

INTRODUCTION AUX ANALYSES MULTIDIMENSIONNELLES EXERCICES DE SYNTHESE CORRIGES

Exercice 1

On a rassemblé les résultats de 15 enfants de 10 ans à 6 subtests du WISC (scores 0 à 5). Les variables observées sont : CUB (Cubes de Kohs), PUZ (Assemblage d'objets), CAL (Calcul mental), MEM (Mémoire immédiate des chiffres), COM (Compréhension de phrases), VOC (Vocabulaire). Le protocole observé est le suivant :

WISC	CUB	PUZ	CAL	MEM	COM	VOC
I1	5	5	4	0	1	1
I2	4	3	3	2	2	1
I3	2	1	2	3	2	2
I4	5	3	5	3	4	3
I5	4	4	3	2	3	2
I6	2	0	1	3	1	1
I7	3	3	4	2	4	4
I8	1	2	1	4	3	3
I9	0	1	0	3	1	0
I10	2	0	1	3	1	0
I11	1	2	1	1	0	1
I12	4	2	4	2	1	2
I13	3	2	3	3	2	3
I14	1	0	0	3	2	2
I15	2	1	1	2	3	2

On traite ces données par une analyse en composantes principales normée. Les principaux résultats de cette ACP sont indiqués ci-dessous :

Corrélations

	CUB	PUZ	CAL	MEM	COM	VOC
CUB	1,0000	0,7320	0,9207	-0,4491	0,3086	0,2735
PUZ	0,7320	1,0000	0,7510	-0,6143	0,2814	0,2850
CAL	0,9207	0,7510	1,0000	-0,3685	0,4077	0,4869
MEM	-0,4491	-0,6143	-0,3685	1,0000	0,3032	0,2023
COM	0,3086	0,2814	0,4077	0,3032	1,0000	0,7819
VOC	0,2735	0,2850	0,4869	0,2023	0,7819	1,0000

Val. Propres (matrice de corrél.) & stat. associées Variables actives seules

	Val. propr	% Total variance	Cumul Val. propr	Cumul %
1	3,2581	54,3020	3,2581	54,3020
2	1,8372	30,6194	5,0953	84,9214
3	0,4430	7,3831	5,5383	92,3044
4	0,2538	4,2292	5,7920	96,5337
5	0,1679	2,7990	5,9600	99,3327
6	0,0400	0,6673	6,0000	100,0000

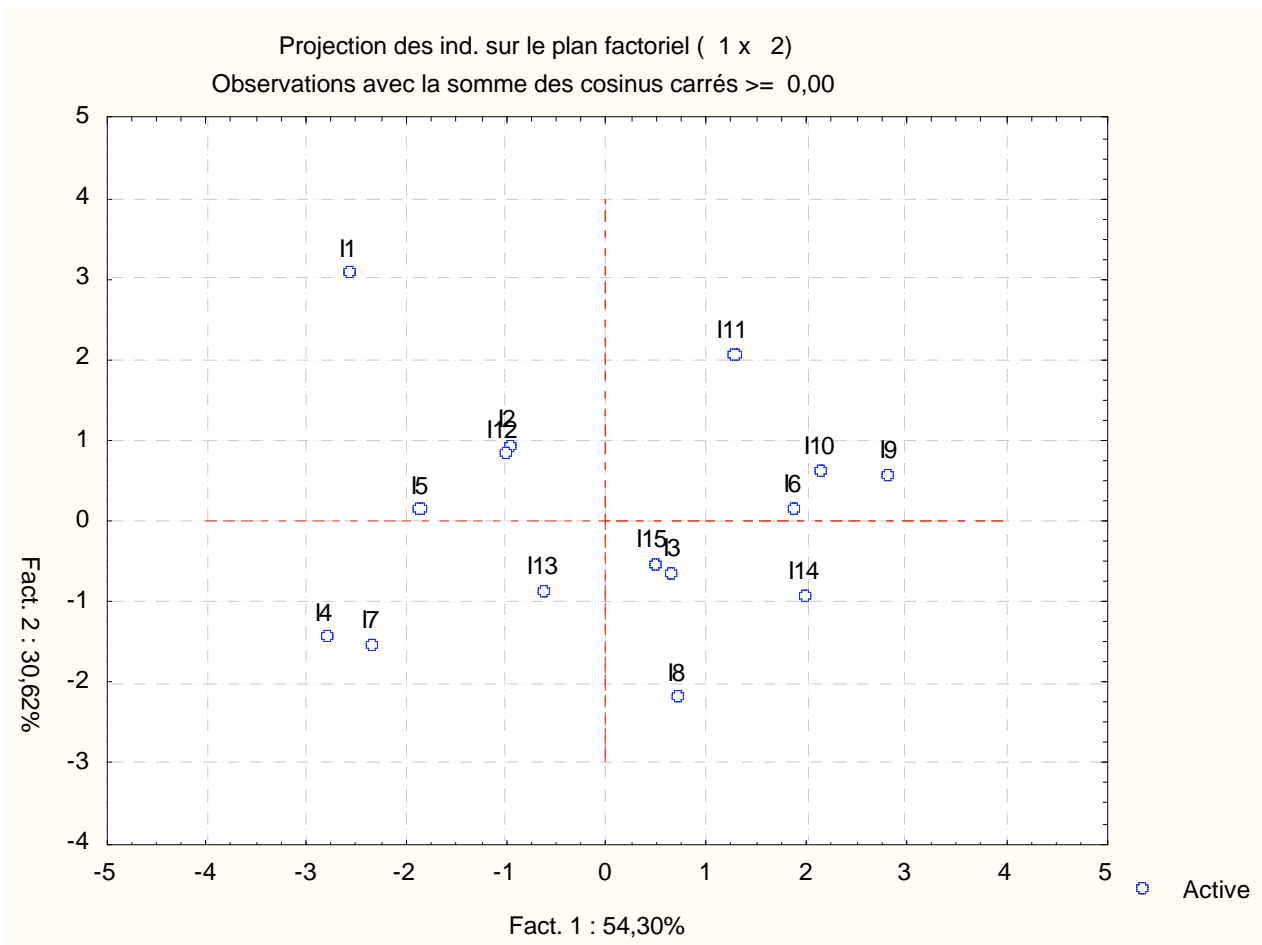
Scores, contributions et qualités de représentation des individus

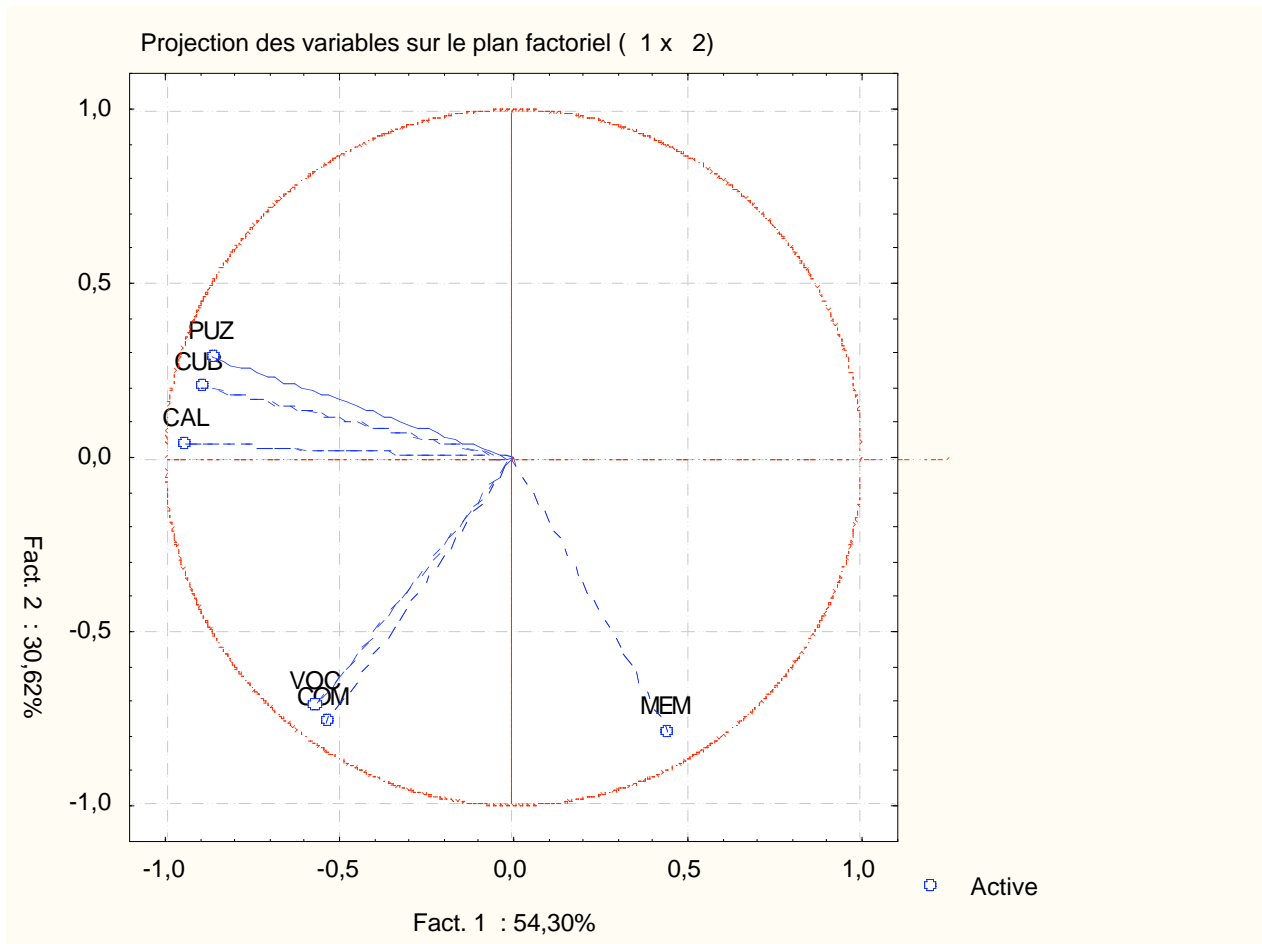
	Score Fact. 1	Score Fact. 2	Contribution Fact.1	Contribution Fact.2	Cos ² Fact. 1	Cos ² Fact. 1&2
I1	-2,5616	3,0568	13,43	33,91	0,4078	0,5807
I2	-0,9661	0,9370	1,91	3,19	0,3907	0,3676
I3	0,6765	-0,6624	0,94	1,59	0,4446	0,4263
I4	-2,7969	-1,4636	16,01	7,77	0,7160	0,1961
I5	-1,8423	0,1211	6,95	0,05	0,8142	0,0035
I6	1,8891	0,1350	7,30	0,07	0,8426	0,0043
I7	-2,3396	-1,5487	11,20	8,70	0,6028	0,2641
I8	0,7275	-2,2054	1,08	17,65	0,0816	0,7499
I9	2,8400	0,5423	16,50	1,07	0,8745	0,0319
I10	2,1733	0,6117	9,66	1,36	0,7433	0,0589
I11	1,2940	2,0373	3,43	15,06	0,2256	0,5592
I12	-0,9947	0,8181	2,02	2,43	0,3120	0,2110
I13	-0,6099	-0,8730	0,76	2,77	0,1949	0,3994
I14	2,0150	-0,9470	8,31	3,25	0,7548	0,1667
I15	0,4957	-0,5591	0,50	1,13	0,1151	0,1464

Saturations, contributions et qualités de représentation des variables

	Saturation Fact. 1	Saturation Fact. 2	Contribution Fact.1	Contribution Fact.2	Cos ² Fact.1	Cos ² Fact.2
CUB	-0,8970	0,2018	0,25	0,02	0,8046	0,8453
PUZ	-0,8652	0,2883	0,23	0,05	0,7485	0,8316
CAL	-0,9458	0,0390	0,27	0,00	0,8945	0,8960
MEM	0,4449	-0,7861	0,06	0,34	0,1980	0,8160
COM	-0,5382	-0,7627	0,09	0,32	0,2897	0,8714
VOC	-0,5683	-0,7156	0,10	0,28	0,3229	0,8350

Représentation des individus dans le premier plan factoriel





- 1) Etude du tableau des valeurs propres
 a) A quoi correspond la somme des valeurs propres ?

La somme des valeurs propres est égale à 6. Dans le cas d'une ACP normée telle que celle qui est effectuée ici, cette somme est égale au nombre de variables. Cette valeur correspond également à la variance du nuage des individus.

- b) On choisit de n'étudier que les deux premières composantes principales. Justifier ce choix en analysant le tableau des valeurs propres.

On peut choisir de ne s'intéresser qu'aux valeurs propres dont la contribution à la variance est supérieure à la moyenne. Ceci revient à étudier les composantes principales correspondant à des valeurs propres supérieures à 1. Or, seules les deux premières valeurs propres vérifient cette propriété sur l'exemple fourni.

- 2) Etude du tableau des corrélations
 a) Quels sont les subtests les plus fortement corrélés entre eux ?

Les corrélations les plus fortes sont celles correspondant aux coefficients de corrélation les plus proches de 1 ou -1. Ici, le coefficient de corrélation dont la valeur absolue est la plus proche de 1 est celui qui relie CUB et CAL.

- b) Comment s'organisent les signes "+" et les signes "-" dans le tableau des coefficients de corrélation. Commenter.

La plupart des coefficients de corrélation sont positifs. Autrement dit, un bon résultat à un test est, en règle général, lié à de bons résultats aux autres tests. On note toutefois une exception remarquable : la variable MEM (mémoire immédiate des chiffres) est corrélée négativement à 3 autres variables : CUB, PUZ et CAL. Mais ce sont les seuls coefficients négatifs du tableau.

3) Etude des qualités de représentation dans le premier plan principal

Quel est l'individu le moins bien représenté par le premier plan principal ? Quel est l'individu le mieux représenté ?

Les qualités de représentation des individus dans le premier plan factoriel sont obtenues en additionnant les résultats des deux colonnes "Cos²" du tableau relatif aux individus. Le sujet le mieux représenté est I1, avec une qualité de représentation de 0,9885, le plus mal représenté est I15, avec une qualité de 0,2615.

4) Etude du nuage des individus.

a) Quels sont les individus dont la contribution à la formation de la première composante principale est supérieure à la moyenne ? Pour chacun d'eux, préciser le signe de la coordonnée correspondante.

Les individus envisagés ici sont ceux dont la contribution est supérieure à 6,67%. Ce sont les individus suivants :

-	+
I4 (16,01%)	I9 (16,5%)
I1 (13,43%)	I10 (9,66%)
I7 (11,2%)	I14 (8,31%)
I5 (6,95%)	

Caractériser cet axe en termes d'opposition entre individus.

Cet axe oppose les individus I4, I1, I7 et I5 d'une part aux individus I9, I10 et I14 d'autre part.

Les sujets du premier groupe obtiennent des scores généralement plus élevés que ceux du 2^e groupe, particulièrement sur les variables CUB, PUZ et CAL.

b) Même question pour la deuxième composante principale.

-	+
I8 (17,65%)	I1 (33,9%)
I4 (7,73%)	I11 (15,06%)

Cet axe oppose les individus I8 et I4 d'une part aux individus I1 et I11 d'autre part. On peut noter que l'individu I1 explique à lui seul plus du tiers de la variance de cet axe.

5) Etude du nuage des variables

a) La représentation graphique des variables montre qu'elles sont toutes très bien représentées dans le plan (CP1, CP2). Justifier cette affirmation.

Les extrémités des vecteurs représentant les variables sont toutes très proches du cercle des corrélations, ce qui montre que les qualités de représentation des variables sont satisfaisantes. D'autre part la dernière colonne du tableau des résultats relatifs aux variables montre que la variable la plus mal représentée est MEM, avec une qualité qui atteint quand même 0,8816.

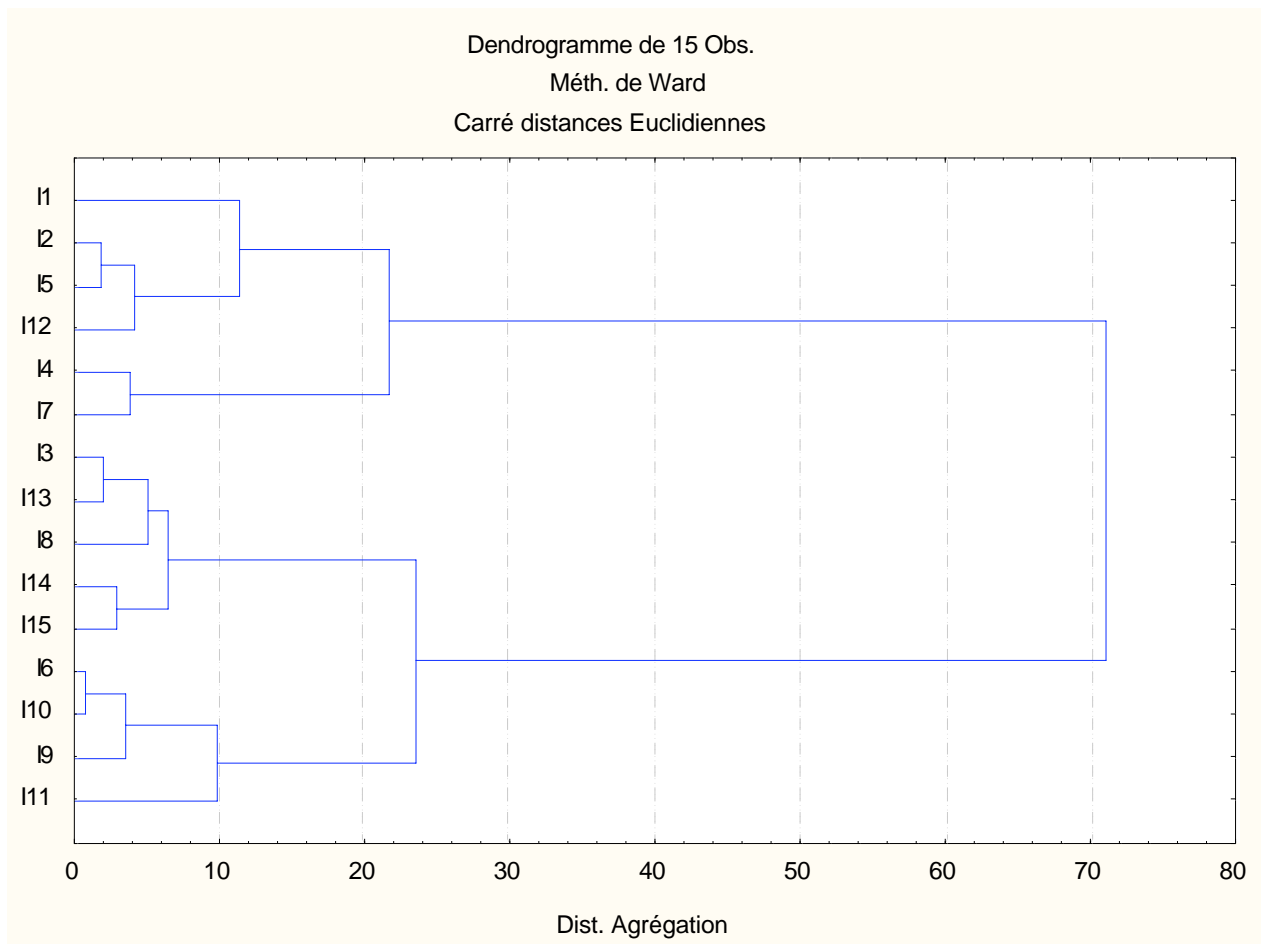
b) Quelles sont les variables qui sont corrélées positivement avec le premier facteur principal ? Quelles sont celles qui sont corrélées négativement ? Comment peut-on caractériser cet axe par rapport aux variables de départ ?

Les corrélations peuvent être lues sur le dessin ou sur le tableau des saturations. Seule la variable MEM est corrélée positivement avec le premier axe, toutes les autres sont corrélées négativement. Ainsi, cet axe oppose le résultat au test MEM à tous les autres tests.

c) Quelles sont les variables qui ont joué un rôle dominant dans la formation du deuxième axe.

Les corrélations les plus fortes sont observées pour les variables MEM, COM et VOC. Ces variables sont toutes trois corrélées négativement avec la variable CP2.

6) En utilisant les scores centrés-réduits des individus, on a obtenu le graphique suivant :



a) Quelle est la méthode dont les résultats sont représentés à l'aide de graphiques de ce type ?

Il s'agit ici d'une classification ascendante hiérarchique.

b) L'examen de ce graphique suggère de réaliser une partition des sujets soit en deux classes, soit en quatre classes. Justifier.

On observe un saut brutal de l'indice d'agrégation lors de la formation de la dernière classe. Le passage de 3 à 2 classes se fait avec des indices d'agrégation peu différents. En revanche, on avait également une forte progression de l'un indice d'agrégation lors du passage de 4 à 3 classes.

c) On retient finalement la partition en 4 classes.
Décrire en extension chacune des classes de la partition.

Classe I : I1, I2, I5, I12

Classe II : I4, I7

Classe III : I3, I13, I8, I14, I15

Classe IV : I6, I10, I9, I11

d) Comment peut-on décrire les positions des 4 classes précédentes sur le graphique de l'ACP ?

Les classes correspondent approximativement aux 4 quadrants définis par les deux premières variables principales. Ainsi :

La classe I correspond à $CP1 < 0$ et $CP2 > 0$

La classe II correspond à $CP1 < 0$ et $CP2 < 0$, mais ne contient pas I13. Il vaudrait mieux la décrire par : $CP1 < -1, CP2 < -1$

La classe III correspond à $CP1 > -1, CP2 < 0$

La classe IV correspond à $CP1 > 0$ et $CP2 > 0$.

e) On sait que la définition d'une hiérarchie de classes, avec l'indice d'agrégation correspondant, permet de

définir sur les individus une distance dite *ultramétrique*. Evaluer à partir du graphique la distance ultramétrique séparant le sujet I15 et le sujet I6.

La distance ultramétrique associée à une CAH est obtenue en considérant, pour chaque couple d'objets, la plus petite classe les contenant et en notant l'indice d'agrégation correspondant. Pour les objets I15 et I6, la classe correspondante est l'avant-dernière, avec un indice d'agrégation proche de 24 (valeur lue sur le graphique).

Exercice 2

Une entreprise veut choisir un nom pour un nouveau produit, qui doit paraître de qualité supérieure, prestigieuse, luxueux, convenant à un public masculin, connaisseur, raffiné, distingué, de niveau socio-économique élevé...¹

L'entreprise hésite entre 12 marques (cf 1ère ligne du tableau ci-dessous). Pour faire un choix entre ces marques, un échantillon de clients potentiels a été interrogé. Onze attributs leur étaient proposés (cf. 1ère colonne du tableau). Pour chaque attribut, la personne interrogée devait désigner une ou plusieurs marques auxquelles l'attribut se rapportait le mieux.

Les résultats obtenus sont les suivants :

	Orly	Alezan	Corsaire	Directoire	Ducat	Fontenoy	Icare	Zodiaque	Pavois	Cocker	Escale	Hotesse	Total
Vieillot	1	2	14	38	18	10	9	5	9	4	0	1	111
Nouveau-riche	20	9	1	11	10	9	1	1	20	9	7	12	110
Sobre-élégant	9	23	1	15	7	11	6	2	7	12	3	17	113
Cocasse	1	3	15	15	6	5	12	18	4	25	2	2	108
Racé	4	33	7	8	3	6	6	4	5	15	5	3	99
Mièvre	3	9	1	7	7	5	12	9	6	9	6	13	87
Distingué	11	9	1	17	4	21	6	1	5	4	5	27	111
Vulgaire	4	4	32	2	6	0	9	7	3	10	10	7	94
Pour un homme	9	12	23	4	7	13	5	5	10	5	13	0	106
Pour une femme	9	3	9	8	4	2	6	8	1	6	23	33	112
Pour une petite nature	7	5	2	7	11	2	6	11	9	24	10	0	94
Total	78	112	106	132	83	84	78	71	79	123	84	115	1145

On traite ce tableau par une méthode d'analyse des données multidimensionnelles. Les principaux résultats sont rassemblés dans les tableaux et graphiques ci-dessous :

Fréquences

	Orly	Alezan	Corsaire	Directoire	Ducat	Fontenoy	Icare	Zodiaque	Pavois	Cocker	Escale	Hotesse	Total
Vieillot	0,09%	0,17%	1,22%	3,32%	1,57%	0,87%	0,79%	0,44%	0,79%	0,35%	0,00%	0,09%	9,69%
Nouveau-riche	1,75%	0,79%	0,09%	0,96%	0,87%	0,79%	0,09%	0,09%	1,75%	0,79%	0,61%	1,05%	9,61%
Sobre-élégant	0,79%	2,01%	0,09%	1,31%	0,61%	0,96%	0,52%	0,17%	0,61%	1,05%	0,26%	1,48%	9,87%
Cocasse	0,09%	0,26%	1,31%	1,31%	0,52%	0,44%	1,05%	1,57%	0,35%	2,18%	0,17%	0,17%	9,43%
Racé	0,35%	2,88%	0,61%	0,70%	0,26%	0,52%	0,52%	0,35%	0,44%	1,31%	0,44%	0,26%	8,65%
Mièvre	0,26%	0,79%	0,09%	0,61%	0,61%	0,44%	1,05%	0,79%	0,52%	0,79%	0,52%	1,14%	7,60%
Distingué	0,96%	0,79%	0,09%	1,48%	0,35%	1,83%	0,52%	0,09%	0,44%	0,35%	0,44%	2,36%	9,69%
Vulgaire	0,35%	0,35%	2,79%	0,17%	0,52%	0,00%	0,79%	0,61%	0,26%	0,87%	0,87%	0,61%	8,21%
Pour un homme	0,79%	1,05%	2,01%	0,35%	0,61%	1,14%	0,44%	0,44%	0,87%	0,44%	1,14%	0,00%	9,26%
Pour une femme	0,79%	0,26%	0,79%	0,70%	0,35%	0,17%	0,52%	0,70%	0,09%	0,52%	2,01%	2,88%	9,78%
Pour une petite nature	0,61%	0,44%	0,17%	0,61%	0,96%	0,17%	0,52%	0,96%	0,79%	2,10%	0,87%	0,00%	8,21%
Total	6,81%	9,78%	9,26%	11,53%	7,25%	7,34%	6,81%	6,20%	6,90%	10,74%	7,34%	10,04%	100 %

¹ Dans la version d'origine de cet exercice (Benzécri, 1980), le produit dont il est question est une cigarette, mais est-il encore "politiquement correct" de bâtir un exercice sur ce thème ?

Fréquences lignes

	Orly	Alezan	Corsaire	Directoire	Ducat	Fontenoy	Icare	Zodiaque	Pavois	Cocker	Escale	Hotesse	Total
Vieillot	0,90%	1,80%	12,61%	34,23%	16,22%	9,01%	8,11%	4,50%	8,11%	3,60%	0,00%	0,90%	100%
Nouveau-riche	18,18%	8,18%	0,91%	10,00%	9,09%	8,18%	0,91%	0,91%	18,18%	8,18%	6,36%	10,91%	100%
Sobre-élégant	7,96%	20,35%	0,88%	13,27%	6,19%	9,73%	5,31%	1,77%	6,19%	10,62%	2,65%	15,04%	100%
Cocasse	0,93%	2,78%	13,89%	13,89%	5,56%	4,63%	11,11%	16,67%	3,70%	23,15%	1,85%	1,85%	100%
Racé	4,04%	33,33%	7,07%	8,08%	3,03%	6,06%	6,06%	4,04%	5,05%	15,15%	5,05%	3,03%	100%
Mièvre	3,45%	10,34%	1,15%	8,05%	8,05%	5,75%	13,79%	10,34%	6,90%	10,34%	6,90%	14,94%	100%
Distingué	9,91%	8,11%	0,90%	15,32%	3,60%	18,92%	5,41%	0,90%	4,50%	3,60%	4,50%	24,32%	100%
Vulgaire	4,26%	4,26%	34,04%	2,13%	6,38%	0,00%	9,57%	7,45%	3,19%	10,64%	10,64%	7,45%	100%
Pour un homme	8,49%	11,32%	21,70%	3,77%	6,60%	12,26%	4,72%	4,72%	9,43%	4,72%	12,26%	0,00%	100%
Pour une femme	8,04%	2,68%	8,04%	7,14%	3,57%	1,79%	5,36%	7,14%	0,89%	5,36%	20,54%	29,46%	100%
Pour une petite nature	7,45%	5,32%	2,13%	7,45%	11,70%	2,13%	6,38%	11,70%	9,57%	25,53%	10,64%	0,00%	100%
Total	6,81%	9,78%	9,26%	11,53%	7,25%	7,34%	6,81%	6,20%	6,90%	10,74%	7,34%	10,04%	100%

Fréquences colonnes

	Orly	Alezan	Corsaire	Directoire	Ducat	Fontenoy	Icare	Zodiaque	Pavois	Cocker	Escale	Hotesse	Total
Vieillot	1,28%	1,79%	13,21%	28,79%	21,69%	11,90%	11,54%	7,04%	11,39%	3,25%	0,00%	0,87%	9,69%
Nouveau-riche	25,64%	8,04%	0,94%	8,33%	12,05%	10,71%	1,28%	1,41%	25,32%	7,32%	8,33%	10,43%	9,61%
Sobre-élégant	11,54%	20,54%	0,94%	11,36%	8,43%	13,10%	7,69%	2,82%	8,86%	9,76%	3,57%	14,78%	9,87%
Cocasse	1,28%	2,68%	14,15%	11,36%	7,23%	5,95%	15,38%	25,35%	5,06%	20,33%	2,38%	1,74%	9,43%
Racé	5,13%	29,46%	6,60%	6,06%	3,61%	7,14%	7,69%	5,63%	6,33%	12,20%	5,95%	2,61%	8,65%
Mièvre	3,85%	8,04%	0,94%	5,30%	8,43%	5,95%	15,38%	12,68%	7,59%	7,32%	7,14%	11,30%	7,60%
Distingué	14,10%	8,04%	0,94%	12,88%	4,82%	25,00%	7,69%	1,41%	6,33%	3,25%	5,95%	23,48%	9,69%
Vulgaire	5,13%	3,57%	30,19%	1,52%	7,23%	0,00%	11,54%	9,86%	3,80%	8,13%	11,90%	6,09%	8,21%
Pour un homme	11,54%	10,71%	21,70%	3,03%	8,43%	15,48%	6,41%	7,04%	12,66%	4,07%	15,48%	0,00%	9,26%
Pour une femme	11,54%	2,68%	8,49%	6,06%	4,82%	2,38%	7,69%	11,27%	1,27%	4,88%	27,38%	28,70%	9,78%
Pour une petite nature	8,97%	4,46%	1,89%	5,30%	13,25%	2,38%	7,69%	15,49%	11,39%	19,51%	11,90%	0,00%	8,21%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Taux de liaison

	Orly	Alezan	Corsaire	Directoire	Ducat	Fontenoy	Icare	Zodiaque	Pavois	Cocker	Escale	Hotesse
Vieillot	-0,868	-0,816	0,362	1,970	1,237	0,228	0,190	-0,274	0,175	-0,665	-1,000	-0,910
Nouveau-riche	1,669	-0,164	-0,902	-0,133	0,254	0,115	-0,867	-0,853	1,635	-0,238	-0,133	0,086
Sobre-élégant	0,169	1,081	-0,904	0,151	-0,145	0,327	-0,221	-0,715	-0,102	-0,011	-0,638	0,498
Cocasse	-0,864	-0,716	0,500	0,205	-0,234	-0,369	0,631	1,688	-0,463	1,155	-0,748	-0,816
Racé	-0,407	2,408	-0,236	-0,299	-0,582	-0,174	-0,110	-0,348	-0,268	0,410	-0,312	-0,698
Mièvre	-0,494	0,058	-0,876	-0,302	0,110	-0,217	1,025	0,668	0,000	-0,037	-0,060	0,488
Distingué	0,455	-0,171	-0,903	0,328	-0,503	1,579	-0,207	-0,855	-0,347	-0,665	-0,386	1,422
Vulgaire	-0,375	-0,565	2,677	-0,815	-0,119	-1,000	0,405	0,201	-0,537	-0,010	0,450	-0,259
Pour un homme	0,246	0,157	1,344	-0,673	-0,089	0,672	-0,308	-0,239	0,367	-0,561	0,672	-1,000
Pour une femme	0,180	-0,726	-0,132	-0,380	-0,507	-0,757	-0,214	0,152	-0,871	-0,501	1,799	1,934
Pour une petite nature	0,093	-0,456	-0,770	-0,354	0,614	-0,710	-0,063	0,887	0,388	1,377	0,450	-1,000

Valeurs propres

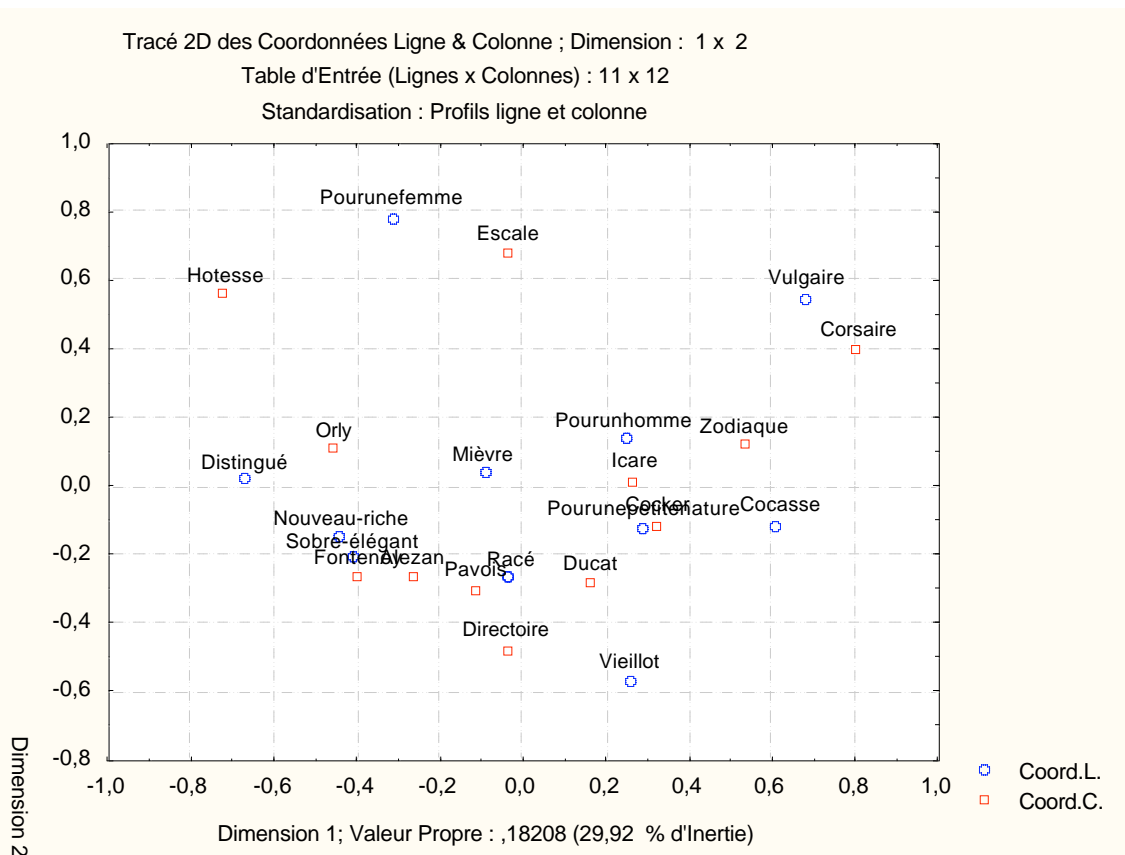
Nombre de Dims.	Valeurs Propres et Inertie de toutes les Dimensions				
	Inertie Totale = ,60855 Chi ² = 696,79 dl = 110 p = 0,0000				
	ValSing.	ValProp.	%age Inertie	%age Cumulé	Chi ²
1	0,4267	0,1821	29,9210	29,9210	208,4852
2	0,3646	0,1330	21,8484	51,7694	152,2368
3	0,3230	0,1043	17,1421	68,9115	119,4437
4	0,2849	0,0812	13,3400	82,2515	92,9514
5	0,2532	0,0641	10,5317	92,7832	73,3831
6	0,1456	0,0212	3,4817	96,2649	24,2602
7	0,1053	0,0111	1,8232	98,0881	12,7041
8	0,0943	0,0089	1,4610	99,5491	10,1800
9	0,0514	0,0026	0,4338	99,9829	3,0223
10	0,0102	0,0001	0,0171	100,0000	0,1192

Résultats relatifs aux lignes

NomLigne	Coordonnées Ligne et Contributions à l'Inertie												
	Standardisation : Profils ligne et colonne												
	N°	Coord. Dim.1	Coord. Dim.2	Coord. Dim.3	Masse	Qualité	Inertie Relative	Inertie Dim.1	Cos ² Dim.1	Inertie Dim.2	Cos ² Dim.2	Inertie Dim.3	Cos ² Dim.3
Vieillot	1	0,266	-0,572	0,662	0,097	0,926	0,144	0,038	0,078	0,238	0,362	0,408	0,486
Nouveau-riche	2	-0,436	-0,157	-0,007	0,096	0,380	0,089	0,100	0,337	0,018	0,043	0,000	0,000
Sobre-élégant	3	-0,409	-0,216	-0,194	0,099	0,853	0,048	0,090	0,567	0,035	0,159	0,035	0,127
Cocasse	4	0,614	-0,118	0,025	0,094	0,640	0,095	0,196	0,616	0,010	0,023	0,001	0,001
Racé	5	-0,031	-0,270	-0,704	0,086	0,805	0,101	0,000	0,001	0,047	0,103	0,411	0,701
Mièvre	6	-0,091	0,033	-0,044	0,076	0,050	0,028	0,003	0,036	0,001	0,005	0,001	0,009
Distingué	7	-0,667	0,014	0,262	0,097	0,824	0,099	0,237	0,714	0,000	0,000	0,064	0,110
Vulgaire	8	0,686	0,544	0,026	0,082	0,843	0,123	0,212	0,517	0,183	0,325	0,001	0,001
Pour un homme	9	0,249	0,131	-0,108	0,093	0,203	0,068	0,031	0,139	0,012	0,038	0,010	0,026
Pour une femme	10	-0,313	0,779	0,183	0,098	0,891	0,133	0,053	0,118	0,447	0,732	0,031	0,040
Pour une petite nature	11	0,292	-0,129	-0,218	0,082	0,281	0,072	0,039	0,160	0,010	0,031	0,037	0,089

Résultats relatifs aux colonnes

Nom Col.	Coordonnées Colonne et Contributions à l'Inertie												
	Standardisation : Profils ligne et colonne												
	N°	Coord. Dim.1	Coord. Dim.2	Coord. Dim.3	Masse	Qualité	Inertie Relative	Inertie Dim.1	Cos ² Dim.1	Inertie Dim.2	Cos ² Dim.2	Inertie Dim.3	Cos ² Dim.3
Orly	1	-0,452	0,101	-0,080	0,068	0,453	0,055	0,076	0,419	0,005	0,021	0,004	0,013
Alezan	2	-0,262	-0,267	-0,723	0,098	0,796	0,134	0,037	0,083	0,053	0,086	0,490	0,627
Corsaire	3	0,808	0,393	0,125	0,093	0,720	0,174	0,332	0,572	0,107	0,135	0,014	0,014
Directoire	4	-0,034	-0,482	0,484	0,115	0,883	0,100	0,001	0,002	0,202	0,439	0,259	0,442
Ducat	5	0,160	-0,288	0,261	0,072	0,645	0,033	0,010	0,094	0,045	0,302	0,047	0,248
Fontenoy	6	-0,395	-0,270	0,152	0,073	0,505	0,060	0,063	0,312	0,040	0,146	0,016	0,046
Icare	7	0,267	0,010	0,055	0,068	0,324	0,026	0,027	0,310	0,000	0,000	0,002	0,013
Zodiaque	8	0,537	0,114	-0,038	0,062	0,517	0,060	0,098	0,492	0,006	0,022	0,001	0,003
Pavois	9	-0,116	-0,305	-0,027	0,069	0,254	0,048	0,005	0,032	0,048	0,220	0,000	0,002
Cocker	10	0,322	-0,117	-0,339	0,107	0,526	0,078	0,061	0,235	0,011	0,031	0,118	0,260
Escale	11	-0,035	0,676	-0,080	0,073	0,767	0,073	0,001	0,002	0,252	0,754	0,005	0,011
Hotesse	12	-0,724	0,552	0,214	0,100	0,901	0,160	0,289	0,540	0,230	0,314	0,044	0,047



1) Quelle est la méthode d'analyse utilisée ici ? Pourquoi utilise-t-on cette méthode ?

La méthode utilisée ici est une analyse factorielle des correspondances. Cette méthode permet d'étudier des données présentées sous forme de tableau de contingence.

2) a) Dans le tableau des fréquences lignes, on lit 20,35% à l'intersection de la ligne "sobre-élégant" et de la colonne "alezan". Comment a été obtenue cette valeur ? Que signifie-t-elle ?

Les fréquences lignes sont obtenues en divisant l'effectif d'une cellule par l'effectif de la ligne correspondante. Ainsi, on a ici : $20,35\% = 23/113$.

b) Comparer qualitativement le profil de la marque "Corsaire" au profil moyen.

Ce sont les profils colonnes qui doivent être utilisés ici. Alors que le profil colonne moyen est relativement uniforme (fréquences de 7,60% à 9,87%), le profil de la modalité "Corsaire" fait apparaître des fréquences très faibles (4 fréquences à moins de 1%) et des fréquences élevées, avec notamment "Vulgaire", à plus de 30%. Ce profil est donc très éloigné du profil moyen.

3) a) La première valeur indiquée dans le tableau des taux de liaison est -0,868. Indiquer le calcul qui a permis d'obtenir cette valeur.

Le taux de liaison est obtenu à partir de la fréquence (0,09%) et des fréquences marginales de la ligne (9,69%) et de la colonne (6,81%) à l'aide du calcul suivant :

$$t = \frac{0,0009 - 0,0969 \times 0,0681}{0,0969 \times 0,0681} = -0,863$$

b) Au vu du tableau des taux de liaison, indiquer une modalité ligne et une modalité colonne qui "s'attirent". Indiquer une modalité ligne et une modalité colonne qui "se repoussent".

Il existe un taux de liaison positif élevé entre "Orly" et "Nouveau-riche", ce qui montre que ces deux modalités s'attirent. Au contraire, le taux de liaison négatif entre "Orly" et "Vieillot" montre que ces modalités se repoussent.

b) Quel est le taux de liaison maximum du tableau ? Comment pourrait-on exprimer d'une autre façon le résultat correspondant à ce taux de liaison.

Le taux maximum (2,677) est observé pour le couple de modalités (Vulgaire, Corsaire). Il signifie que la fréquence de ce couple de modalités dépasse de 267% la fréquence attendue sous hypothèse d'indépendance des lignes et colonnes.

4) Le coefficient Phi-2 de ce tableau vaut 0,60855. Relier ce résultat aux valeurs propres d'une part, au khi-2 du tableau d'autre part.

On sait que ce coefficient est égal à la somme des valeurs propres. Il est relié au khi-2 par la relation $\text{Phi}^2 = \text{Khi}^2/N$. On vérifie, sur les tableaux ci-dessus, que : $0,60855 = 696,79/1145$.

5) Au vu du tableau des valeurs propres, combien de variables factorielles faudrait-il étudier dans le cadre d'une étude complète des données observées ?

L'AFC fournit ici 10 valeurs propres, soit une moyenne de 10% d'inertie associée à chaque axe. D'autre part, on observe une brusque diminution de l'inertie associée à partir du 6^e axe. Autrement dit, la méthode de la moyenne, aussi bien que la méthode du coude, nous conduiraient ici à étudier les 5 premiers axes.

6) a) En utilisant le tableau des résultats relatifs aux lignes, calculer la distance du point représentant la modalité "Nouveau-riche" à l'origine du repère, dans l'espace de dimension 3 défini par les 3 premiers axes factoriels.

On a : $d^2 = (-0,436)^2 + (-0,157)^2 + (-0,007)^2 = 0,2148$ d'où : $d=0,4635$.

b) Rappeler la définition de la notion de qualité de représentation d'un point.

La qualité de représentation d'un point M est le rapport $\frac{OH^2}{OM^2}$, où H désigne la projection du point M dans le sous-espace retenu pour la représentation.

c) En utilisant la qualité de représentation indiquée dans le tableau, calculer la distance du point représentant la modalité "Nouveau-riche" à l'origine du repère dans l'espace de dimension 10 défini par l'ensemble des axes factoriels.

La qualité de représentation de "Nouveau-riche" dans l'espace des trois premiers facteurs est 0,38. Le carré de la distance dans l'espace de dimension 10 est donc : $D^2 = \frac{0,2148}{0,38} = 0,5652$ d'où $D=0,7518$.

d) Mêmes questions pour le point représentant la modalité "Sobre-élégant".

On a : $d^2 = (-0,409)^2 + (-0,216)^2 + (-0,194)^2 = 0,2516$ d'où : $d=0,5016$.

La qualité de représentation de "Sobre-élégant" dans l'espace des trois premiers facteurs est 0,853. Le carré de la distance dans l'espace de dimension 10 est donc : $D^2 = \frac{0,2516}{0,853} = 0,2950$ d'où $D=0,5431$.

e) Dans la représentations en 3 dimensions, "Nouveau-riche" est plus près du profil moyen que "Sobre-élégant". En est-il de même dans la réalité, c'est-à-dire dans l'espace de dimension 10 ?

On constate que, dans l'espace de dimension 3, "Nouveau-riche" est plus près de l'origine que "Sobre-élégant", alors que dans l'espace de dimension 10, c'est l'inverse qui se produit. Mais, la qualité de représentation de "Nouveau-riche" dans l'espace de dimension 3 est nettement moins bonne que celle de "Sobre-élégant".

7) Etude de la première variable factorielle

a) On considère le nuage des attributs. Quels sont les individus dont la contribution est supérieure à la moyenne ? Pour chacun d'eux, préciser le signe de la coordonnée correspondante. Comment peut-on interpréter cet axe en termes d'opposition entre attributs.

Le nombre d'attributs est de 11. Le seuil envisagé correspond donc à 9,09% de contribution à l'inertie du premier axe.

-	+
Distingué (23,7%)	Vulgaire (21,2%)
Nouveau-riche (10,0%)	Cocasse (19,6%)

On voit que cet axe oppose essentiellement les attributs antinomiques distingué / vulgaire.

b) Même question pour le nuage des marques envisagées.

Le seuil est ici de 1/12, c'est-à-dire 8,33%.

-	+
Hôtesse (28,9%)	Corsaire (33,2%)
	Zodiaque (19,8%)

Les marques qui s'opposent ici sont d'une part "Hôtesse", fortement liée aux attributs "distingué" et "pour une femme" (cf. le tableau des taux de liaison) à des marques telles que "Corsaire" ou "Zodiaque", liées l'une à "vulgaire", l'autre à "cocasse".

8) Mener une étude analogue pour la deuxième variable.

On considère la deuxième variable et le nuage des attributs :

-	+
Vieillot (23,8%)	Pour une femme (44,7%)
	Vulgaire (18,3%)

S'agissant du nuage des marques envisagées :

-	+
Directoire (20,2%)	Escale (25,2%)
	Hôtesse (23%)
	Corsaire (10,7%)

Une part importante de l'inertie de cet axe est associée à l'attribut "Pour une femme", qui est associé à des marques telles que "Escale" ou "Hôtesse". On y retrouve, dans une moindre mesure, "Corsaire", associé à "Vulgaire". A ces attributs s'oppose "Vieillot", lié à "Directoire"

9) Faire une synthèse des deux études précédentes en décrivant les résultats obtenus dans le premier plan factoriel.

L'étude des deux premiers axes (et du graphique), permet déjà de cerner trois zones :

- Les marques "Zodiaque", "Corsaire" liées à des attributs tels que "vulgaire" et "cocasse" ;
- Une forte liaison entre "Escale" et "Pour une femme", à laquelle s'ajoutent des marques telles que "Hôtesse" ou "Orly", qui sont également liées à "distingué";
- Des marques telles que "Directoire", et dans une moindre mesure, "Pavois" ou "Ducat", liées à l'attribut "vieillot".

10) A votre avis, quelle est la marque qui a finalement été retenue par l'entreprise ?

Compte tenu du cahier des charges posé au départ (produit qui doit paraître de qualité supérieure, prestigieux, luxueux, convenant à un public masculin, connaisseur, raffiné, distingué, de niveau socio-économique élevé...), aucune des marques citées dans l'analyse précédente ne convient. La zone la plus intéressante du graphique est ici celle contenant les points "distingué", "nouveau-riche", "sobre élégant", "Fontenoy", "Alezan". Mais "Alezan" n'est pas assez lié à "distingué", et c'est finalement la marque "Fontenoy" qui a été retenue.