

Examen d'Analyse de données

Session 1 – durée 3 heures

Les documents sont interdits. Les calculatrices sont autorisées.

Exercice 1: - Etude d'un tableau à l'aide d'une ACP (7 points)

Soit le tableau $T = D$ décrivant pour les 6 individus A,B,...,F les valeurs prises par les variables quantitatives x, y, z

A	2	1	0
B	1	1	0
C	0	0	1
D	2	0	1
E	1	0	1
F	0	1	0

$x \quad y \quad z$

1/ Construction de la matrice des corrélations

- a. Calculer la moyenne, la variance et les écart-type des variables.
- b. Déterminer le tableau centré réduit.

c. En déduire que la matrice des corrélations est $M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix}$

- d. Interpréter cette matrice.

2/ Ajustement du nuage des individus

- a. Déterminer les valeurs propres de M.
- b. Interpréter ces valeurs propres en terme d'inertie. En déduire le % d'inertie projeté sur chaque axe.
- c. Déterminer un vecteur propre unitaire associé à λ_1 et λ_2 .
- d. Quelle est la signification des axes définis par ces vecteurs?

3/ Représentation des individus

- a. Calculer les composantes principales F_1 et F_2 .
- b. Quelle est la signification de F_1 et F_2 ?
- c. Placer les individus dans le plan (u_1, u_2) (prendre pour unité 4cm)

4/ Représentation des variables

- a. Calculer les facteurs principaux G_1 et G_2 .
- b. Quelle est la signification de G_1 et G_2 ?
- c. Placer les variables dans le plan (v_1, v_2) et tracer le cercle des corrélations.

Exercice 2 : Analyse de documents (7 points)

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...

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 :

1. 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%

2. Fréquences lignes

	Orly	Alezan	Corsaire	Directoire	Ducat	Fonteno y	Icare	Zodiaqu e	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%

3. 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%

4. Valeurs propres

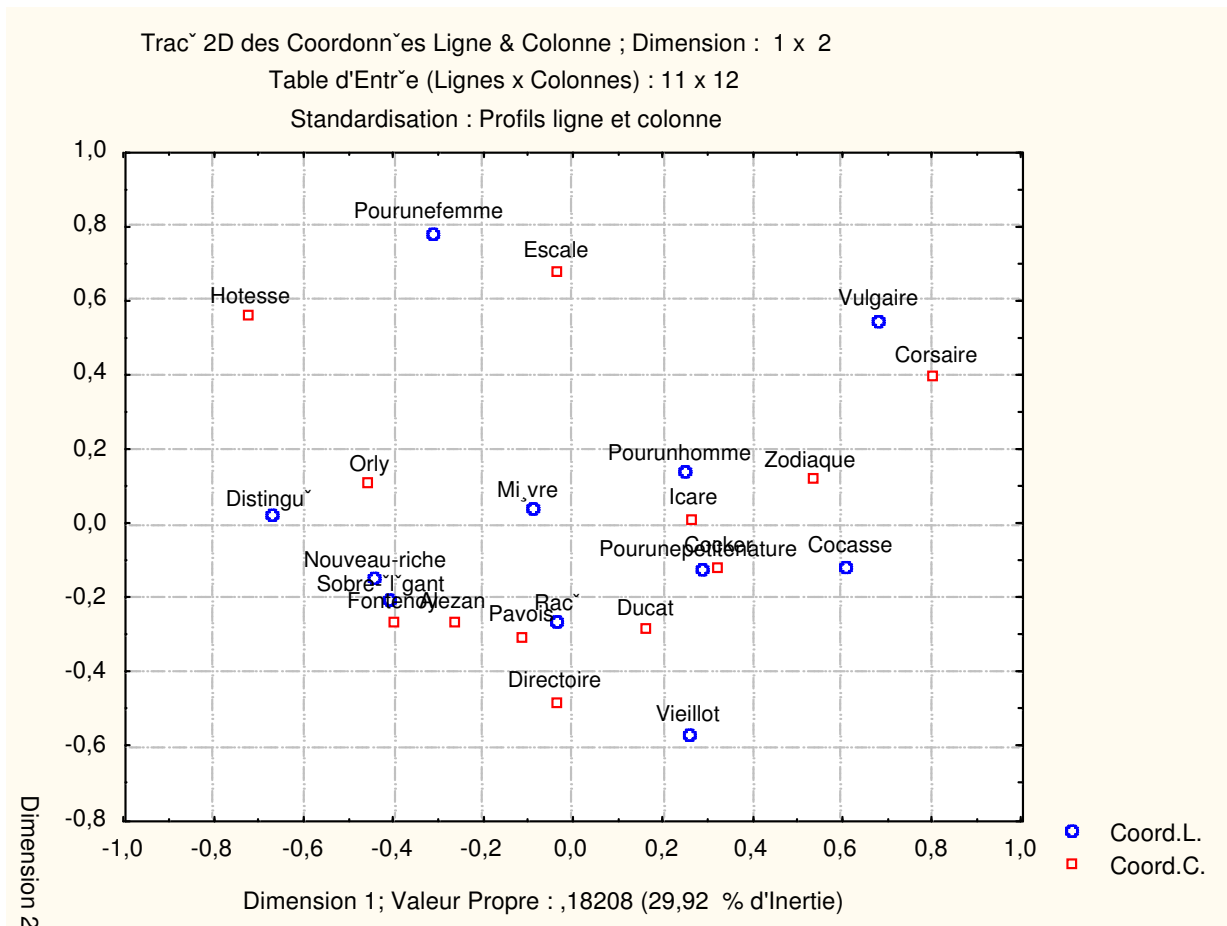
Nombre de Dims.	Valeurs Propres et Inertie de toutes les Dimensions				
	Inertie Totale = ,60855 Chi ² = 696,79 dl = 110 p = 0,0000				
	ValProp.	%age Inertie	%age Cumulé	Chi ²	
1	0,1821	29,9210	29,9210	208,4852	
2	0,1330	21,8484	51,7694	152,2368	
3	0,1043	17,1421	68,9115	119,4437	
4	0,0812	13,3400	82,2515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

5. Résultats relatifs aux lignes

NomLigne	Coordonnées Ligne et Contributions à l'Inertie												
	Standardisation : Profils ligne et colonne												
	N°	Coord F 1	Coord F.2	Coord F 3	Masse		Inertie	Ctr F1	Cos ² F1	Ctr F 2	Cos ² F 2	Ctr F.3	Cos ² F3
Vieillot	1	0,266	-0,572	0,662	0,097		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		0,089	0,100	0,337	0,018	0,043	0,000	0,000
Sobre-élégant	3	-0,409	-0,216	-0,194	0,099		0,048	0,090	0,567	0,035	0,159	0,035	0,127
Cocasse	4	0,614	-0,118	0,025	0,094		0,095	0,196	0,616	0,010	0,023	0,001	0,001
Racé	5	-0,031	-0,270	-0,704	0,086		0,101	0,000	0,001	0,047	0,103	0,411	0,701
Mièvre	6	-0,091	0,033	-0,044	0,076		0,028	0,003	0,036	0,001	0,005	0,001	0,009
Distingué	7	-0,667	0,014	0,262	0,097		0,099	0,237	0,714	0,000	0,000	0,064	0,110
Vulgaire	8	0,686	0,544	0,026	0,082		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		0,068	0,031	0,139	0,012	0,038	0,010	0,026
Pour une femme	10	-0,313	0,779	0,183	0,098		0,133	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		0,072	0,039	0,160	0,010	0,031	0,037	0,089

6. Résultats relatifs aux colonnes

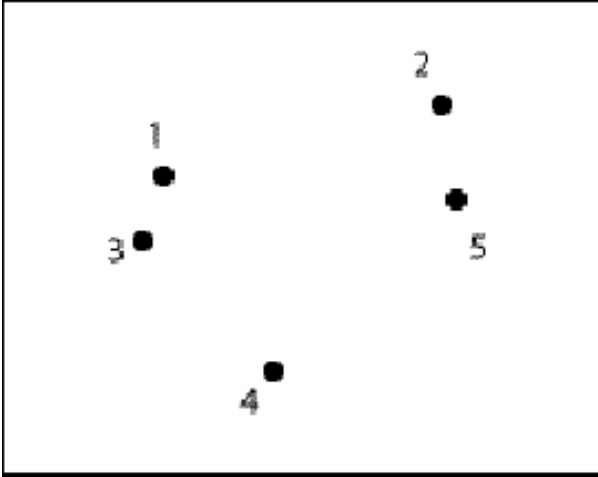
Nom Col.	Coordonnées Colonne et Contributions à l'Inertie												
	Standardisation : Profils ligne et colonne												
	N°	Coord. F 1	Coord. F.2	Coord. F 3	Masse		Inertie	Ctr F1	Cos ² F1	Ctr F 2	Cos ² F 2	Ctr F.3	Cos ² F3
Orly	1	-0,452	0,101	-0,080	0,068		0,055	0,076	0,419	0,005	0,021	0,004	0,013
Alezan	2	-0,262	-0,267	-0,723	0,098		0,134	0,037	0,083	0,053	0,086	0,490	0,627
Corsaire	3	0,808	0,393	0,125	0,093		0,174	0,332	0,572	0,107	0,135	0,014	0,014
Directoire	4	-0,034	-0,482	0,484	0,115		0,100	0,001	0,002	0,202	0,439	0,259	0,442
Ducat	5	0,160	-0,288	0,261	0,072		0,033	0,010	0,094	0,045	0,302	0,047	0,248
Fontenoy	6	-0,395	-0,270	0,152	0,073		0,060	0,063	0,312	0,040	0,146	0,016	0,046
Icare	7	0,267	0,010	0,055	0,068		0,026	0,027	0,310	0,000	0,000	0,002	0,013
Zodiaque	8	0,537	0,114	-0,038	0,062		0,060	0,098	0,492	0,006	0,022	0,001	0,003
Pavois	9	-0,116	-0,305	-0,027	0,069		0,048	0,005	0,032	0,048	0,220	0,000	0,002
Cocker	10	0,322	-0,117	-0,339	0,107		0,078	0,061	0,235	0,011	0,031	0,118	0,260
Escale	11	-0,035	0,676	-0,080	0,073		0,073	0,001	0,002	0,252	0,754	0,005	0,011
Hotesse	12	-0,724	0,552	0,214	0,100		0,160	0,289	0,540	0,230	0,314	0,044	0,047



- 1) a) Quelle est la méthode d'analyse utilisée ici ? Pourquoi utilise-t-on cette méthode ?
b) Quelle information nous apporte le test du Khi deux présenté en 4.
- 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 ?
b) Comparer qualitativement le profil de la marque "Corsaire" au profil moyen.
- 3) a) Au vu du tableau des valeurs propres, combien de variables factorielles faudrait-il étudier dans le cadre d'une étude complète des données observées ?
b) Quel résultat manque-t-il pour aider au choix des axes ?
c) Dans cette étude, on se limite aux trois premiers axes. Quel pourcentage d'inertie est expliqué par ces trois axes?
- 4) a) Rappeler la définition de la notion de qualité de représentation d'un point.
b) Les points sont-ils bien représentés dans le plan F1-F2?
- 5) 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.
b) Même question pour le nuage des marques envisagées.
- 6) Mener une étude analogue pour la deuxième variable.
- 7) Faire une synthèse des deux études précédentes en décrivant les résultats obtenus dans le premier plan factoriel.
- 8) A votre avis, quelle est la marque qui a finalement été retenue par l'entreprise ?

Exercice 3 : Classification (6 points)

Cinq objets sont représentés en grandeur nature. La distance utilisée entre les objets est la distance euclidienne (mesurée en mm au double-décimètre).



I. Construction d'une CAH

1/ Construire le tableau de distance. Les distances seront exprimées en mm (arrondir au mm près).

2/ Effectuer une CAH sur ce tableau en utilisant comme critère d'agrégation celui du "saut minimal".

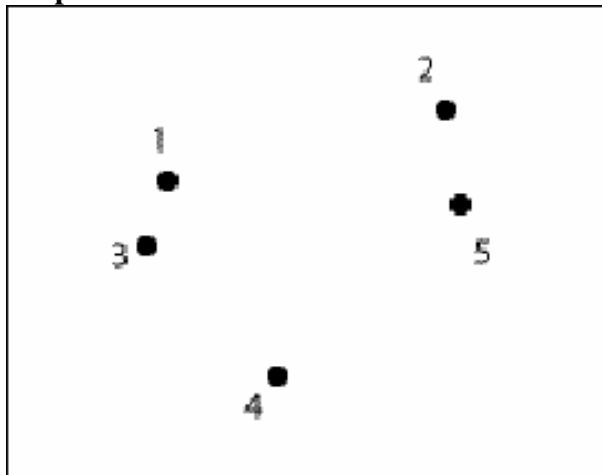
- Construire tous les tableaux intermédiaires
- Construire le dendrogramme résultant.
- Construire le tableau des distances ultramétriques correspondant.
- Quelle est la partition qui vous semble préférable. En déduire les groupes.

II Classification à l'aide de centres mobiles

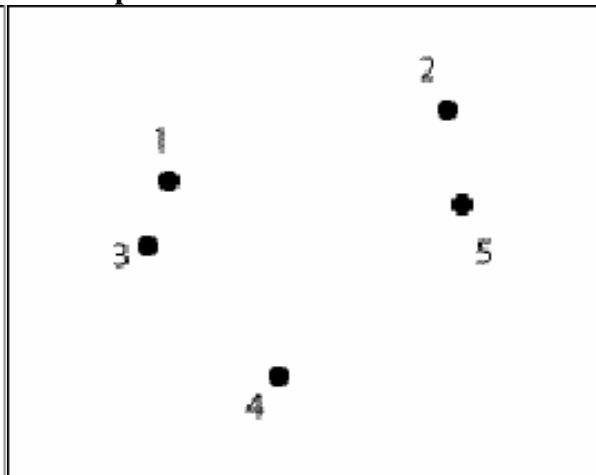
Construire une partition en deux classes des objets à l'aide de la méthode des centres mobiles.

On choisira comme premiers centres mobiles les objets 1 et 3. Décrire sur les différentes figures les étapes successives.

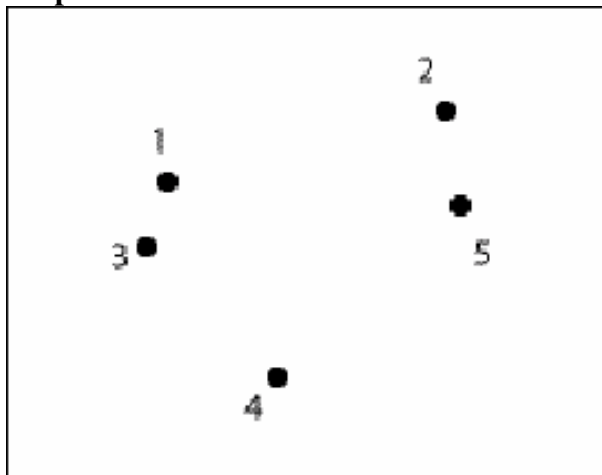
Étape 1



Étape 2



Étape 3



Étape 4

