

4.2 L'ACM avec Statistica

4.2.1 Présentation des données étudiées

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

<http://www.info.univ-angers.fr/pub/gh/Datasets/pbio.htm>.

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. Les données figurent (sous plusieurs formes) dans le classeur PBIO-2007.stw du serveur de TD. En particulier, le protocole observé se trouve dans la feuille PBIO-Protocole-Num (les modalités y sont représentées par leur code numérique) et dans la feuille PBIO-Protocole-Alpha (dans cette feuille, des étiquettes de texte, représentant des abréviations des libellés de réponses, ont été introduites).

4.2.2 Exploration préalable des données

Avant de réaliser des analyses multivariées sur ces données, explorez-les en utilisant le menu Statistiques - Statistiques Elémentaires - Tables de fréquences pour obtenir les tris à plat des différentes variables.

	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

On peut également utiliser le menu Statistiques - Statistiques Elémentaires - Statistiques Descriptives - Tableaux et tris croisés pour obtenir des tableaux de contingence tels que :

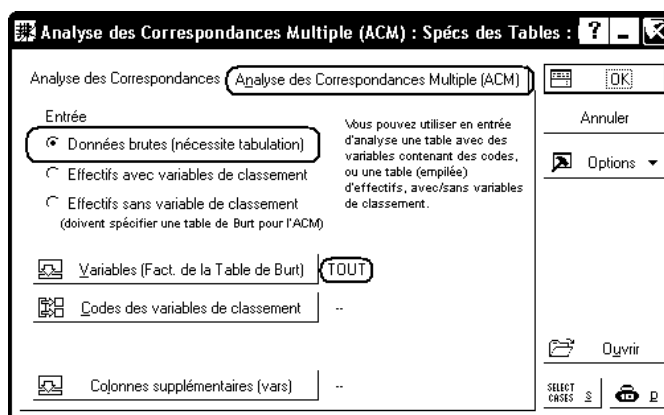
Table de Fréquences - Synthèse (PBIO-2.sta)								
Table :CONSOM(7) x AGE(6)								
	CONSOM	AGE	AGE	AGE	AGE	AGE	AGE	Totaux
		- de 25 ans	entre 25 et 35 ans	entre 35 et 45 ans	entre 45 et 55 ans	entre 55 et 65 ans	+ de 65 ans	Ligne
	jamais	15	39	19	21	9	8	111
	une fois	3	5	0	3	3	1	15
	rarement	16	25	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

4.2.3 ACM menée à partir d'un tableau protocole

4.2.3.1 ACM à partir d'un tableau protocole : étude sur l'ensemble des variables

On peut essayer de réaliser une ACM avec, comme données actives, l'ensemble des données fournies. Gardez ici PBIO-Protocole-Alpha comme feuille active. 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.



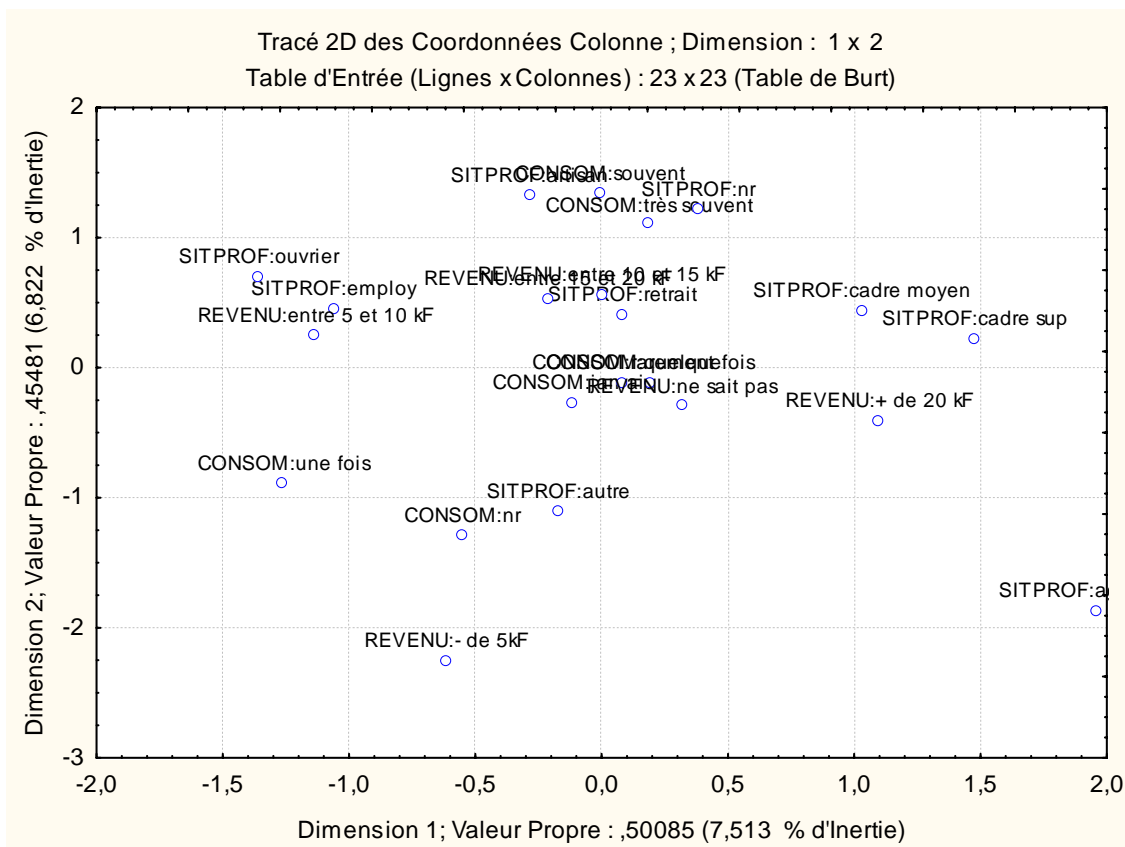
Cependant, les modalités sont alors trop nombreuses et les résultats sont difficilement interprétables. Par ailleurs, certains problèmes cités plus haut se rencontrent sur ces données :

- 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.

4.2.3.2 ACM menée sur les variables CONSOM, SITPROF et REVENU

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 :



Calcul des taux d'inertie modifiés

On sait que Statistica ne fournit pas directement ce résultat. Nous allons donc devoir recourir à Excel pour réaliser le calcul.

Dans Statistica, affichez le tableau des 20 valeurs propres correspondant à cette analyse. Sélectionnez l'ensemble du tableau et utilisez le menu Edition - Copier avec noms.

Sans quitter Statistica, ouvrez un classeur Excel et collez le contenu du tableau. Calculez ensuite successivement :

- la moyenne des valeurs propres
- pour les valeurs propres supérieures à cette moyenne, les valeurs propres modifiées (écart à la moyenne précédente élevé au carré)
- la somme de ces valeurs propres
- le taux d'inertie modifié correspondant.

Vous devriez aboutir à un tableau du type suivant :

	ValSing.	ValProp.	%age	%age	Chi ²	VP modif.	Tx modifiés
1	0,7077	0,5008	7,5127	7,51	665,74	0,02806	42,04%
2	0,6744	0,4548	6,8222	14,33	604,55	0,01476	22,11%
3	0,6571	0,4317	6,4760	20,81	573,87	0,00968	14,51%
4	0,6435	0,4141	6,2119	27,02	550,47	0,00653	9,78%
5	0,6318	0,3992	5,9884	33,01	530,66	0,00434	6,50%
6	0,6171	0,3808	5,7116	38,72	506,13	0,00225	3,37%
7	0,6004	0,3604	5,4065	44,13	479,10	0,00073	1,10%
8	0,5938	0,3526	5,2896	49,42	468,74	0,00037	0,56%
9	0,5816	0,3382	5,0737	54,49	449,61	0,00002	0,04%
10	0,5779	0,3340	5,0103	59,50	443,99	0,00000	0,00%
11	0,5662	0,3206	4,8088	64,31	426,13		
12	0,5583	0,3117	4,6755	68,99	414,32		

13	0,5506	0,3031	4,5470	73,53	402,93		
14	0,5481	0,3004	4,5055	78,04	399,26		
15	0,5384	0,2899	4,3484	82,39	385,34		
16	0,5242	0,2747	4,1212	86,51	365,20		
17	0,5114	0,2616	3,9235	90,43	347,68		
18	0,4919	0,2419	3,6288	94,06	321,56		
19	0,4853	0,2355	3,5321	97,59	313,00		
20	0,4005	0,1604	2,4062	100,00	213,22		
	Moyenne	0,3333			Total	0,06675	

Exploitation de ce résultat : après transformation, il reste 10 valeurs propres. Parmi celles-ci, les 3 premières (et éventuellement la 4ème) ont des taux d'inertie modifiés supérieurs ou voisins de $100\%/10 = 10\%$. L'étude devrait donc se concentrer sur les 3, voire 4 premiers axes.

Tableau des résultats relatifs aux lignes

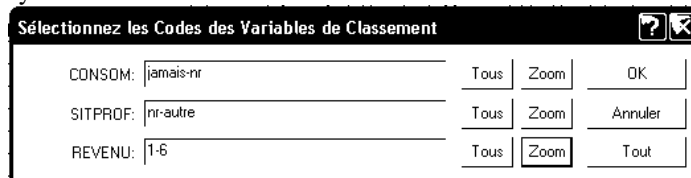
L'interprétation utilisera également les tableaux des coordonnées, qualités de représentation et contributions :

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
CONSUM:jamais	1	-0,1147	-0,2745	0,0883	0,0319	0,0368	0,0023	0,0047	0,0146	0,0271
CONSUM:une fois	2	-1,2597	-0,8938	0,0119	0,0886	0,0482	0,0378	0,0589	0,0210	0,0297
CONSUM:rarement	3	0,0862	-0,1216	0,0708	0,0060	0,0394	0,0011	0,0020	0,0023	0,0040
CONSUM:quelquefois	4	0,1921	-0,1167	0,0931	0,0196	0,0360	0,0069	0,0143	0,0028	0,0053
CONSUM:souvent	5	-0,0048	1,3379	0,0191	0,1088	0,0471	0,0000	0,0000	0,0751	0,1088
CONSUM:très souvent	6	0,1899	1,1087	0,0390	0,1676	0,0442	0,0028	0,0048	0,1054	0,1628
CONSUM: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

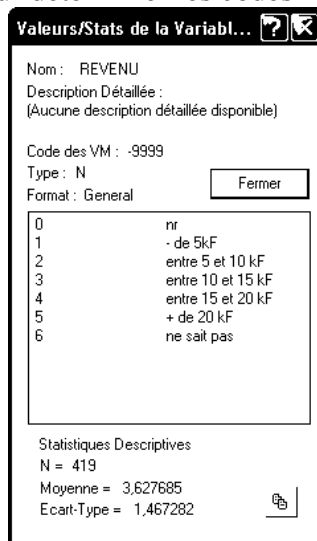
4.2.3.3 Elimination des modalités de faible effectif

L'étude précédente montre que les modalités "non réponse", notamment la modalité "nr" de la variable "REVENU", occupent des positions excentrées sur le graphique. Lorsqu'on travaille à partir d'un tableau protocole, on peut facilement éliminer ces modalités de la façon suivante :

- On sélectionne les variables comme précédemment
- On utilise ensuite le bouton "Codes des variables de classement" et l'on complète le dialogue, par exemple, comme suit :



On pourra utiliser le bouton "Zoom" pour déterminer les codes numériques des modalités à sélectionner :



On pourra observer que le fait d'éliminer cette modalité ne modifie pas vraiment les résultats produits, mais améliore la lisibilité des graphiques.

4.2.4 ACM menée à partir d'un tableau de Burt

Statistica permet également de générer un tableau de Burt à partir du protocole et/ou de réaliser une ACM à partir d'un tableau de Burt.

4.2.4.1 Générer le tableau de Burt correspondant aux variables CONSOM, AGE et ETAT CIVIL

Faites une nouvelle Analyse des correspondances multiples en sélectionnant les variables CONSOM, AGE et ETAT CIVIL comme variables de l'étude. Sélectionnez ensuite l'onglet "Etude" puis le bouton "Effectifs observés".

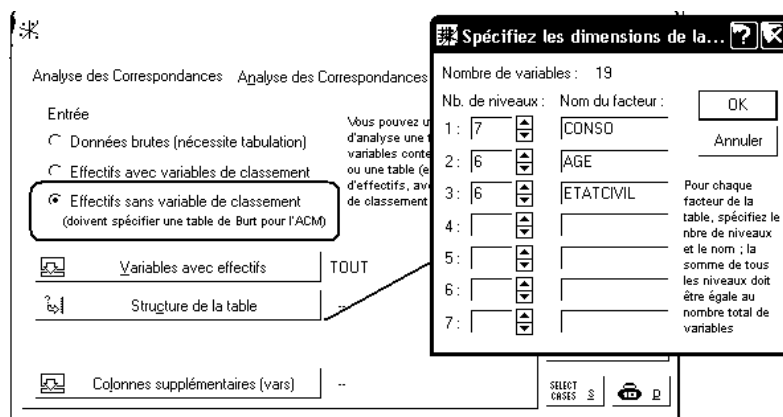
La feuille de résultats obtenue est constituée du tableau de Burt relatif à ces trois variables, et des marges ligne et colonne correspondantes. On peut ensuite la copier (menu local "Copier le document du classeur") et la coller dans le classeur, au même niveau hiérarchique que les autres feuilles de données. On peut également supprimer la dernière ligne, la dernière colonne et renommer les variables, faute de quoi, toutes les modalités d'une même question seront étiquetées de façon identique dans le graphique produit. On obtient ainsi un résultat tel que :

	jamais	une fois	rarement	quelquefois	souvent	très souvent	NR	- de 25 ans	de 25 à 35 ans	de 35 à 45 ans	de 45 à 55 ans	de 55 à 65 ans	+ de 65 ans	autre	marié	célibataire	divorcé	concubinage	veuf
CONSOM:jamais	111	0	0	0	0	0	0	15	39	19	21	9	8	3	58	30	4	12	4
CONSOM:une fois	0	15	0	0	0	0	0	3	5	0	3	3	1	0	8	5	0	1	1
CONSOM:rarement	0	0	89	0	0	0	0	16	25	17	19	4	8	0	45	32	3	7	2
CONSOM:quelquefois	0	0	0	117	0	0	0	13	28	16	38	12	10	1	72	25	9	7	3
CONSOM:souvent	0	0	0	0	24	0	0	2	4	3	10	2	3	1	9	7	2	4	1
CONSOM:très souvent	0	0	0	0	0	49	0	0	11	13	8	9	8	0	30	9	2	6	2

s souvent																			
CONSOM:nr	0	0	0	0	0	0	14	3	4	2	2	2	1	1	6	6	1	0	0
AGE:- de 25 ans	15	3	16	13	2	0	3	52	0	0	0	0	0	1	4	41	0	6	0
AGE:entre 25 et 35 ans	39	5	25	28	4	11	4	0	116	0	0	0	0	3	45	46	2	19	1
AGE:entre 35 et 45 ans	19	0	17	16	3	13	2	0	0	70	0	0	0	41	15	4	8	2	
AGE:entre 45 et 55 ans	21	3	19	38	10	8	2	0	0	0	101	0	0	2	81	5	6	4	3
AGE:entre 55 et 65 ans	9	3	4	12	2	9	2	0	0	0	0	41	0	0	31	4	5	0	1
AGE:+ de 65 ans	8	1	8	10	3	8	1	0	0	0	0	0	39	0	26	3	4	0	6
ETATCIVIL: autre	3	0	0	1	1	0	1	1	3	0	2	0	0	6	0	0	0	0	0
ETATCIVIL: marie	58	8	45	72	9	30	6	4	45	41	81	31	26	0	228	0	0	0	0
ETATCIVIL: celibataire	30	5	32	25	7	9	6	41	46	15	5	4	3	0	0	114	0	0	0
ETATCIVIL: divorce	4	0	3	9	2	2	1	0	2	4	6	5	4	0	0	0	21	0	0
ETATCIVIL: concubinage	12	1	7	7	4	6	0	6	19	8	4	0	0	0	0	0	0	37	0
ETATCIVIL: veuf	4	1	2	3	1	2	0	0	1	2	3	1	6	0	0	0	0	0	13

4.2.4.2 ACM menée variables CONSOM, AGE et ETAT CIVIL

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.

4.2.5 Améliorer les graphiques fournis par Statistica

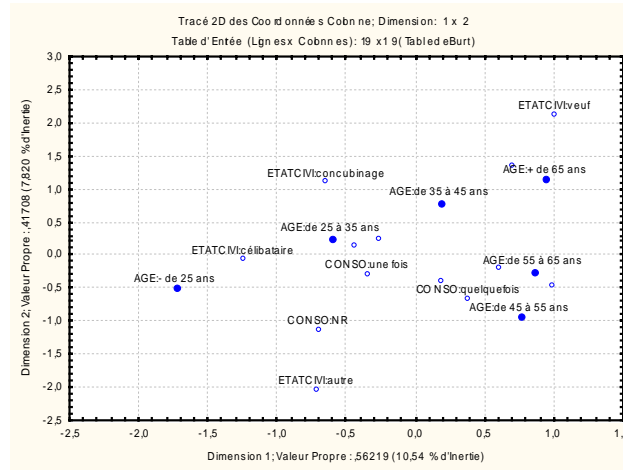
4.2.5.1 L'édition de graphiques dans Statistica

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.
- 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...").

- On peut utiliser l'outil "habillage / balayage" : . Par exemple, pour éviter d'afficher toutes les étiquettes de points, on sélectionne les points pour lesquels on souhaite garder les étiquettes, puis on clique sur "Etiqueter les points sélectionnés".
- L'éditeur de données graphiques est activé à l'aide d'un clic droit sur le graphique, puis en sélectionnant l'item "Editeur de données graphiques" dans le menu local. Il permet notamment de modifier certaines étiquettes, et même de déplacer ou de supprimer certains points.

Exemple : Utilisez les outils décrits ci-dessus pour obtenir, pour l'ACM faite au paragraphe précédent, une représentation graphique analogue à celle ci-dessous (les points représentant les modalités de la variable AGE sont marqués, et les étiquettes des modalités des autres variables ont été supprimées lorsqu'elles risquaient de nuire à la lisibilité).



4.2.5.2 Coller ou exporter le graphique dans d'autres logiciels

Coller un graphique dans Word

A l'aide du bouton droit de la souris, on peut copier le graphique dans Statistica.

Si on utilise le menu "coller" de Word :

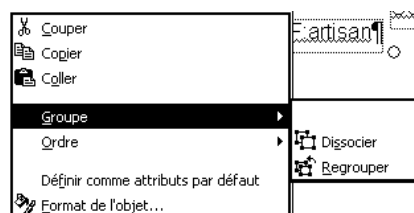
- L'objet inséré est du type "graphique Statistica"
- C'est un bloc lié au paragraphe actif au moment de l'insertion, et sa position peut varier en fonction des modifications ultérieures du texte.

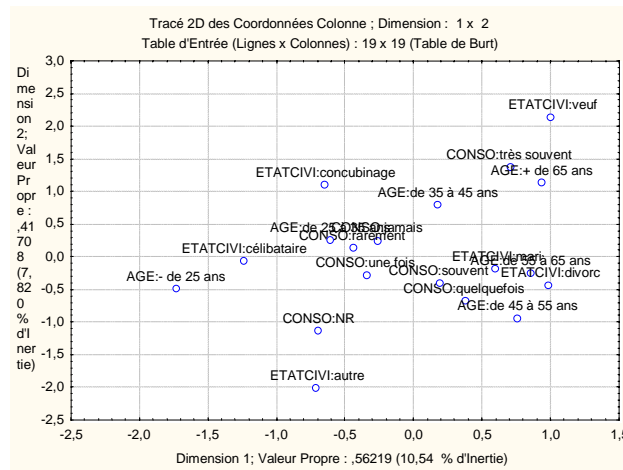
Dans Word, on peut aussi utiliser le menu Edition - Collage spécial, et le coller comme objet "Graphique Statistica" ou comme objet "image", en désactivant la boîte à cocher "dissocier du texte". Dans les deux cas, un double clic sur le graphique permet de l'éditer. Mais, s'il s'agit d'un objet de type "graphique Statistica", ce sont les outils d'édition de Statistica qui sont disponibles, alors que s'il s'agit d'une image, ce sont les outils d'édition de dessins de Word qui sont disponibles.

Par exemple, reprenez le graphique de l'ACM précédente, collez-le comme image dans Word, et déplacez les étiquettes de points de façon à améliorer sa lisibilité.

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 :





4.2.5.3 Exporter un graphique

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)".

On peut alors ouvrir 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.

4.2.6 ACM menée à partir d'un tableau disjonctif

4.2.6.1 L'exemple choisi

L'exemple qui suit est tiré de [Rouanet - Le Roux] qui fait lui-même référence à une célèbre enquête britannique (D.V. Glass, 1954, Social Mobility in Britain, London, Routledge & Kegan Paul).

Une enquête a été menée auprès de 3450 individus. Les variables qui ont été observées sont les suivantes :

- A : statut social du père du répondant (deux modalités : a1 élevé, a2 faible)
- B : niveau d'études du répondant (deux modalités : b1 élevé, b2 faible)
- C : statut du répondant (trois modalités : c1, c2, c3 de niveaux décroissants).

Le tableau des effectifs est donné par :

A	B	C	n _k
a1	b1	c1	106
a1	b1	c2	88
a1	b1	c3	8
a1	b2	c1	18
a1	b2	c2	45
a1	b2	c3	11
a2	b1	c1	106
a2	b1	c2	776
a2	b1	c3	133
a2	b2	c1	27
a2	b2	c2	1274
a2	b2	c3	858
TOTAL			3450

Nous avons en tout 7 modalités, et $2 \times 2 \times 3 = 12$ patrons de réponses possibles. Tous sont représentés parmi les réponses observées. Le début du tableau disjonctif complet est le suivant :

	a1	a2	b1	b2	c1	c2	c3
sujet 1	1	0	1	0	1	0	0
sujet 2	1	0	1	0	0	1	0
...	...	0

Les données observées peuvent également être décrites à l'aide du *tableau disjonctif des patrons* :

	a1	a2	b1	b2	c1	c2	c3	TOTAL
a1b1c1	106	0	106	0	106	0	0	318
a1b1c2	88	0	88	0	0	88	0	264
a1b1c3	8	0	8	0	0	0	8	24
a1b2c1	18	0	0	18	18	0	0	54
a1b2c2	45	0	0	45	0	45	0	135
a1b2c3	11	0	0	11	0	0	11	33
a2b1c1	0	106	106	0	106	0	0	318
a2b1c2	0	776	776	0	0	776	0	2328
a2b1c3	0	133	133	0	0	0	133	399
a2b2c1	0	27	0	27	27	0	0	81
a2b2c2	0	1274	0	1274	0	1274	0	3822
a2b2c3	0	858	0	858	0	0	858	2574
TOTAL	276	3174	1217	2233	257	2183	1010	10350

Enfin, le tableau de Burt est ici :

	A:a1	A:a2	B:b1	B:b2	C:c1	C:c2	C:c3	Total
A:a1	276	0	202	74	124	133	19	828
A:a2	0	3174	1015	2159	133	2050	991	9522
B:b1	202	1015	1217	0	212	864	141	3651
B:b2	74	2159	0	2233	45	1319	869	6699
C:c1	124	133	212	45	257	0	0	771
C:c2	133	2050	864	1319	0	2183	0	6549
C:c3	19	991	141	869	0	0	1010	3030
Total	828	9522	3651	6699	771	6549	3030	31050

4.2.7 Lien entre l'ACM et l'AFC

Rappel: 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).

Menée à partir d'un tableau disjonctif, l'ACM est donc une AFC. Et, c'est bien l'onglet "Analyse des correspondances" qui nous servira ici. Grâce à la propriété d'équivalence distributionnelle, il revient au même d'utiliser le tableau disjonctif complet ou le tableau disjonctif des patrons, mais le premier comporte 3450 lignes alors que le second n'en comporte que 12.

Nous pouvons donc, comme en AFC, nous intéresser aux profils ligne et colonne, aux taux de liaison et au Φ^2 d'un des tableaux disjonctifs vu comme un tableau de contingence. On retrouverait les mêmes résultats en effectuant une AFC sur le tableau disjonctif complet.

Rappel : le coefficient Phi-2 vaut :

$$\Phi^2 = \frac{K}{Q} - 1$$

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

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

4.2.8 Valeurs propres

Les fichiers contenus dans le répertoire Statut du serveur de TD permettent de retrouver les résultats qui suivent.

Chargez Statistica et ouvrez le classeur Statut-2007.stw. Utilisez la feuille de données statut-disjonctif-patrons comme feuille active.

Exécutez ensuite une AFC en indiquant que la feuille de données est un tableau de contingence.

Vous devriez obtenir le tableau des valeurs propres (non nulles) suivant :

Valeurs Propres et Inertie de toutes les Dimensions (statut-disjonctif-patrons.sta)						
Table d'Entrée (Lignes x Colonnes) : 12 x 7						
Inertie Totale = 1,3333 Chi2 = 13800, dl = 66 p = 0,0000						
	ValSing.	ValProp.	%age Inertie	%age Cumulé	Chi_	
1	0,7466	0,5574	41,81	41,81	5769,17	
2	0,5983	0,3579	26,84	68,65	3704,29	
3	0,4824	0,2328	17,46	86,11	2409,25	
4	0,4304	0,1852	13,89	100,00	1917,27	

4.2.8.1 Calcul des taux d'inertie modifiés

Dans notre exemple, deux valeurs propres sont supérieures à la moyenne (1/3). Les valeurs modifiées conduisent aux résultats suivants :

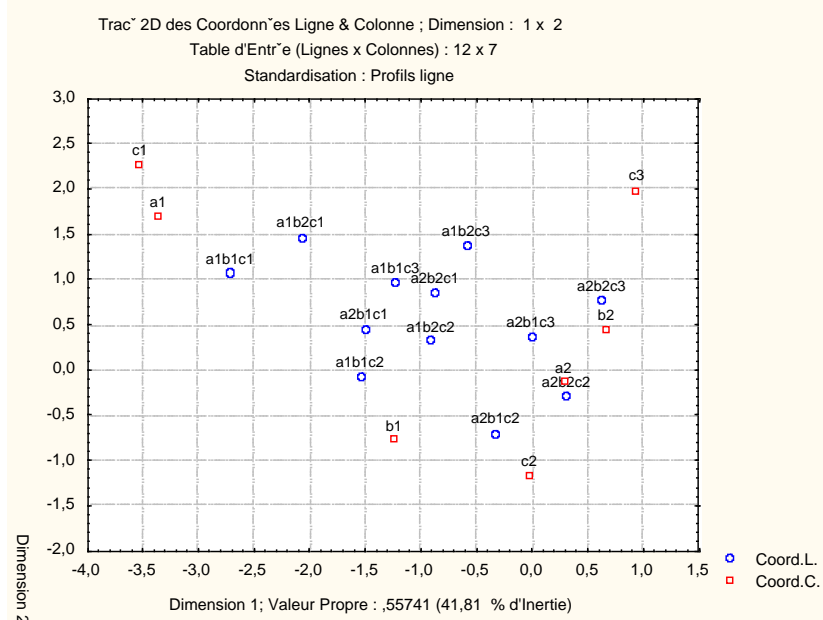
Valeur propre	λ'_l	Taux modifié
0,5574	0,1130	98,81%
0,3579	0,0014	1,19%

On pourrait donc se borner à étudier l'axe 1.

4.2.9 Résultats graphiques

Les graphiques produits possèdent des propriétés géométriques intéressantes. Cependant, nous avons jusqu'à présent utilisé l'option "Centrer-réduire les données - Profils ligne et colonne" (sous l'onglet "Options" de la fenêtre de dialogue). Or, la mise en évidence de ces propriétés nécessite d'utiliser, selon le cas, l'option "Profils ligne (interpréter dist. lignes)" ou l'option "Profils colonne (interpréter dist. colonnes)".

Par exemple, en utilisant les profils lignes, on obtient :

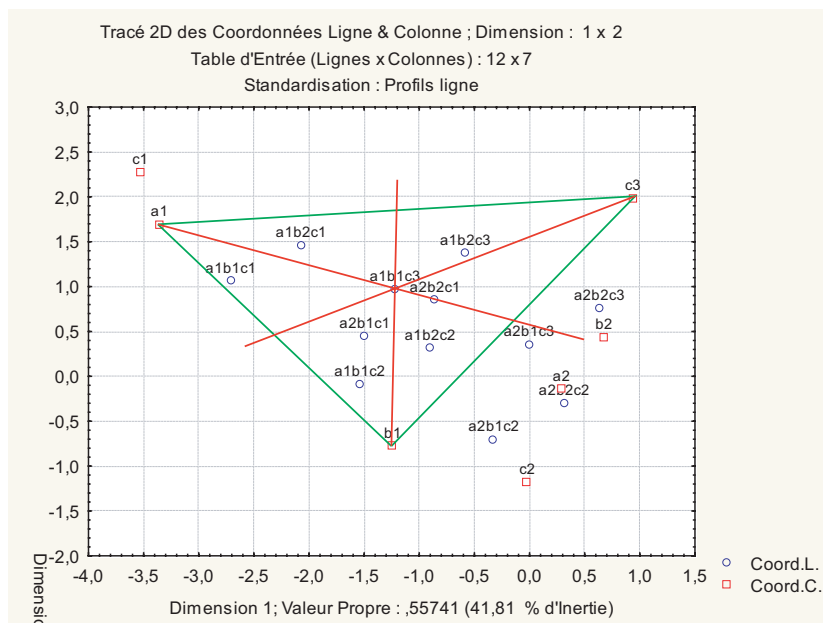


Avec ce choix d'échelles, le nuage des patrons est entièrement contenu à l'intérieur de celui des modalités. Ce graphique met particulièrement bien en évidence les propriétés suivantes :

Le point représentant chaque patron est l'équibarycentre des modalités correspondant à ce patron.

Cette propriété est vraie aussi bien pour les individus que pour les patrons. Elle est vraie dans l'espace multidimensionnel, et elle est conservée par les projections sur les plans factoriels.

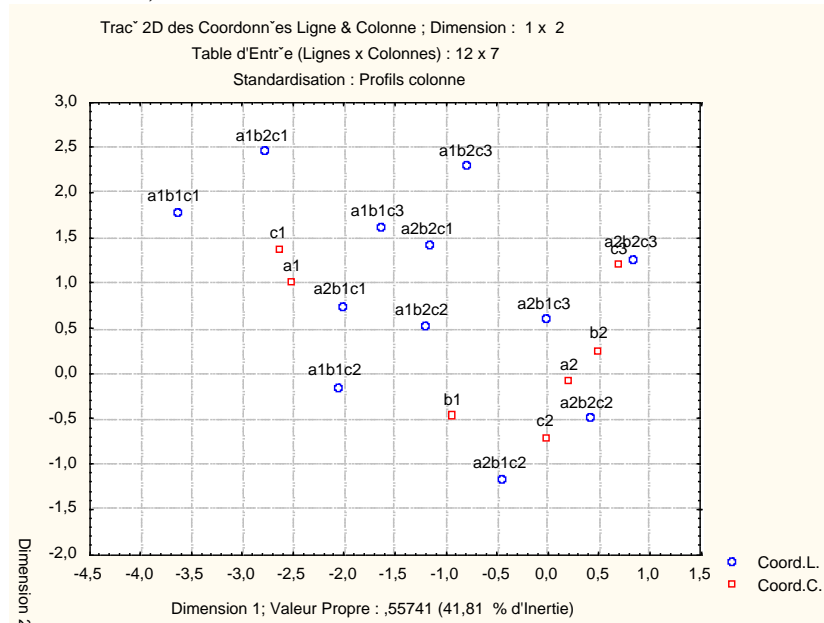
De manière plus claire, chaque patron est le centre de gravité du triangle formé par ses trois modalités. Ou encore, par exemple, considérons les modalités a1, b1, c3 et le patron a1b1c3. Chaque droite joignant l'une de ces modalités au point représentant le patron passe par le milieu du segment défini par les deux autres modalités :



On constate également sur le graphique que les droites (a1 a2) et (b1 b2) passent par l'origine du repère, qui est également à l'intérieur du triangle formé par les trois points c1, c2 et c3.

Le sous-nuage des modalités d'une question a pour point moyen le point moyen du nuage (moyenne pondérée par les fréquences des modalités).

En utilisant les profils colonnes, on obtient :



Chaque modalité est obtenue comme barycentre des patrons qui l'ont choisie, mais chaque patron doit être pondéré par sa fréquence. La constatation la plus immédiate que l'on peut faire sur le graphique ci-dessus est la suivante : chaque modalité se trouve à l'intérieur du polygone (convexe) défini à partir des 4 ou 6 patrons qui l'ont choisie.

On voit également apparaître sur cette représentation la propriété *d'équipollence*, vérifiée par les patrons de réponses. Par exemple les 4 segments suivants :

- le segment qui joint a1b1c1 à a1b1c2
- le segment qui joint a1b2c1 à a1b2c2
- le segment qui joint a2b1c1 à a2b1c2
- le segment qui joint a2b2c1 à a2b2c2

sont parallèles, de même longueur et de même sens

4.2.10 Comparaison entre analyse d'un tableau de Burt et celle d'un TDC

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.

Qu'en est-il avec Statistica ?

Rendez active la feuille de données Statut-Burt.

Effectuez l'analyse en choisissant l'onglet "Analyse des Correspondances Multiple" et l'item "Tableau de Burt".

On constate que l'on obtient, pour les modalités, des résultats identiques aux précédents. En particulier, les valeurs propres sont celles qui ont indiquées plus haut. Ce sont également celles que l'on obtiendrait

en effectuant l'analyse à partir de l'onglet "Analyse des Composantes Multiples" et du tableau protocole ou du tableau des effectifs.

En revanche, nous pouvons effectuer une AFC à l'aide de l'onglet "Analyse des correspondances", 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 :

Nombre de Dims.	Valeurs Propres et Inertie de toutes les Dimensions (statut-Burt.sta)				
	Inertie Totale = ,52730 Chi2 = 16373, dl = 36 p = 0,0000				
	ValSing.	ValProp.	%age Inertie	%age Cumulé	Chi2
1	0,5574	0,3107	58,9236	58,9236	9647,3580
2	0,3579	0,1281	24,2925	83,2162	3977,3286
3	0,2328	0,0542	10,2761	93,4922	1682,4691
4	0,1852	0,0343	6,5078	100,0000	1065,4910

4.2.11 Synthèse : menu à utiliser selon la forme des données d'entrée

On a observé plusieurs (2, 3 ou plus) variables nominales sur une population, et on souhaite explorer ces données à l'aide d'une ACM. Selon la forme sous laquelle ces données sont disponibles, on utilisera sous Statistica les menus suivants :

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

4.3 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 : $I = \frac{9-2}{2} = 3,5$.

4.3.1 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_ = 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

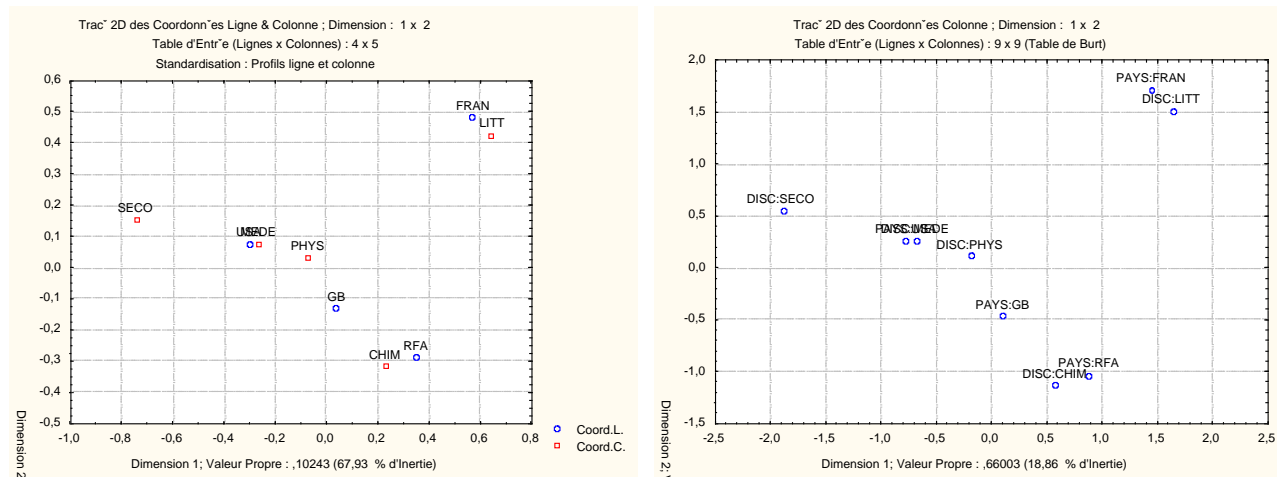
Nombre de Dims.	Valeurs Propres et Inertie de toutes les Dimensions (Nobel-Burt.sta)				
	Inertie Totale = 3,5000				
	ValSing.	ValProp.	%age Inertie	%age Cumulé	Chi2
1	0,812420	0,660027	18,85792	18,8579	407,5699
2	0,780833	0,609699	17,41999	36,2779	376,4924
3	0,712309	0,507383	14,49667	50,7746	313,3117
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 $1/2$.

Remarque : Le passage des valeurs propres λ_i aux valeurs propres μ_i est exactement la transformation des valeurs propres proposée par Benzécri pour l'ACM.

4.3.1.1 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.

4.4 Exercices

4.4.1 Exercice 1

Les fichiers Beverage.sta et Beverag2.sta sont des fichiers d'exemples fournis avec Statistica pour illustrer l'ACM. Ils ont été recopiés dans le répertoire Boissons du serveur de TD.

Le fichier d'exemple Beverage.sta contient des données collectées sur un groupe d'étudiants en maîtrise de gestion, hommes et femmes, de l'Université de Columbia, auxquels on a demandé d'indiquer la fréquence avec laquelle ils avaient acheté et consommé différents types de soda durant du mois écoulé. Les données pour les 34 individus ont été codifiées dans un tableau disjonctif complet (binaire) : un 1 a été saisi si l'individu a répondu avoir acheté ou consommé au moins une fois au cours du mois la boisson respective, et un 0 a été saisi si l'individu respectif a répondu avoir acheté ou consommé moins d'une fois dans le mois. Pour chacun des 8 sodas populaires utilisés dans cette étude, une seconde variable a été créée, codifiée comme l'inverse de la première variable respective, c'est-à-dire qu'un 1 a été saisi si la boisson respective n'a pas été consommée ni achetée, et 0 a été saisi si elle a été consommée ou achetée au cours du mois. Ci-dessous, observez une liste partielle des données codifiées de cette manière, pour 8 sodas courants. Ouvrez le fichier de données Beverage.sta situé dans le répertoire Boissons.

1) Réalisez une ACM sur ces données et retrouver ainsi les principales conclusions données dans l'aide de Statistica :

Il s'avère que toutes les boissons sont raisonnablement bien représentées par la solution à deux dimensions, mis à part Pepsi Light dont la valeur de Qualité est inférieure à 0,5.

Un examen attentif du graphique suggère que le premier axe oppose essentiellement les boissons allégées aux boissons classiques, alors que la seconde dimension oppose les colas aux autres sodas.

En outre, si vous examinez attentivement les statistiques des coordonnées de lignes, vous allez constater que les individus contribuant le plus à l'inertie de la seconde dimension sont les observations numéro 13 et 28. Ces points "définissent" presque à eux seuls la direction de la seconde dimension.

2) Reprenez ensuite l'étude à l'aide du fichier Beverag2.sta, dont les données sont saisies sous forme de tableau protocole.

4.4.2 Exercice 2

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 une feuille de données croisant la fréquentation (variable 1) et l'âge (variable 13) sous la forme d'un tableau de contingence et cinq feuilles de données contenant les tableaux de Burt obtenus en sélectionnant 3 ou 4 des items du questionnaire. Choisissez l'un de ces tableaux de Burt et analysez-le à l'aide d'une ACM.

Travail à rendre par mail à votre enseignant (Francois.Carpentier@univ-brest.fr) :

- Un classeur Statistica contenant les résultats numériques des analyses et les graphiques.
- Un fichier Word contenant une (brève) interprétation des résultats.

4.4.3 Exercice à rendre

Le fichier Enquete-Eleves-US.stw contient les données d'une étude trouvée (en 2006) sur le site <http://stat.genopole.cnrs.fr/teaching/>. Ces pages semblent malheureusement ne plus être accessibles en 2007.

Il s'agit d'une étude sur des élèves provenant de différentes écoles du milieu rural au milieu urbain. Les différentes questions et leurs modalités sont :

Gender: Boy or girl

Grade: 4, 5 or 6

Age: Age in years

Race: White, Other

Urban/Rural: Rural, Suburban, or Urban school district

School: Brentwood Elementary, Brentwood Middle, Ridge, Sand, Eureka, Brown, Main, Portage, Westdale Middle"

Goals: Student's choice in the personal goals question where options were 1 = Make Good Grades, 2 = Be Popular, 3 = Be Good in Sports.

Grades: Rank of "make good grades" (1=most important for popularity, 4=least important)

Sports: Rank of "being good at sports" (1=most important for popularity, 4=least important)

Looks: Rank of "being handsome or pretty" (1=most important for popularity, 4=least important)

Money: Rank of "having lots of money" (1=most important for popularity, 4=least important)

A partir d'un questionnement de votre choix sur ces données, faites une sélection de variables (au moins 4 variables) et procédez à une ACM en utilisant ces variables.

Interprétez ensuite (brièvement) des résultats obtenus.

Travail à rendre par mail à votre enseignant (Francois.Carpentier@univ-brest.fr) :

- Un classeur Statistica contenant les résultats numériques des analyses et les graphiques.
- Un fichier Word ou un rapport Statistica indiquant le questionnement initial et fournissant l'interprétation des résultats.