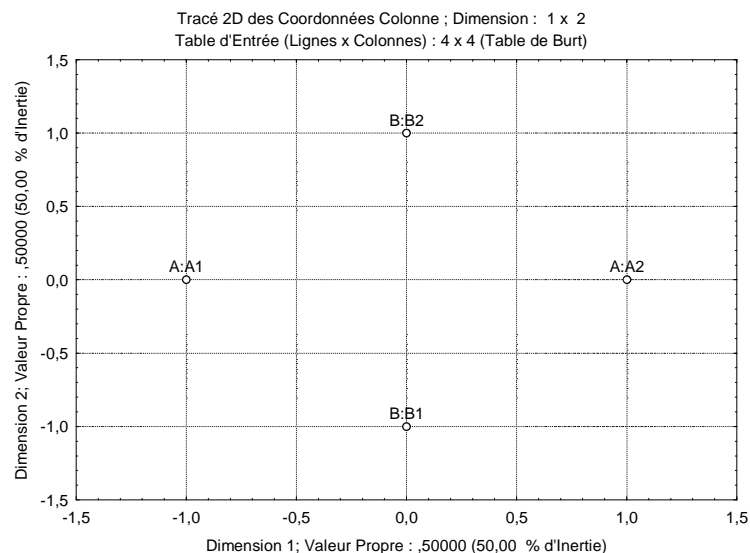
2.4.5 Exploration de l'ACM sur des mini-exemples

Etudions comment seront disposées les modalités colonnes lorsque la situation comporte 2 questions A et B à 2 modalités chacune (respectivement A1 et A2, B1 et B2). L'espace de représentation est alors de dimension 2, autrement dit, l'ACM produit une représentation non déformée dans un plan.

Cas 1 : les effectifs des modalités sont donnés par :

	A1	A2	Total
B1	50	50	100
B2	50	50	100
Total	100	100	200

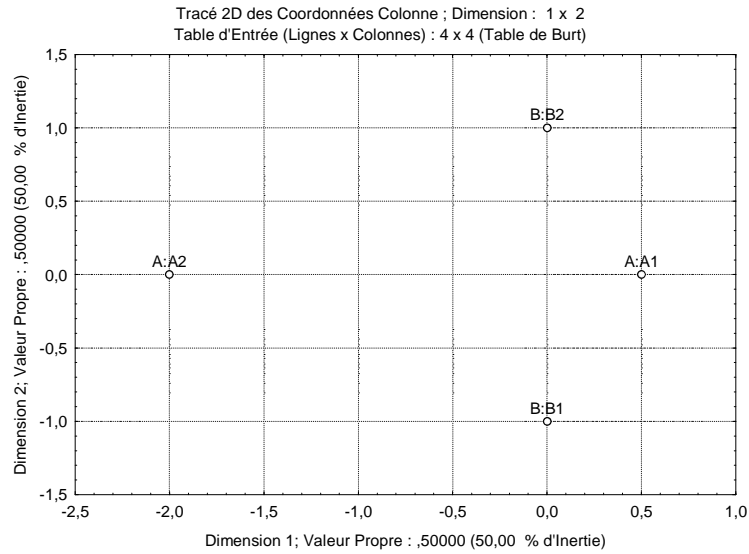
La représentation est alors :



Cas 2 : les effectifs des modalités sont donnés par :

	A1	A2	Total
B1	80	20	100
B2	80	20	100
Total	160	40	200

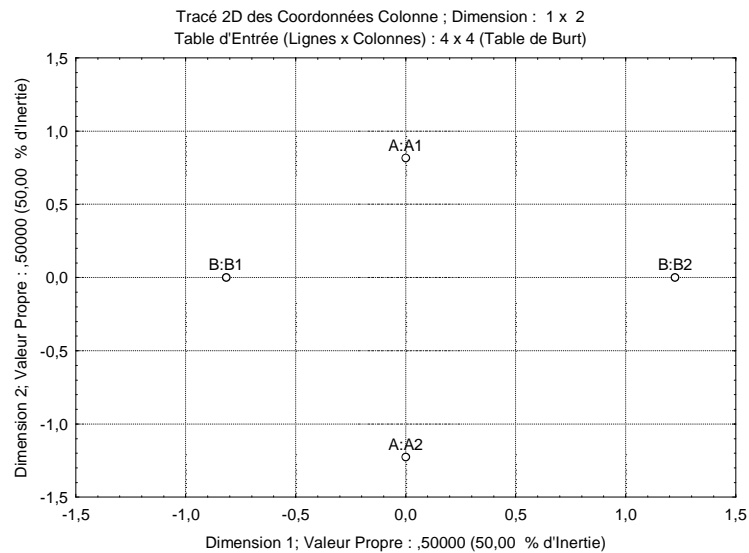
A2 est alors plus éloigné de O que A1. B1 et B2 sont à égale distance de O, et cette distance est intermédiaire entre celle de A1 et celle de A2. La représentation est alors :



Cas 3 : les effectifs des modalités sont donnés par :

	A1	A2	Total
B1	72	48	120
B2	48	32	80
Total	120	80	200

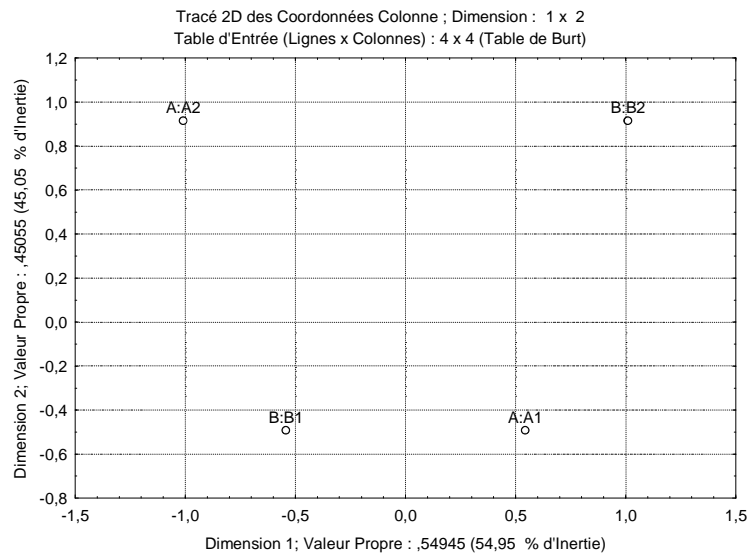
A2 est alors plus éloigné de O que A1. A1 et B1 sont à égale distance de O, et cette distance est intermédiaire entre celle de A1 et celle de A2. La représentation est alors :



Cas 4 : les effectifs des modalités sont donnés par :

	A1	A2	Total
B1	80	50	130
B2	50	20	70
Total	130	70	200

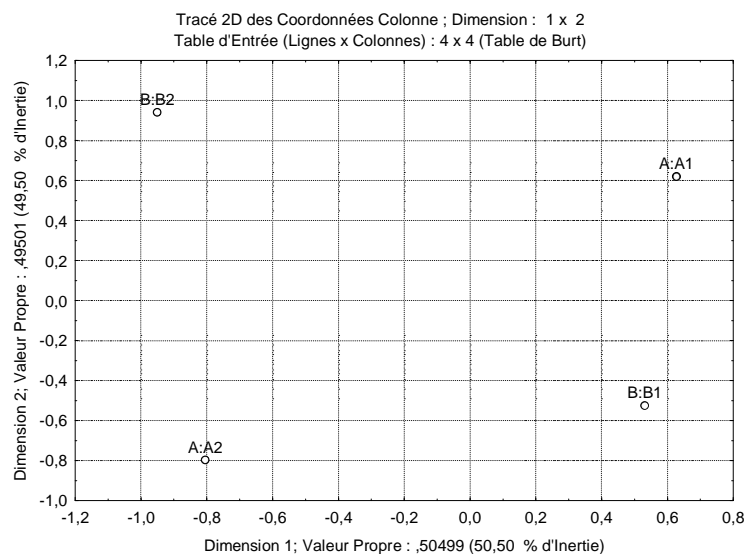
La situation est en apparence analogue à la précédente. En fait l'inertie due aux combinaisons de modalités l'emporte ici sur celle liée aux questions, et on obtient :



Cas 5 : les effectifs des modalités sont donnés par :

	A1	A2	Total
B1	73	56	129
B2	40	32	72
Total	113	88	201

C'est la situation la plus générale. On obtient :



2.4.6 ACM avec Statistica

Comme l'indiquent Rouanet et Le Roux :

Effectuer l'analyse des correspondances multiples, c'est effectuer l'analyse factorielle des correspondances du tableau disjonctif complet, muni des relations $K < Q >$ (modalités emboîtées dans les questions) et $I < K < q >$ (individus emboîtés dans les modalités de chaque question).

Quelle que soit la forme des données d'entrée, l'ACM sera réalisée à partir du menu Statistiques - Techniques exploratoires multivariées - Analyse des correspondances. Mais, selon la structure des

données, c'est l'onglet "Analyse de correspondances" ou l'onglet "Analyse des correspondances multiples (ACM)" qui sera utilisé, selon le tableau suivant :

Format des données	Onglet "Analyse des Correspondances"	Onglet "Analyse des Correspondances Multiple"	Observations
Tableau protocole	Non	Oui	AFC impossible si plus de 2 variables
Tableau d'effectifs	Non	Oui	AFC impossible si plus de 2 variables
Tableau Disjonctif Complet	Oui	Non	
Tableau Disjonctif des patrons	Oui	Non	
Tableau de Burt	Oui	Oui	Les deux analyses ne fournissent pas les mêmes résultats

Exemple.

Ref. L'exemple qui suit est accessible sur Internet à partir des adresses :

<http://www.skeptron.uu.se/broadly/sec/k-10-gda.htm>

http://www.math-info.univ-paris5.fr/~lerb/livres/MCA/MCA_en.html

Le Roux, B., Rouanet, H., Savage, M., Warde, A., Class and Cultural Division in the UK, Sociology 2008, No 42, pp.1042-1071

<http://soc.sagepub.com/content/42/6/1049>

Il s'agit vraisemblablement de données recueillies dans le cadre de l'enquête "Cultural Capital and Social Exclusion" (CCSE) administrée en 2003 et 2004 au Royaume-Uni par le National Centre for Social Research.

Le questionnaire comportait notamment les questions suivantes :

Q1 : do you prefer leisure activities you ca

Leisure:friends

Leisure:family

Leisure:alone

Leisure:partner

Q2 : Would you say that during your free time

lack time

always sth to do

Stimes nothing to do

often do nothing

Q3 : If you had more time, 1st choice would b

home DIY

artistic activities

to rest

develop knowledge

physical activities

take care of family

to take courses

Q4 : When you go out in the evening, do you u

GoingOut:friends
GoingOut:alone
GoingOut:partner
don't go out
GoingOut:family

Q5 : Time watching TV (hours by week)

TV:never
TV:<10h
TV:[10;19h[
TV:[19;30h[
TV:>=30h

Q6 : # of books or comic trips during last 12

no book
1-4 books
5-12 books
13 -39 books
40 books or more

QS1 : Gender

women
men

QS2 : Education level

no degree
CEP
CAP-BEP
BEPC
Bac
Bac+2
>Bac+2
Students

QS3 : Age

<=25
25-35
35-45
45-55
55-65
>65

QS4 : PCS

Femmes au foyer
Retraités
Etudiants, élèves
Autres inactifs
Cadres et profession
Employés
Ouvriers non qualifi
Professions interméd
Artisans, commerçant

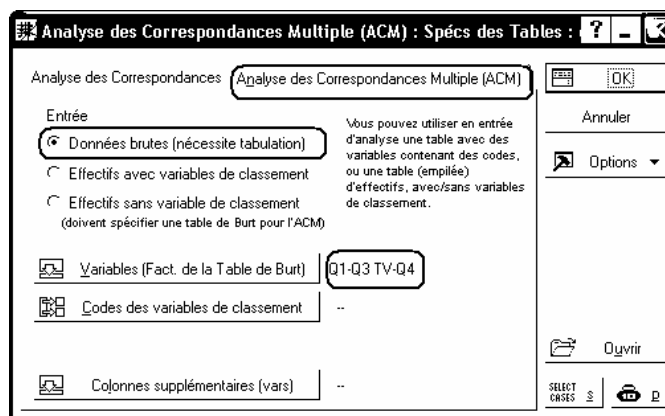
Ouvriers qualifiés
Agriculteurs

Ouvrez le fichier Culture.stw. Les données y sont saisies sous forme de tableau protocole. Deux jeux de données sont disponibles :

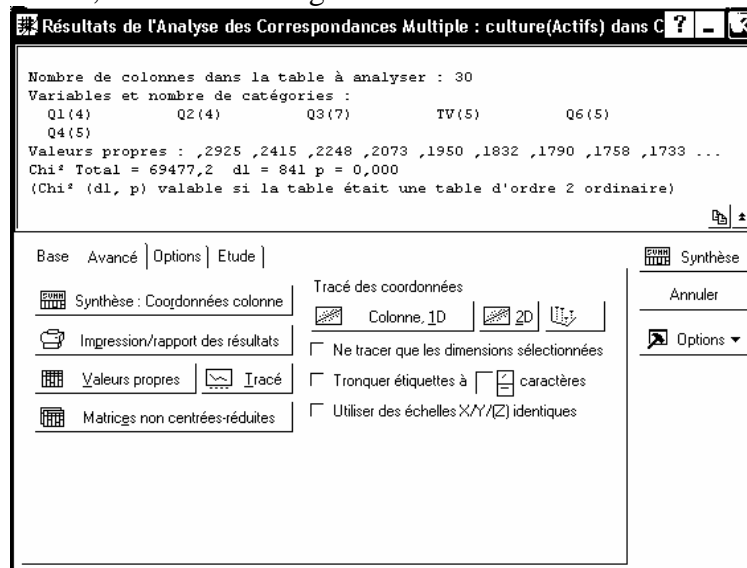
- la feuille de données Culture contient les résultats relatifs à 3002 observations, mais un certain nombre de réponses aux questions qui feront l'objet de l'ACM (Q1, Q2, Q3, Q4, TV et Q6) sont incomplètes, et la variable ISUP indique que les individus correspondants sont rendus "inactifs".

- la feuille de données Culture (Actifs) contient les réponses des 2720 individus actifs.

Réalisons, par exemple, une ACM sur les variables Q1, Q2, Q3, Q4, TV et Q6, à partir du tableau protocole. Après avoir déclaré cette feuille de données comme 'feuille active', on sélectionne l'onglet "Analyse des correspondances multiple" et on complète le premier dialogue comme suit :



Une fois ce dialogue validé, un second dialogue s'affiche :



Le bouton "Effectifs Observés" de l'onglet "Etude" permet d'obtenir un tableau similaire au tableau de Burt. Les pourcentages ligne, pourcentages colonne, khi-2, etc utilisent ce dernier tableau.

L'onglet "Avancé" permet d'obtenir les autres résultats :

Nombre de Dims.	Valeurs Propres et Inertie de toutes les Dimensions (culture(Actifs) dans Culture.stw Table d'Entrée (Lignes x Colonnes) : 30 x 30 (Table de Burt) Inertie Totale = 4,0000				
	ValSing.	ValProp.	%age Inertie	%age Cumulé	Chi²
1	0,5409	0,2925	7,3131	7,3131	5080,9201
2	0,4914	0,2415	6,0374	13,3504	4194,5915
3	0,4742	0,2248	5,6210	18,9714	3905,2809
4	0,4553	0,2073	5,1814	24,1528	3599,8892
5	0,4416	0,1950	4,8745	29,0273	3386,6677
6	0,4281	0,1832	4,5808	33,6081	3182,5987
7	0,4230	0,1790	4,4738	38,0818	3108,2367
8	0,4193	0,1758	4,3960	42,4779	3054,2426
9	0,4162	0,1733	4,3313	46,8092	3009,2644
10	0,4108	0,1688	4,2197	51,0288	2931,6975
11	0,4068	0,1655	4,1363	55,1651	2873,7570
12	0,4007	0,1605	4,0132	59,1783	2788,2769
13	0,3991	0,1593	3,9821	63,1605	2766,6609
14	0,3952	0,1562	3,9042	67,0646	2712,5159
15	0,3905	0,1525	3,8123	70,8769	2648,6693
16	0,3881	0,1506	3,7658	74,6427	2616,3419
17	0,3801	0,1445	3,6113	78,2540	2509,0210
18	0,3777	0,1427	3,5665	81,8205	2477,8935
19	0,3691	0,1362	3,4062	85,2267	2366,5577
20	0,3618	0,1309	3,2734	88,5000	2274,2318
21	0,3535	0,1249	3,1235	91,6236	2170,1442
22	0,3469	0,1203	3,0083	94,6319	2090,0819
23	0,3417	0,1168	2,9198	97,5517	2028,5967
24	0,3129	0,0979	2,4483	100,0000	1701,0171

Taux d'inertie modifiés

On ne conserve que les valeurs propres λ supérieures à λ_m et on calcule pour chacune d'entre elles :

$\lambda' = (\lambda - \lambda_m)^2$. Le taux d'inertie modifié est alors calculé par : $\frac{\lambda'}{\sum \lambda'}$ et on conserve les valeurs

propres dont le taux modifié est supérieur à la moyenne (des taux modifiés). Pour l'exemple traité, l'application de cette méthode donne les résultats suivants :

	ValSing.	ValProp.	Val. Prop. Modifiées	%age modifié	Cumuls
1	0,5409	0,2925	0,0158	56,92%	56,92%
2	0,4914	0,2415	0,0056	20,12%	77,04%
3	0,4742	0,2248	0,0034	12,16%	89,20%
4	0,4553	0,2073	0,0016	5,92%	95,12%
5	0,4416	0,1950	0,0008	2,88%	98,00%
6	0,4281	0,1832	0,0003	0,99%	98,98%
7	0,4230	0,1790	0,0002	0,54%	99,53%
8	0,4193	0,1758	0,0001	0,30%	99,83%
9	0,4162	0,1733	0,0000	0,16%	99,98%
10	0,4108	0,1688	0,0000	0,02%	100,00%
11	0,4068	0,1655			
12	0,4007	0,1605			
13	0,3991	0,1593			
14	0,3952	0,1562			

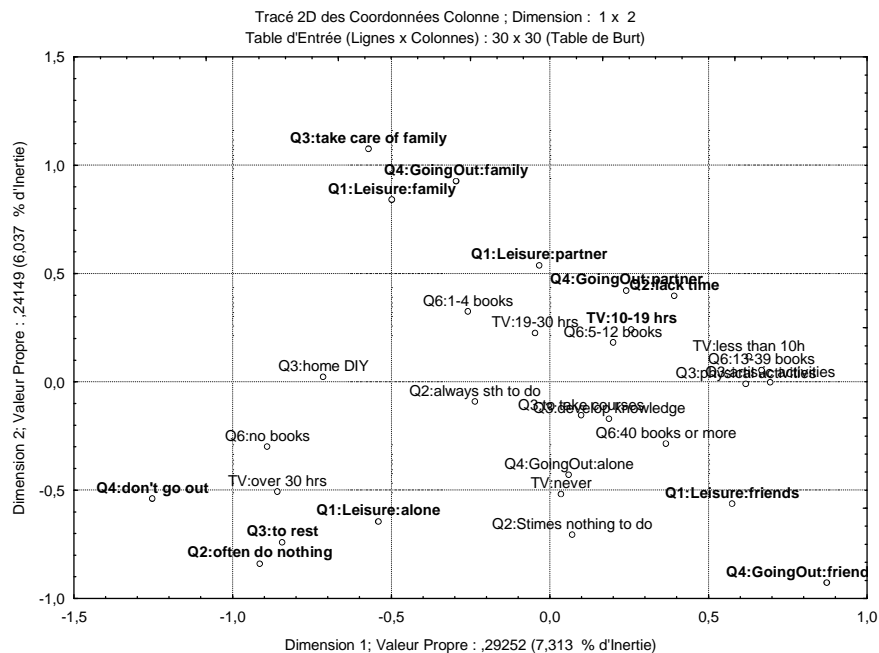
15	0,3905	0,1525			
16	0,3881	0,1506			
17	0,3801	0,1445			
18	0,3777	0,1427			
19	0,3691	0,1362			
20	0,3618	0,1309			
21	0,3535	0,1249			
22	0,3469	0,1203			
23	0,3417	0,1168			
24	0,3129	0,0979			
		0,1667	0,0278		

Le taux d'inertie modifié moyen est de $100\%/10 = 10\%$. Les trois premières valeurs propres modifiées dépassent ce taux. Nous étudierons ici les deux premiers axes.

Les coordonnées, contributions et qualités de représentation sont données dans le tableau ci-dessous.

Coordonnées Colonne et Contributions à l'Inertie (culture(Actifs) dans Culture.stw) Table d'Entrée (Lignes x Colonnes) : 30 x 30 (Table de Burt) Inertie Totale = 4,0000										
NomLigne	ligne numé	Coord. Dim. 1	Coord. Dim. 2	Masse	Qualité	Inertie Relative	Inertie Dim. 1	Cosinus ² Dim. 1	Inertie Dim. 2	Cosinus ² Dim. 2
Q1:Leisure:partner	1	-0,033	0,537	0,031	0,066	0,034	0,000	0,000	0,037	0,066
Q1:Leisure:friends	2	0,575	-0,563	0,065	0,413	0,025	0,073	0,211	0,085	0,202
Q1:Leisure:family	3	-0,498	0,841	0,044	0,343	0,031	0,037	0,089	0,129	0,254
Q1:Leisure:alone	4	-0,540	-0,645	0,027	0,134	0,035	0,027	0,055	0,046	0,079
Q2:always sth to do	5	-0,236	-0,091	0,071	0,047	0,024	0,013	0,041	0,002	0,006
Q2:lack time	6	0,392	0,396	0,069	0,218	0,024	0,036	0,108	0,045	0,110
Q2:Stimes nothing to d	7	0,071	-0,707	0,015	0,049	0,038	0,000	0,000	0,031	0,049
Q2:often do nothing	8	-0,914	-0,840	0,012	0,123	0,039	0,035	0,067	0,036	0,056
Q3:home DIY	9	-0,715	0,022	0,026	0,094	0,035	0,045	0,094	0,000	0,000
Q3:artistic activities	10	0,695	-0,003	0,024	0,082	0,036	0,040	0,082	0,000	0,000
Q3:to rest	11	-0,843	-0,741	0,019	0,159	0,037	0,045	0,089	0,042	0,069
Q3:develop knowledge	12	0,187	-0,171	0,028	0,013	0,035	0,003	0,007	0,003	0,006
Q3:physical activities	13	0,619	-0,010	0,035	0,102	0,033	0,046	0,102	0,000	0,000
Q3:take care of family	14	-0,571	1,076	0,019	0,195	0,037	0,022	0,043	0,093	0,152
Q3:to take courses	15	0,099	-0,155	0,016	0,004	0,038	0,001	0,001	0,002	0,003
TV:less than 10h	16	0,628	0,117	0,027	0,078	0,035	0,036	0,075	0,002	0,003
TV:19-30 hrs	17	-0,046	0,225	0,043	0,019	0,031	0,000	0,001	0,009	0,018
TV:10-19 hrs	18	0,258	0,241	0,049	0,051	0,030	0,011	0,027	0,012	0,024
TV:over 30 hrs	19	-0,859	-0,507	0,032	0,240	0,034	0,082	0,178	0,034	0,062
TV:never	20	0,035	-0,519	0,016	0,028	0,038	0,000	0,000	0,018	0,028
Q6:no books	21	-0,891	-0,299	0,037	0,251	0,032	0,100	0,226	0,014	0,025
Q6:5-12 books	22	0,199	0,182	0,039	0,022	0,032	0,005	0,012	0,005	0,010
Q6:1-4 books	23	-0,258	0,325	0,030	0,037	0,034	0,007	0,014	0,013	0,023
Q6:40 books or more	24	0,366	-0,286	0,026	0,041	0,035	0,012	0,025	0,009	0,015
Q6:13-39 books	25	0,667	0,053	0,034	0,117	0,033	0,053	0,116	0,000	0,001
Q4:don't go out	26	-1,253	-0,540	0,029	0,388	0,034	0,154	0,327	0,035	0,061
Q4:GoingOut:friends/o	27	0,873	-0,928	0,036	0,452	0,033	0,095	0,212	0,129	0,240
Q4:GoingOut:alone	28	0,060	-0,428	0,012	0,015	0,039	0,000	0,000	0,009	0,015
Q4:GoingOut:partner	29	0,242	0,421	0,056	0,119	0,028	0,011	0,029	0,041	0,089
Q4:GoingOut:family	30	-0,296	0,926	0,033	0,237	0,033	0,010	0,022	0,119	0,215

Dans le graphique suivant, les modalités qui ont une contribution supérieure à 3,3% à la formation du premier axe sont indiquées en caractères rouges, celles qui ont une contribution supérieure à la moyenne sur le second axe sont représentées en caractères gras.



2.4.7 Autres exemples d'ACM

Les autres exemples d'ACM que nous traiterons sont donnés à l'aide d'un tableau de Burt. En effet, c'est généralement sous cette forme que l'on trouve des données susceptibles de servir de base à un exercice.

2.4.7.1 Le cas "Aspirations des Français"

Ouvrez le classeur Aspi.stw. La présentation du cas, rappelée dans un rapport contenu dans le classeur est la suivante :

Source : Morineau A., Morin S., Pratique du traitement des enquêtes - Exemple d'utilisation du système SPAD, Cisia-Ceresta, Montreuil, 2000

On travaille sur des données extraites d'une enquête d'opinion réalisée en 1978, concernant les conditions de vie et les aspirations des Français.

Les questions prises en compte ici, et leurs modalités, sont les suivantes :

- 1- Sexe de la personne interrogée :
 - masc : masculin
 - femi : féminin
- 2- Possédez-vous des valeurs mobilières
 - vmo1 : oui
 - vmo2 : non
- 3- Taille d'agglomération
 - agg1 : moins de 2000 h
 - agg2 : de 2000 à 20000 h
 - agg3 : de 20000 à 100000 h
 - agg4 : plus de 100000h

agg5 : Paris

4- Diplôme de l'enquête :

die1 : aucun

die2 : CEP ou fin d'études

die3 : BEPC - BE - BEPS

die4 : bac - brevet sup.

die5 : université, gde école

5- Statut du logement

slo1 : en accession

slo2 : propriétaire

slo3 : locataire

slo4 : logé gratuit, autre

6- Age de l'enquête

agc1 : moins de 25 ans

agc2 : 25 à 34 ans

agc3 : 35 à 49 ans

agc4 : 50 à 64 ans

agc5 : plus de 65 ans

7- Type d'emploi

emp1 : ouvriers

emp2 : employés


emp3 : cadres

emp4 : autres

empNR : non réponse

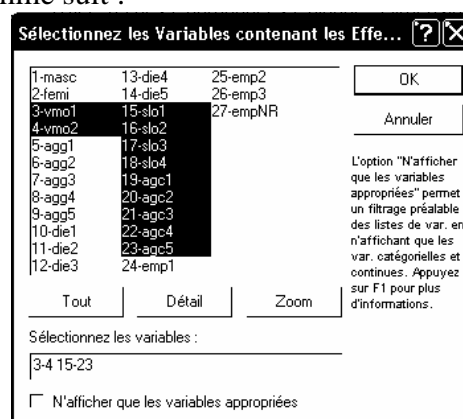
Remarque : pour une ACM sur la totalité des 27 modalités du TDB, les auteurs retiennent 5 axes principaux.

Faites tout d'abord une ACM sur la totalité du tableau de Burt (27 modalités - remarquez que seules 4 modalités de la variable "Type d'emploi" sont présentes.

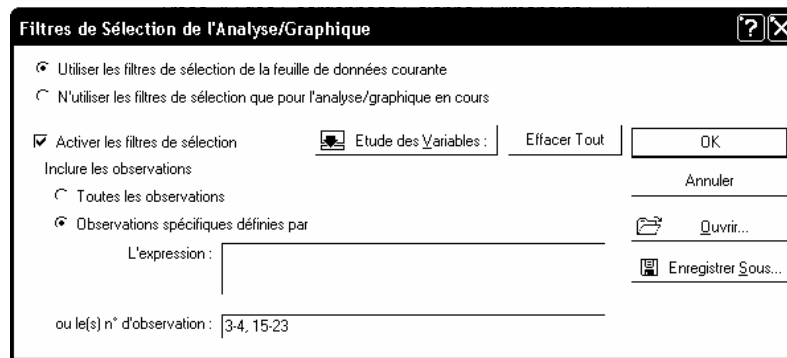
Remarque : le graphique ainsi obtenu est assez peu lisible. Il est cependant possible de l'améliorer en utilisant l'outil "Balayage/Habillage" :  . A l'aide de cet outil, il est par exemple possible de supprimer certains points qui se superposent au centre du graphique. Attention cependant à ce que le graphique conserve une certaine honnêteté intellectuelle !

Réalisez ensuite une ACM en ne prenant en compte que certaines variables, par exemple, la variable 2 (valeurs mobilières), la variable 5 (statut du logement) et la variable 6 (âge de l'enquête). Pour cela :

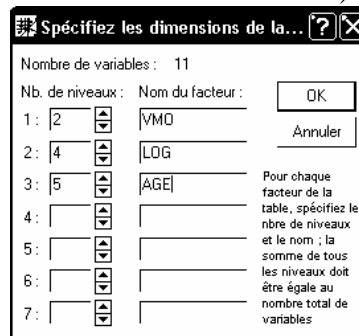
- Sélectionnez les variables comme suit :



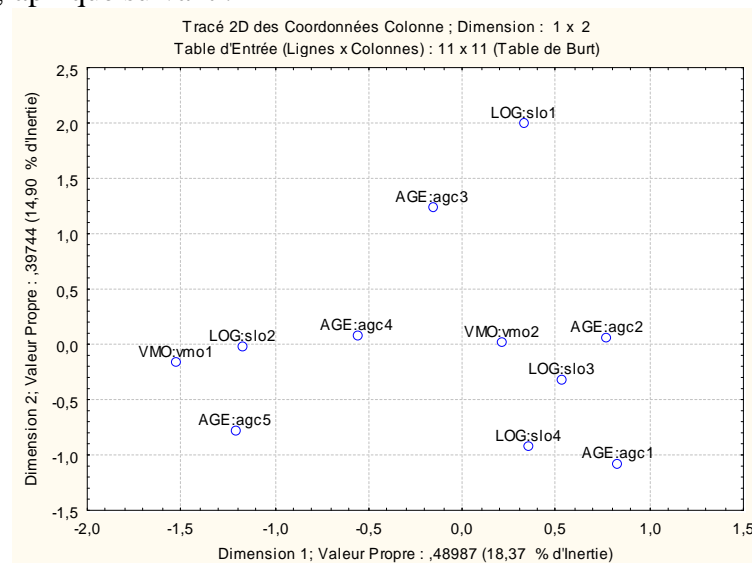
- Sélectionnez ensuite les observations correspondantes, par exemple en les désignant par leurs numéros. Pour cela, cliquez sur le bouton "Select Cases" et complétez le dialogue comme suit :



- Structurez enfin les variables (bouton "Structure de la table") de la façon suivante :



On obtient ainsi le graphique suivant :



La possession de valeurs mobilières est ainsi plutôt associée à l'occupation d'un logement en propriété, et à une personne relativement âgée (agc4, agc5), alors que la non-possession est plutôt le fait de personnes jeunes, locataires. L'âge agc3 est dans une certaine mesure associé à l'accession à la propriété alors que le dernier statut du logement est plutôt le fait des moins de 25 ans (qui, par ailleurs, ne possèdent généralement pas de valeurs mobilières).

2.4.7.2 Le cas "Avignon"

Source : Croutsche, J.-J., Pratiques statistiques en gestion et études de marchés, Editions ESKA, Paris, 1997

Une enquête sur la fréquentation du centre ville d'Avignon. On trouvera ci-dessous le texte d'une partie des questions posées, ainsi que le codage des modalités de réponse.

- 1- Combien de fois par mois allez-vous dans le centre ville pour faire des achats ?
 - a1 : Plus de 3 fois par mois
 - a2 : de 2 à 3 fois
 - a3 : de 1 à 2 fois
 - a4 : Autre
- 2- Votre fréquentation du centre ville est-elle plus ou moins importante qu'il y a 5 ans ?
 - f1 : Beaucoup moins importante
 - f2 : Un peu moins importante
 - f3 : Identique
 - f4 : Un peu plus importante
 - f5 : Beaucoup plus importante
- 3-
- 4-
- 5- Etes-vous satisfait de la propreté du centre ville ?
 - p1 : très satisfait
 - p2 : satisfait
 - p3 : moyennement satisfait
 - p4 : peu satisfait
 - p5 : très peu satisfait
- 6- Que pensez-vous de la sécurité dans le centre ville ?
 - s1 : Très faible
 - s2 : Faible
 - s3 : Normale
 - s4 : Importante
 - s5 : Très importante
- 7- Si vous observez des problèmes de sécurité : vous arrive-t-il de ne pas vous rendre dans le centre ville à cause de ce problème ?
 - r1 : oui
 - r2 : non
- 8-
- 9-
- 10-
- 11- Où habitez-vous ?
 - h1 : Avignon intra-muros
 - h2 : Avignon extra-muros
 - h3 : autre
- 12-
- 13- Dans quelle tranche d'âge vous situez-vous ?
 - â1 : 15-19 ans
 - â2 : 20-30 ans
 - â3 : 31-40 ans
 - â4 : 41-50 ans
 - â5 : 51-60 ans
 - â6 : Plus de 60 ans
- 14-

Dans le classeur Avignon.stw se trouvent diverses feuilles de données contenant les tableaux de Burt obtenus en sélectionnant 3 ou 4 des items du questionnaire. Analysez chacun des aspects ainsi définis à l'aide d'une ACM.