

Université Panthéon-Assas

SESSION : Septembre 2018.
 ANNEE D'ETUDE : Master 1 Economie Managériale et Industrielle
 MATIERE : ANALYSE DE DONNEES (4001) Enseignant : Mr FAKHFAKH
calculatrice autorisée.

Exercice N°1

Une enquête a été effectuée auprès de 56 marques d'eau. Nous disposons ainsi des principales caractéristiques de chaque marque en terme de teneur (en mg/litre) des composants suivants :

CA : calcium. MG : magnésium. NA : sodium. K : potassium.
 SUL : sulfates. NO3 : nitrates. HCO3 : carbonates. CL : chlorures.

Un premier résumé de ces caractéristiques est donné dans le tableau suivant :

	Moyenne	Ecart Type	Matrice de corrélation							
			CA	MG	NA	K	SUL	NO3	HCO3	CL
CA	102,449	120,656	1,000	0,717	0,092	0,037	0,918	-0,015	0,109	0,264
MG	25,052	27,040	0,717	1,000	0,551	0,070	0,652	-0,108	0,561	0,420
NA	83,896	184,108	0,092	0,551	1,000	-0,009	0,079	-0,110	0,831	0,533
K	48,156	292,690	0,037	0,070	-0,009	1,000	0,099	0,754	-0,038	-0,028
SUL	143,525	333,325	0,918	0,652	0,079	0,099	1,000	-0,029	-0,068	0,342
NO3	4,992	10,356	-0,015	-0,108	-0,110	0,754	-0,029	1,000	-0,095	-0,101
HCO3	407,956	567,256	0,109	0,561	0,831	-0,038	-0,068	-0,095	1,000	0,080
CL	46,507	137,193	0,264	0,420	0,533	-0,028	0,342	-0,101	0,080	1,000

Les deux premiers vecteurs propres sont les suivants :

	CA	MG	NA	K	SUL	NO3	HCO3	CL
Prin1	0,431	0,527	0,384	0,002	0,409	-0,085	0,316	0,337
Prin2	0,332	0,036	-0,362	0,485	0,381	0,479	-0,380	-0,058

Matrice de corrélation des variables avec les 4 premières composantes principales

	Prin1	Prin2	Prin3	Prin4
CA	0,762	0,457	-0,350	-0,209
MG	0,931	0,050	0,034	-0,176
NA	0,679	-0,497	0,482	0,134
K	0,004	0,668	0,661	0,041
SUL	0,722	0,524	-0,393	-0,030
NO3	-0,150	0,659	0,650	0,035
HCO3	0,558	-0,522	0,510	-0,381
CL	0,595	-0,080	0,002	0,793

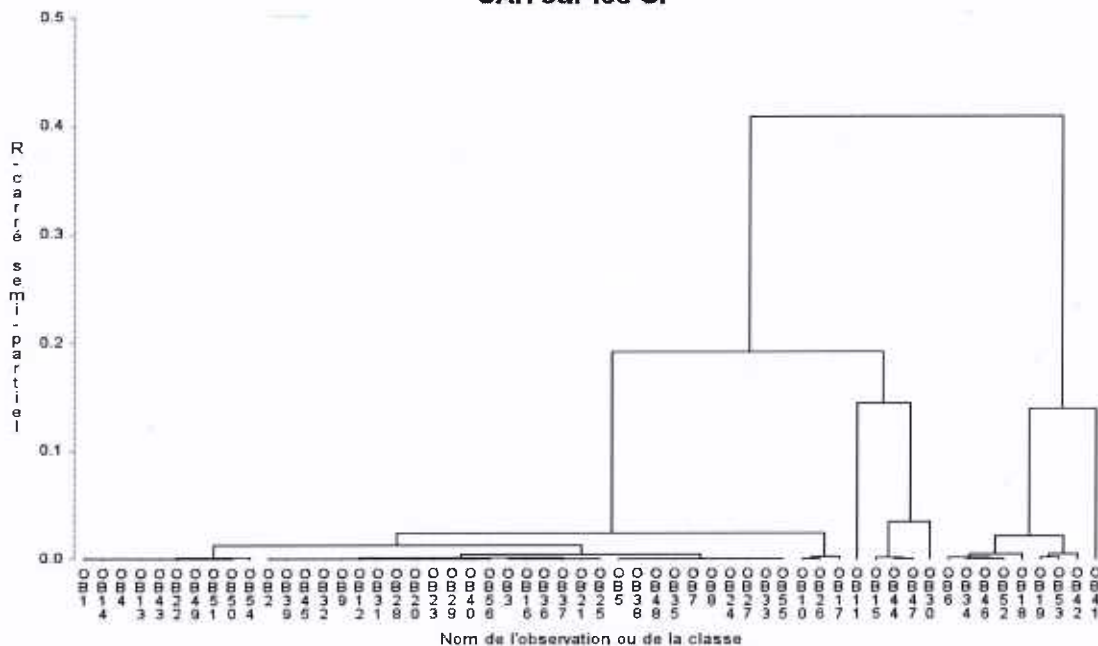
- 1) Examiner les moyennes et les écart-types. Quels sont les critères qui paraissent les plus importants.
- 2) a. A partir de la matrice de corrélation, représenter le dendrogramme des variables. Interpréter.
 b. Est-il indispensable d'effectuer l'analyse sur les variables centrées et réduites.
- 3) Les quatre premières valeurs propres obtenues dans cet exemple, classées par ordre décroissant, sont les suivantes : (3.117), (1.893), (1.630) et (0.870).
 Déterminer la valeur de l'inertie totale et donner la contribution de chaque axe à l'inertie totale. Interpréter.
- 4) a. Quelles sont les variables qui contribuent le plus à la définition de la première composante principale ?
 Que traduit cette première composante principale.
 b. Même question que (a.) pour les axes 2 et 3.
 c. Quel est le nombre d'axes que l'on peut retenir? Justifier votre réponse.
- 5) Représenter le cercle de corrélation (pour les deux premiers axes). L'interpréter.

6) Nous avons décidé d'établir une classification ascendante hiérarchique sur les individus en fonction des mêmes variables.

a. Un résumé sur les derniers nœuds de la CAH est donné dans le tableau suivant, ainsi que l'arbre correspondant à cette classification. Donner le nombre de classes à retenir. Justifier votre réponse.

Classes	Ainé	Bénjamin	Perte Inertie
8	CL14	CL10	.964
7	CL9	CL11	.942
6	CL8	CL12	.918
5	CL17	OB30	.884
4	CL7	OB41	.745
3	OB11	CL5	.601
2	CL6	CL3	.409
1	CL2	CL4	.000

CAH sur les CP



b. Nous avons décidé de retenir une typologie à 5 classes dont les caractéristiques sont données dans le tableau suivant :

Cluster	Nb d'indiv	Moyennes des variables par cluster							
		CA	MG	NA	K	SUL	NO3	HCO3	CL
1	6,0	108,9	55,9	351,8	38,9	64,2	3,3	1392,4	74,4
2	1,0	130,0	31,0	0,0	2195,0	387,0	63,0	10,2	1,4
3	4,0	471,3	78,0	168,0	7,6	1238,8	1,2	281,3	250,2
4	1,0	99,0	88,1	968,0	103,0	18,0	1,0	3380,5	88,0
5	44,0	67,5	14,5	21,5	3,1	52,1	4,3	226,7	24,3

Interpréter chacune des classes. Discuter de la pertinence de cette analyse à la lumière de ces résultats. Comment peut-on améliorer la qualité de cette analyse.

Exercice N°2

Nous cherchons dans cet exemple à étudier les conditions favorisant l'efficacité des changements technologiques et organisationnels sur la performance des entreprises. Nous nous intéressons plus particulièrement aux effets de l'implication des représentants des salariés (à travers les syndicats). Nous disposons pour cela d'un échantillon de 1727 entreprises observées en 2004 (enquête REPONSE). Les caractéristiques des entreprises que nous étudions sont : leur performance (Perf+ : supérieure à celle des concurrents, 'Perf-' inférieure, ou 'perf=' identique), le type d'implication des salariés dans la préparation des changements technologiques (Consult : consulte ses salariés, Inform : informe ses salariés, Negotiate :

négoce avec les salariés et Novoice pour aucune des trois précédentes). Nous prenons en compte aussi la présence des 5 principaux syndicats dans l'entreprise (que nous noterons de S1 à S5). S1++ désigne le cas où le syndicat S1 est présent, et S1—lorsque le syndicat S1 n'est pas présent.

I- Nous avons procédé à une ACM sur le nuage des profils des colonnes. Les 8 premières valeurs propres prennent les valeurs suivantes: 0.382 0.150 0.145 0.143 0.141 0.132 0.095 0.088

Le tableau suivant donne les coordonnées et les cosinus carrés sur les quatre premiers axes principaux.

	Coordonnées				Cos Carrés			
	Dim1	Dim2	Dim3	Dim4	Dim1	Dim2	Dim3	Dim4
Perf++	-0,066	1,168	-0,388	0,320	0,002	0,496	0,055	0,037
perf--	-0,006	0,244	1,497	0,085	0,000	0,013	0,504	0,002
perf—	0,034	-0,649	-0,312	-0,184	0,001	0,513	0,118	0,041
CONSULT	-0,069	-0,501	1,501	0,482	0,001	0,051	0,455	0,047
INFORM	0,141	0,407	0,034	-1,780	0,006	0,049	0,000	0,934
NEGOTIATE	0,311	-0,882	-0,581	0,449	0,034	0,275	0,119	0,071
NOVOICE	-0,296	0,646	-0,315	0,603	0,046	0,219	0,052	0,190
S1++	0,714	0,054	0,031	-0,048	0,516	0,003	0,001	0,002
S1--	-0,722	-0,054	-0,032	0,048	0,516	0,003	0,001	0,002
S2++	1,148	0,120	-0,137	0,016	0,463	0,005	0,007	0,000
S2--	-0,403	-0,042	0,048	-0,006	0,463	0,005	0,007	0,000
S3++	1,013	0,094	0,057	0,107	0,561	0,005	0,002	0,006
S3--	-0,544	-0,052	-0,032	-0,059	0,561	0,005	0,002	0,006
S4++	0,745	-0,029	-0,008	0,030	0,568	0,001	0,000	0,001
S4--	-0,763	0,030	0,008	-0,031	0,568	0,001	0,000	0,001
S5++	0,877	0,059	0,069	0,074	0,504	0,002	0,003	0,004
S5--	-0,575	-0,039	-0,045	-0,049	0,504	0,002	0,003	0,004

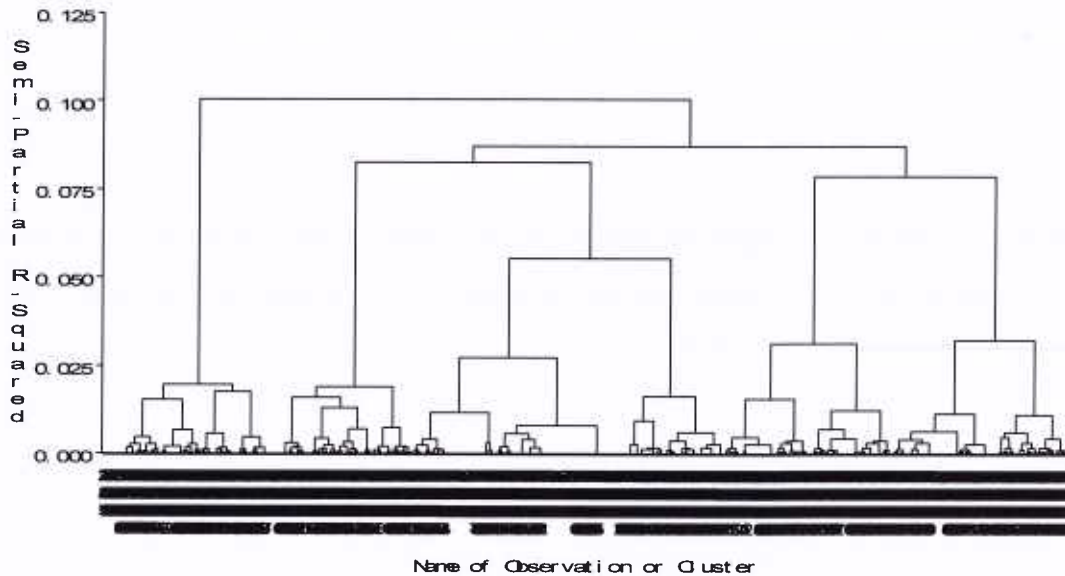
1. Donner la valeur de l'inertie totale. En déduire la contribution de chacun des 8 premiers axes à l'inertie.
2. Quel est le nombre d'axes à retenir. Justifier votre réponse.
3. Représenter le premier plan factoriel (axes 1 et 2). Interpréter.

II. Nous avons ensuite effectué une analyse sur le nuage des profils des lignes.

1- Quelle est la valeur d'inertie totale dans ce cas et quelles seraient les huit premières valeurs propres ?
 2- Vu le nombre important d'individus, nous allons nous concentrer sur les résultats de la classification ascendante hiérarchique. L'arborescence correspondante à cette classification est donnée dans la figure suivante.

- a- Sur quelles variables cette classification est-elle réalisée.
- b- Quel est le nombre de classes que vous retiendriez ?

CAH sur les individus



3- Nous cherchons à donner une interprétation à chacune des classes. Pour cela, nous donnons dans le tableau suivant l'ensemble des caractéristiques de chacune de ces classes.

	Ensemble	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6
performance							
Perf++	26.69	47.32	66.42	13.31	21.28	12.28	10.92
perf--	18.36	10.73	4.85	50.00	9.91	2.01	40.80
perf==	54.95	41.95	28.73	36.69	68.80	85.71	48.28
Implication des salariés							
CONSULT	16.79	0	53.73	0	0	0	83.91
INFORM	22.76	82.44	0	64.79	0	1.25	0
NEGOTIATE	26.11	0	10.82	11.54	73.18	31.58	3.45
NOVOICE	34.34	17.56	35.45	23.67	26.82	67.17	12.64
Syndicats							
S1++	50.32	85.85	16.04	34.62	89.80	21.55	79.89
S2++	26.00	59.51	4.48	6.51	58.31	4.51	43.10
S3++	35.38	77.07	6.34	9.47	74.64	10.53	60.92
S4++	50.61	85.85	6.72	24.85	93.29	32.58	83.91
S5++	39.61	82.44	6.72	10.65	65.89	27.07	72.99
Nb Firmes	1727	205	268	338	343	399	174

Interpréter attentivement ce tableau : donner une interprétation à chacune de ces classes ainsi qu'une interprétation générale.

Questions de cours

1. Peut-on obtenir une (ou des) valeur propre négative dans les cas suivants :
 - a. Le cas d'une ACP
 - b. Le cas d'une AFC
 - c. Le cas d'une ACM
2. Nous considérons un échantillon de N individus caractérisés par « m » variables. Chaque variable « j » de ces « m » variables est composée de P_j modalités. Soit P le nombre total de modalités.
 - a. Quelle est l'expression de l'inertie totale ?
 - b. On suppose que toutes les variables sont binaires. Quelle serait la valeur de l'inertie totale.