

Analyse Factorielle des Correspondances Multiples

Pr Roch Giorgi

 roch.giorgi@univ-amu.fr

Introduction (1)

- Généralisation de l'analyse factorielle des correspondances simples avec p ($p > 2$) variables qualitatives simultanément observées
- Mettre en évidence
 - ✓ Les relations entre les variables qualitatives
 - ✓ Les relations entre les individus statistiques
 - ✓ Les relations entre les variables telles qu'elles apparaissent à partir des relations entre modalités

Introduction (2)

- Réduction de dimension
- Construction de scores pour effectuer analyse de classification
- Souvent, analyse de questionnaires fermés, sondages

Problèmes - Objectifs

- Etude des individus : partition entre les individus
 - ✓ Individus différents si ils ne prennent pas les mêmes niveaux
- Etude des variables : variables synthétiques, liens entre les variables (étude des niveaux)
- Etude des catégories
 - ✓ Deux niveaux de variables différentes sont similaires si des personnes qui prennent ces niveaux sont les mêmes
 - ✓ Deux niveaux sont similaires si les personnes qui prennent ces niveaux se comportent de la même manière (ils prennent les mêmes niveaux pour les autres variables)

⇒ Caractérisation des groupes d'individus par les niveaux

Autour des Données (1)

- Exemple : 3 questions à 10 sujets : Sexe (F ou H), niveau de revenus (M ou E), préférence sur un sujet donné (A, B ou C)

Tableau d'effectifs

ID	Sexe	Revenu	Préférence
1	F	M	A
2	F	E	B
3	F	E	C
4	H	M	B
5	H	E	C
...

Tableau disjonctif complet

ID	Sexe:F	Sexe:H	Rev:M	Rev:E	Préf:A	Préf:B	Préf:C
1	1	0	1	0	1	0	0
2	1	0	0	1	0	1	0
3	1	0	0	1	0	0	1
4	0	1	1	0	0	1	0
5	0	1	0	1	0	0	1

Autour des Données (2)

Tableau de Burt

	Sexe:F	Sexe:H	Rev:M	Rev:E	Préf:A	Préf:B	Préf:C
Sexe:F	5	0	2	3	2	1	2
Sexe:H	0	5	3	2	1	3	1
Rev:M	2	3	5	0	3	2	0
Rev:E	3	2	0	5	0	2	3
Préf:A	2	1	3	0	3	0	0
Préf:B	1	3	2	2	0	4	0
Préf:C	2	1	0	3	0	0	3

Juxtaposition de tableaux de contingence des variables prises 2 à 2x2

ACM et Tableau de Données

- Sur tableau disjonctif complet (matrice d'indicatrices)
- Sur tableau de Burt
 - ✓ Une AFC sur un tableau de Burt équivaut à une ACM sur un tableau disjonctif
 - ✓ Donne des résultats sur les niveaux

Remarque

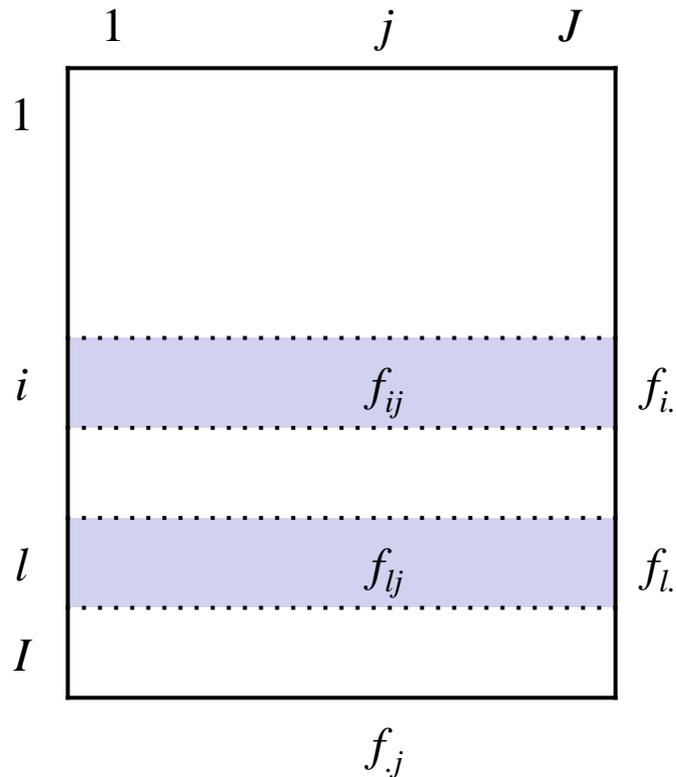
- ACM peut être considérée de différentes manières
 - ✓ ACP sur un tableau de données avec des poids particuliers pour les variables
 - ✓ AFC sur une matrice d'indicatrice
 - ✓ AFC sur un tableau de Burt

ACM – Principes Généraux (1)

- Comme en AFC, on s'intéresse aux
 - ✓ Profils-lignes
 - ✓ Profils-colonnes
 - ✓ Taux de liaison
- Propriété d'équivalence distributionnelle
 - ✓ Le regroupement de 2 lignes ayant un profil de réponse similaire n'impacte pas les autres profils-lignes, ni les autres profils-colonnes
- Profils-colonnes et profils-lignes moyens
 - ✓ Distance entre les individus

ACM – Principes Généraux (2)

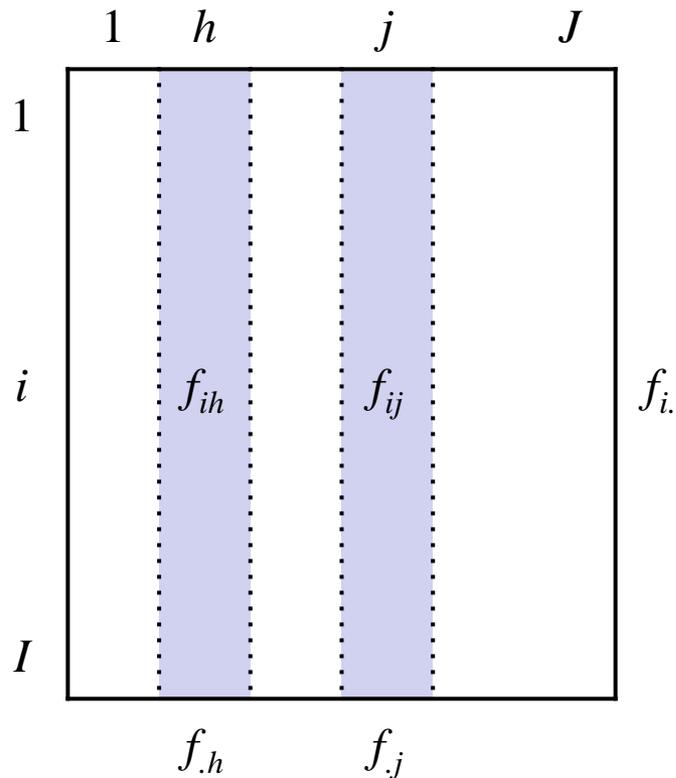
- Distance entre 2 profils-lignes



$$d^2(i, l) = \sum_{j=1}^J \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{lj}}{f_{l.}} \right)^2$$

ACM – Principes Généraux (3)

- Distance entre 2 profils-colonnes



$$d^2(h, j) = \sum_{i=1}^I \frac{1}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ih}}{f_{.h}} \right)^2$$

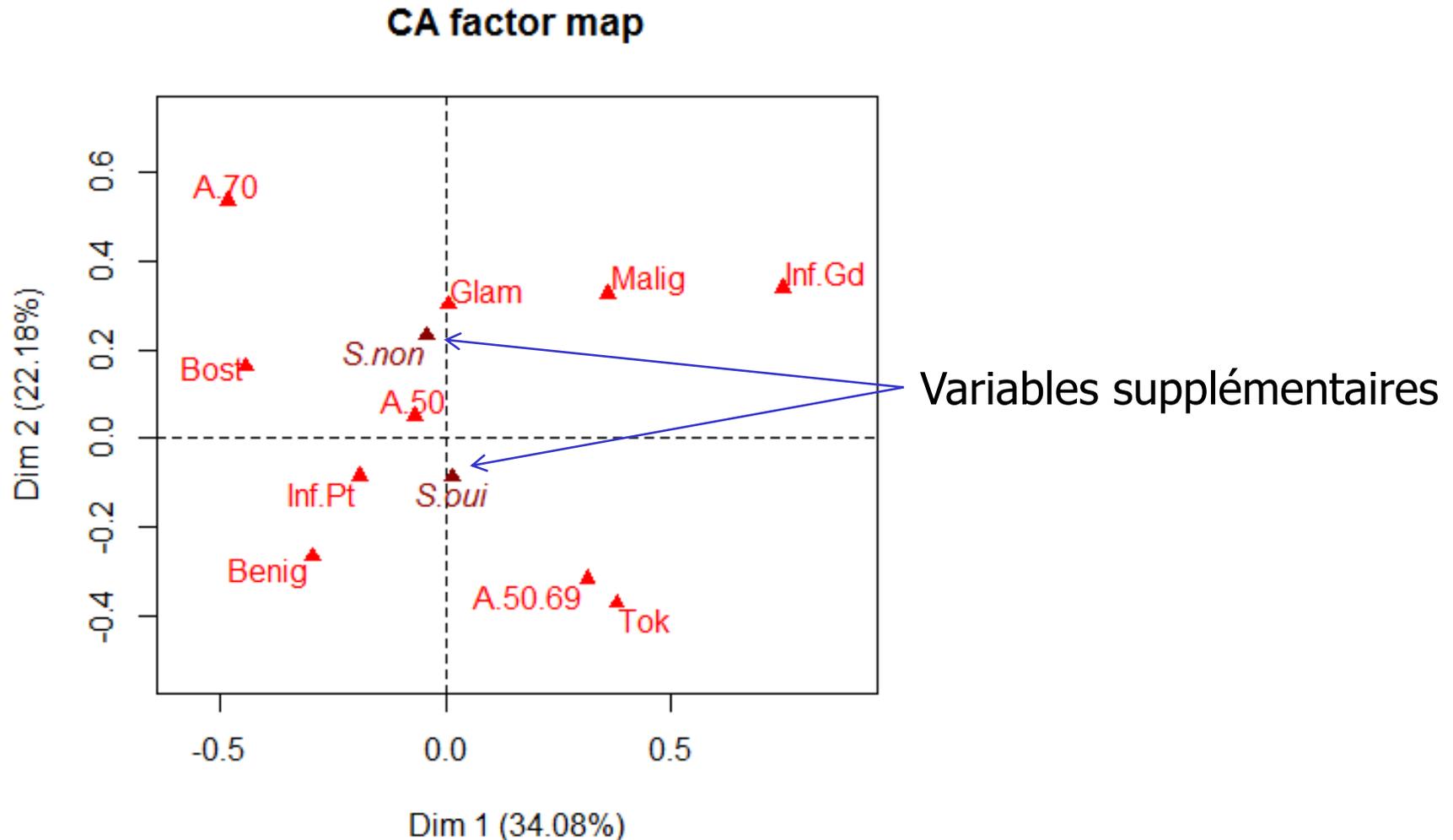
Exemple 1 : ACM sur Tableau de Burt (1)

- Enquête réalisée dans 3 centres hospitaliers sur des patientes atteintes d'un cancer du sein
 - Données sous forme d'un tableau de Burt
 - Objectif : étudier la survie à 3 ans
- ⇒ mise en œuvre d'une analyse par AFC

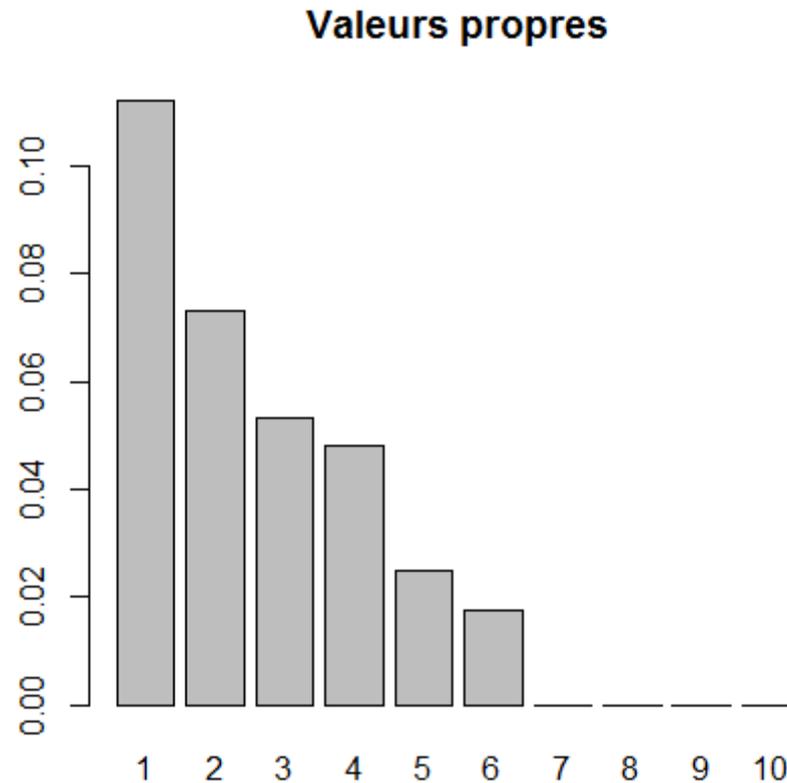
Exemple 1 : ACM sur Tableau de Burt (2)

	Bost	Glam	Tok	A.50	A.50-69	A.70	S.non	S.oui	Inf.Gd	Inf.Pt	Benig	Malig
Bost	253	0	0	122	58	73	82	171	33	220	159	94
Glam	0	221	0	109	71	41	68	153	37	184	102	119
Tok	0	0	290	120	151	19	60	230	84	206	159	131
A.50	122	109	120	351	0	0	93	258	71	280	200	151
A.50-69	58	71	151	0	280	0	68	212	63	217	146	134
A.70	73	41	19	0	0	133	49	84	20	113	74	59
S.non	82	68	60	93	68	49	210	0	44	166	97	113
S.oui	171	153	230	258	212	84	0	554	110	444	323	231
Inf.Gd	33	37	84	71	63	20	44	110	154	0	32	122
Inf.Pt	220	184	206	280	217	113	166	444	0	610	388	222
Benig	159	102	159	200	146	74	97	323	32	388	420	0
Malig	94	119	131	151	134	59	113	231	122	222	0	344

Exemple 1 : ACM sur Tableau de Burt (3)



Exemple 1 : ACM sur Tableau de Burt (4)



valeurs propres possibles = # modalités - # variables

Exemple 1 : ACM sur Tableau de Burt (5)

Coordonnées des variables sur les Dim.

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
Bost	-0,4422	0,1628	0,2949	-0,2417	-0,2671
Glam	0,0061	0,3016	-0,4883	0,3967	0,0378
Tok	0,3811	-0,3718	0,1149	-0,0914	0,2042
A.50	-0,0679	0,0504	-0,3209	-0,3534	0,0220
A.50.69	0,3151	-0,3179	0,1611	0,3055	-0,2115
A.70	-0,4841	0,5362	0,5078	0,2893	0,3871
Inf.Gd	0,7514	0,3392	0,1851	-0,2657	0,1336
Inf.Pt	-0,1897	-0,0856	-0,0467	0,0671	-0,0337
Benig	-0,2954	-0,2668	-0,0180	-0,0155	0,1007
Malig	0,3606	0,3258	0,0220	0,0190	-0,1230

Exemple 1 : ACM sur Tableau de Burt (6)

cosinus carré de la coordonnée d'une variable sur les Dim.

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
Bost	0,4411	0,0598	0,1962	0,1318	0,1609
Glam	0,0001	0,1812	0,4751	0,3135	0,0028
Tok	0,3940	0,3751	0,0358	0,0227	0,1131
A.50	0,0195	0,0107	0,4345	0,5269	0,0020
A.50.69	0,2672	0,2719	0,0698	0,2512	0,1203
A.70	0,2299	0,2820	0,2530	0,0821	0,1470
Inf.Gd	0,6174	0,1258	0,0375	0,0772	0,0195
Inf.Pt	0,6174	0,1258	0,0375	0,0772	0,0195
Benig	0,4629	0,3778	0,0017	0,0013	0,0538
Malig	0,4629	0,3778	0,0017	0,0013	0,0538

Exemple 1 : ACM sur Tableau de Burt (7)

Contribution des variables sur les Dim.

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
Bost	14,4431	3,0072	13,4867	10,0619	23,7514
Glam	0,0024	9,0165	32,3036	23,6698	0,4159
Tok	12,2981	17,9864	2,3453	1,6495	15,9126
A.50	0,4732	0,4004	22,1543	29,8294	0,2241
A.50.69	8,1198	12,6933	4,4520	17,7906	16,4816
A.70	9,1028	17,1517	21,0216	7,5769	26,2299
Inf.Gd	25,3883	7,9495	3,2346	7,4006	3,6190
Inf.Pt	6,4095	2,0069	0,8166	1,8683	0,9136
Benig	10,6994	13,4124	0,0834	0,0689	5,6066
Malig	13,0633	16,3756	0,1019	0,0841	6,8453

- A partir de <http://factominer.free.fr/factomethods/analyse-des-correspondances-multiples.html>
- Etude sur des consommateurs de thé (n=300)
 - ✓ Manière de consommer
 - ✓ Leur image du thé
 - ✓ Plus caractéristiques descriptives
- Objectifs
 - ✓ Y a t-il des groupes particuliers de consommateurs ?
 - ✓ Y a t-il des relations entre des variables, des associations entre les modalités ?
 - ✓ Quels sont les caractéristiques des groupes de consommateurs selon les modalités ?

- Variables

- ✓ Les 18 premières sont les variables « actives » (manière de consommer, image du thé)
- ✓ Suivies de 18 variables descriptives supplémentaires (l'âge quantitative, puis d'autres variables qualitatives, dont l'âge : sexe, catégorie socio-professionnelle, pratique régulière d'un sport,...)

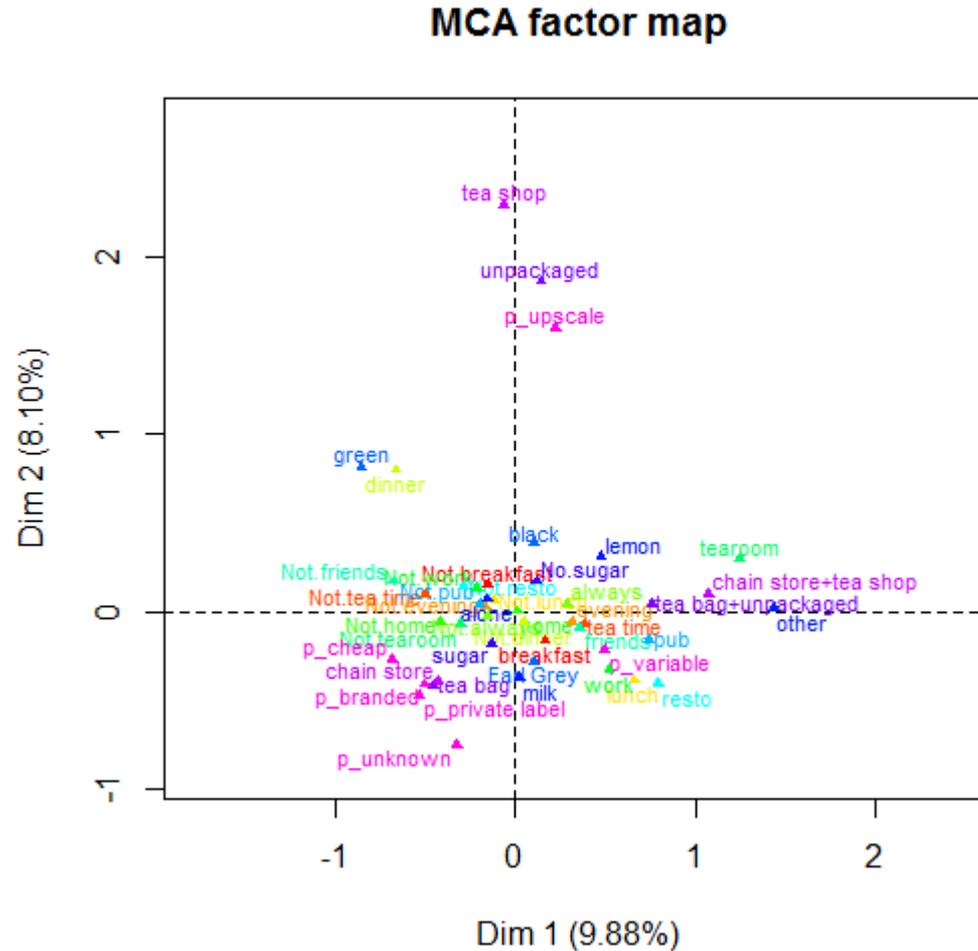
Tableau des effectifs

	breakfast	tea.time	evening	lunch	dinner	...
1	breakfast	Not.tea time	Not.evening	Not.lunch	Not.dinner	...
2	breakfast	Not.tea time	Not.evening	Not.lunch	Not.dinner	...
3	Not.breakfast	tea time	evening	Not.lunch	dinner	...
4	Not.breakfast	Not.tea time	Not.evening	Not.lunch	dinner	...
5	breakfast	Not.tea time	evening	Not.lunch	Not.dinner	...
6	Not.breakfast	Not.tea time	Not.evening	Not.lunch	dinner	...
..

Exemple 2

(6)

Modalités

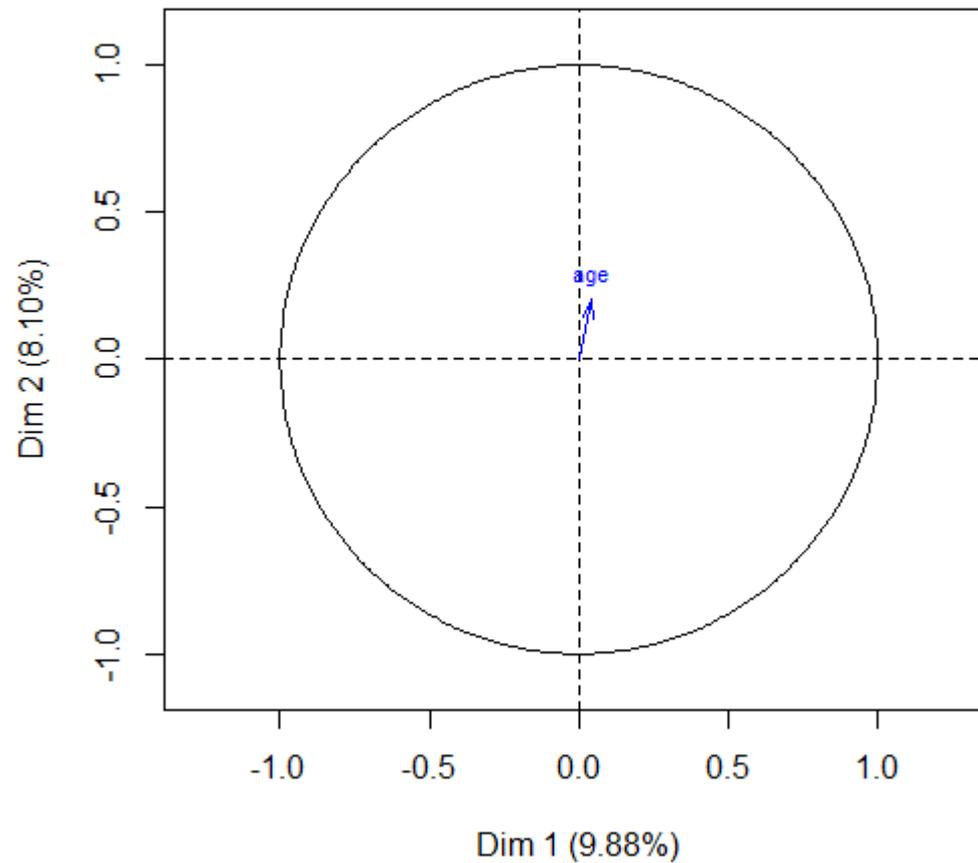


Exemple 2

(7)

Age

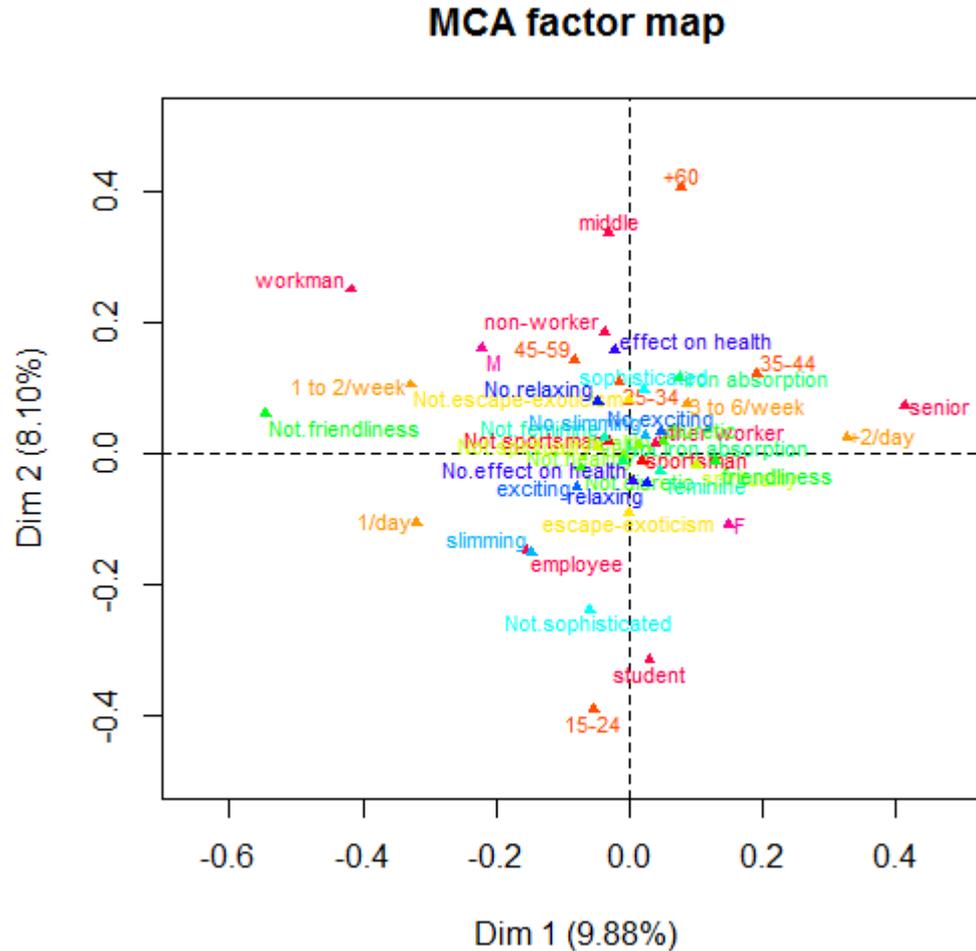
Supplementary variables on the MCA factor map



Exemple 2

(8)

Variables illustratives



Exemple 2

(9)

Description Dim 1 : variables qualitatives

	R2	p.value
where	0,4179	0,0000
tearoom	0,3719	0,0000
how	0,2988	0,0000
friends	0,2432	0,0000
resto	0,2265	0,0000
tea.time	0,1920	0,0000
price	0,2161	0,0000
pub	0,1472	0,0000
work	0,1115	0,0000
How	0,1029	0,0000
Tea	0,0895	0,0000
lunch	0,0746	0,0000
frequency	0,0944	0,0000
...

Exemple 2

(10)

Description Dim 1 : modalités

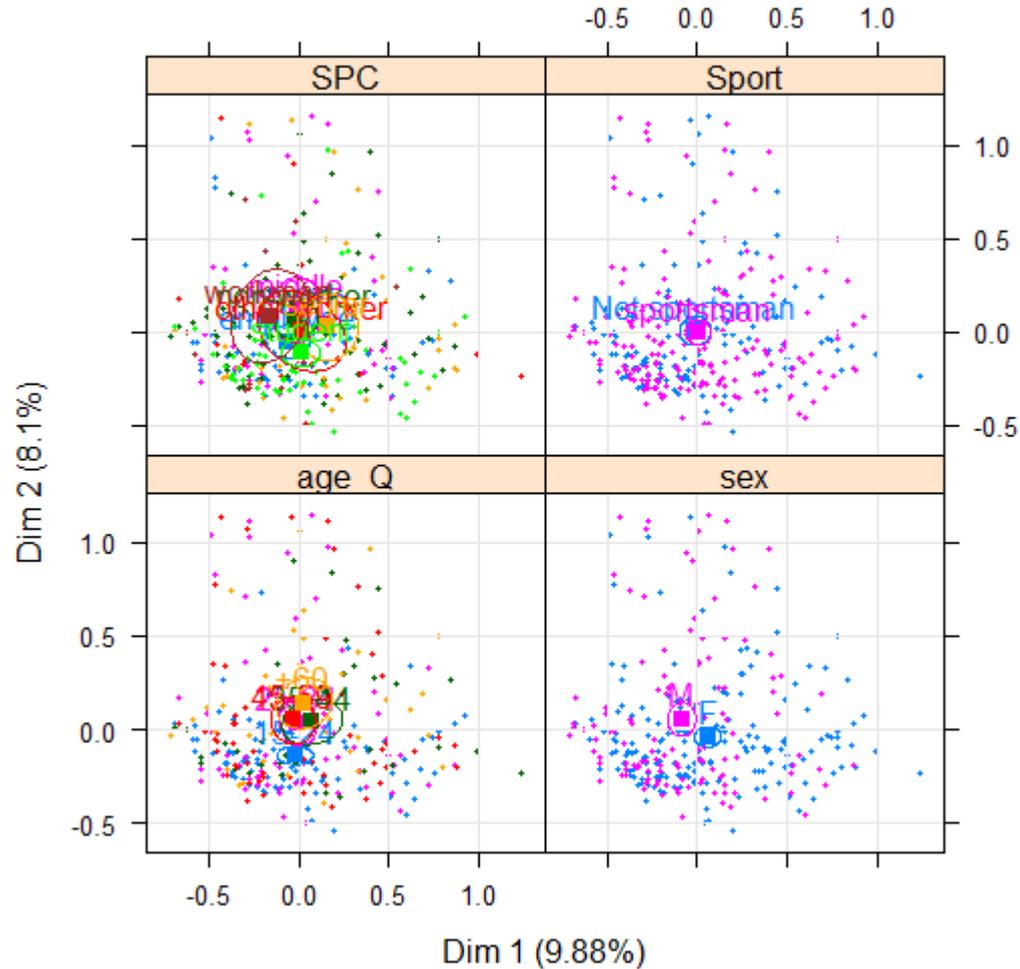
	Estimate	p.value
tearoom	0,2973	0,0000
chain store+tea shop	0,3385	0,0000
friends	0,1995	0,0000
resto	0,2080	0,0000
tea time	0,1701	0,0000
tea bag+unpackaged	0,2346	0,0000
pub	0,1814	0,0000
work	0,1417	0,0000
p_variable	0,2760	0,0000
lunch	0,1486	0,0000
friendliness	0,1302	0,0000
other	0,3819	0,0000
...

... pour autres Dim. et variable quantitative

Exemple 2

(11)

Ellipses de confiance



Mise en Pratique

- Analyse univariée
- Regrouper / Eliminer modalités trop rares
- Eliminer variables non discriminantes ou redondantes
- Recoder en classes (effectifs voisins de préférence) les variables quantitatives
- Si beaucoup de variables, les regrouper par thèmes
- Faire AFC de chaque groupe \Rightarrow sélection des variables les plus pertinentes pour l'objectif de l'étude
- Faire l'ACM des quelques variables importantes
- Si présence d'une variable à expliquer, réaliser une modélisation des données (logit, log-linéaire,...)

Source : *Baccini A, Besse P. Statistique descriptive multidimensionnelle. 1999.*

Sources

- Besse P, Baccini A. Statistique descriptive multidimensionnelle. 1999.
- Josse J, Husson F, Lê S. Multiple correspondance analysis. 11 aout 2008.
- FactoMineR : <http://factominer.free.fr/>