

TD : Analyse des données
Analyse factorielle des correspondances (AFC)

Exercice 1

On veut étudier la liaison entre les caractères : " X : être fumeur" (plus de 20 cigarettes par jour, pendant 10 ans) et " Y : avoir un cancer de la gorge", sur une population de 1000 personnes, dont 500 sont atteintes d'un cancer de la gorge. Voici les résultats observés :

<i>Observé</i>	cancer	non cancer	total
fumeur	342	258	600
non fumeur	158	242	400
total	500	500	1000

- 1) Donner la distribution de X conditionnellement à Y sous forme de tableau.
- 2) Faire un test d'indépendance pour établir la liaison entre ces caractères.

Exercice 2

Une enquête a été effectuée auprès de cent fumeurs afin de choisir les noms de deux nouvelles marques de cigarettes. La première marque est destinée à une clientèle masculine : l'homme ciblé est un connaisseur distingué, raffiné mais viril, de niveau socio-économique élevé. La seconde vise un public féminin, élégant, assuré et dynamique.

Douze marques possibles ont été retenues : Orly (On), Alezan (Ale), Corsaire (Cor), Directoire (Dir), Ducat (Duc), Fontenoy (Fon), Icare (rca), Zodiac (Zod), Pavois (Pav), Cocker (Coc), Escale (Esc), Hôtesse (Hôt). Pour évaluer leur image auprès du public, onze attributs ont été proposés aux fumeurs : vieillot désuet (Vi-D), nouveau riche (N-R), sorbe élégant (S-E), cocasse ridicule (C-R), racé (Rac), mièvre (Miè), distingué (Dis), vulgaire commun (V-C), pour un homme (PuH), pour une femme (PuF) et une petite nature (PPN).

Chaque personne interrogée devait successivement pour chaque attribut, désigner une ou plusieurs marques lui correspondant.

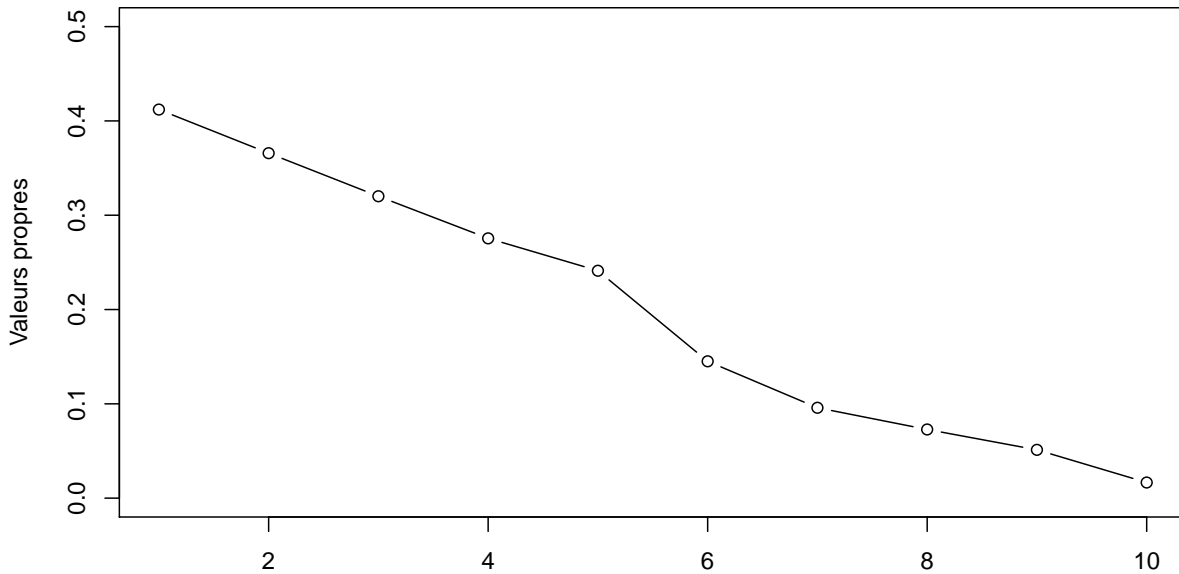
Le tableau des données est le suivant :

	On	Ale	Cor	Dir	Duc	Fon	rca	Zod	Pav	Coc	Esc	Hôt
Vi-D	1	2	14	38	18	10	9	5	9	4	0	1
N-R	20	9	1	11	10	9	1	1	20	9	7	12
S-E	9	23	1	15	7	11	6	2	7	12	3	17
C-R	1	3	15	15	6	5	12	18	4	25	2	2
Rac	4	33	7	8	3	6	6	4	4	15	5	3
Miè	3	9	1	7	7	5	12	9	6	9	6	13
Dis	11	9	1	17	4	21	6	1	5	4	5	27
V-C	4	4	32	2	6	0	9	7	3	10	12	7
PuH	9	12	23	4	7	13	5	5	10	5	13	9
PuF	9	3	9	8	4	2	6	8	1	6	23	33
PPN	7	5	2	7	11	2	6	11	9	24	10	5

- 1) Le tableau des données est-il un tableau de contingence ? Quelles sont les marques les plus fréquemment citées ? Que peut-on dire des attributs ?
- 2) Les publics visés ont-ils des points communs ? Quels sont les attributs correspondants ? Pourquoi t-on proposé des attributs ne leur correspond pas ?
- 3) En examinant les valeurs propres et leur diagramme, déterminer les axes principaux qu'il faut conserver dans les analyses.
- 4) A l'aide du premier plan factoriel, expliquer la différence entre les marques les plus distantes les unes des autres, et la ressemblance entre les marques les plus proches. Quelles sont les marques bien représentées sur ce plan ? Quelles sont les marques importantes dans l'interprétation des deux premiers axes ?
- 5) Quelle interprétation peut-on proposer à l'attribut "Pour une femme" tel qu'il apparaît le long de l'axe 1 ?

Annexe : listening du logiciel

Eboulis des valeurs propres



Etude des lignes du tableau

	Axe 1				Axe 2		
	Poids	coord	ctr	cos2	coord	ctr	cos2
Vi-D	0.096	0.203	2.320	0.044	-0.691	34.157	0.514
N-R	0.095	-0.480	12.878	0.400	-0.056	0.226	0.006
S-E	0.097	-0.427	10.458	0.635	-0.158	1.823	0.087
C-R	0.093	0.616	20.809	0.599	-0.272	5.160	0.117
Rac	0.084	-0.104	0.536	0.014	-0.204	2.626	0.055
Miè	0.075	-0.045	0.089	0.009	-0.001	0.000	0.000
Dis	0.096	-0.611	21.054	0.644	0.058	0.244	0.006
V-C	0.083	0.767	28.689	0.646	0.433	11.597	0.206
PuH	0.099	0.117	0.795	0.046	0.237	4.167	0.189
PuF	0.097	-0.104	0.615	0.015	0.741	39.584	0.736
PPN	0.085	0.187	1.756	0.083	-0.081	0.416	0.015

Etude des colonnes du tableau

		Axe 1			Axe 2		
	Poids	coord	ctr	cos2	coord	ctr	cos2
On	0.067	-0.479	9.102	0.469	0.220	2.430	0.099
Ale	0.097	-0.349	6.941	0.143	-0.150	1.633	0.026
Cor	0.091	0.835	37.507	0.629	0.270	4.979	0.066
Dir	0.114	-0.070	0.326	0.009	-0.555	26.220	0.563
Duc	0.072	0.104	0.453	0.039	-0.323	5.569	0.381
Fon	0.072	-0.460	9.009	0.424	-0.199	2.147	0.080
rca	0.067	0.290	3.340	0.357	-0.082	0.337	0.028
Zod	0.061	0.565	11.490	0.543	-0.009	0.004	0.000
Pav	0.067	-0.218	1.876	0.108	-0.224	2.521	0.115
Coc	0.106	0.286	5.097	0.188	-0.151	1.806	0.052
Esc	0.074	0.083	0.302	0.012	0.711	27.986	0.851
Rôt	0.111	-0.471	14.556	0.341	0.542	24.368	0.449

CA factor map

