

Rédigé par : l'équipe enseignante

Durée : 1h30

A l'intention de : Elèves d'ING1-GI

Document ou matériel autorisés : Calculatrice uniquement

Il s'agit d'étudier un jeu de données constitué de 29 fromages pour lesquels on connaît 10 caractéristiques : Calories, sodium, calcium, lipides, rétinol (vitamine A), folates (Vitamine B9), protéines, cholestérol, magnésium

	calories	sodium	calcium	lipides	retinol	folates	proteines	cholesterol	magnesium
CarredelEst	314	353,5	72,6	26,3	51,6	30,3	21	70	20
Babybel	314	238	209,8	25,1	63,7	6,4	22,6	70	27
Beaufort	401	112	259,4	33,3	54,9	1,2	26,6	120	41
Bleu	342	336	211,1	28,9	37,1	27,5	20,2	90	27

...

1. Partie I : Indicateurs numériques

Concernant la variable lipides, nous obtenons les indicateurs suivants :

- Médiane : $m=26.30$
- Ecart-inter quantile : $Q3-Q1=5.70$
- Moyenne : $\bar{x}=24.16$
- Variance : $s^2=66.09$

Expliquez l'impact des opérations suivantes sur ces indicateurs. Calculer leur nouvelle valeur quand c'est possible.

- 1.1 On ajoute 10 à la variable
- 1.2 On multiplie la variable par 2
- 1.3 On ajoute un fromage avec un teneur en lipides de 70

2. Partie II : Analyse en composantes principales

Une première étude a été menée au travers une ACP. A partir des résultats fournis en annexe, répondez aux questions suivantes.

- 2.1 Combien d'axes retenir-vous dans votre étude ? Quel est le pourcentage de variance expliqué ?
- 2.2 Expliquez pourquoi les variables lipides, calories, cholesterol, proteines, magnesium se retrouvent dans un même groupe ?
- 2.3 Expliquez la contradiction entre le faible coefficient de corrélation entre le rétinol et le calcium et leur représentation sur le cercle de corrélation.
- 2.4 Quelles variables contribuent à la formation des axes 1 et 2 avec quel pourcentage ?
- 2.5 Que se passe-t-il dans l'ACP quand il y a un ou des individus atypiques ? Pensez-vous qu'il y a un ou des individus atypiques ici ? Justifier votre réponse.
- 2.6 Donner une interprétation des fromages suivants : Beaufort, Chaource, Pont Leveque

3. Partie III : Clustering

On décide de former des groupes de fromages. On utilise la classification hiérarchique ascendante pour déterminer les classes.

- 1) Que mesure la distance de Ward ?
- 2) A l'aide des graphiques ci-dessus, déterminez le nombre de classes à retenir.
- 3) Représentez les classes retenues sur le graphique des individus de l'ACP. Donner une interprétation de ces classes.
Attention : N'oubliez pas de rendre le graphique des individus avec votre nom et votre groupe dans votre copie
- 4) Chaque fromage est maintenant associé à un groupe. On souhaite déterminer si la teneur en lipide est liée au groupe. Quel critère allez-vous calculer ? Que représente-t-il ? Comment le calcule-t-on ?

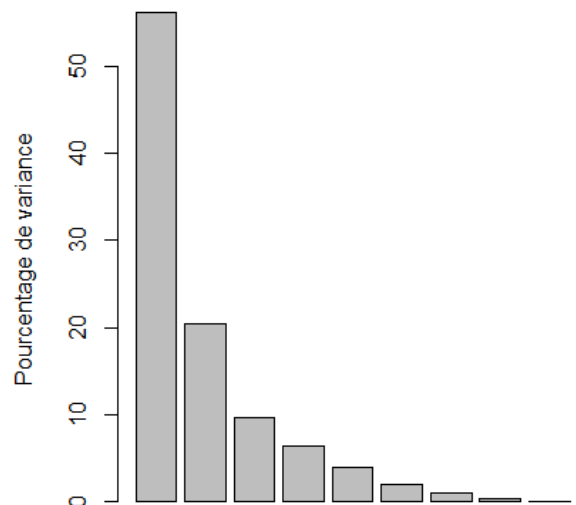
4. Annexes

Matrice des corrélations

	<i>calories</i>	<i>sodium</i>	<i>calcium</i>	<i>lipides</i>	<i>retinol</i>	<i>folates</i>	<i>proteines</i>	<i>cholesterol</i>	<i>magnesium</i>
<i>calories</i>	1								
<i>sodium</i>	0,45	1							
<i>calcium</i>	0,43	0,01	1						
<i>lipides</i>	0,98	0,48	0,34	1					
<i>retinol</i>	-0,04	0,14	-0,29	-0,02	1				
<i>folates</i>	-0,32	0,14	-0,64	-0,28	0,52	1			
<i>proteines</i>	0,89	0,28	0,61	0,81	-0,04	-0,35	1		
<i>cholesterol</i>	0,96	0,33	0,43	0,96	-0,09	-0,37	0,82	1	
<i>magnesium</i>	0,75	0,03	0,71	0,69	-0,10	-0,45	0,79	0,75	1

Résultats sur les valeurs propres de l'ACP

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	5,05	56,10	56,10
comp 2	1,84	20,49	76,59
comp 3	0,87	9,64	86,24
comp 4	0,58	6,42	92,65
comp 5	0,36	3,95	96,60
comp 6	0,18	1,95	98,55
comp 7	0,10	1,08	99,63
comp 8	0,03	0,32	99,95
comp 9	0,00	0,05	100,00

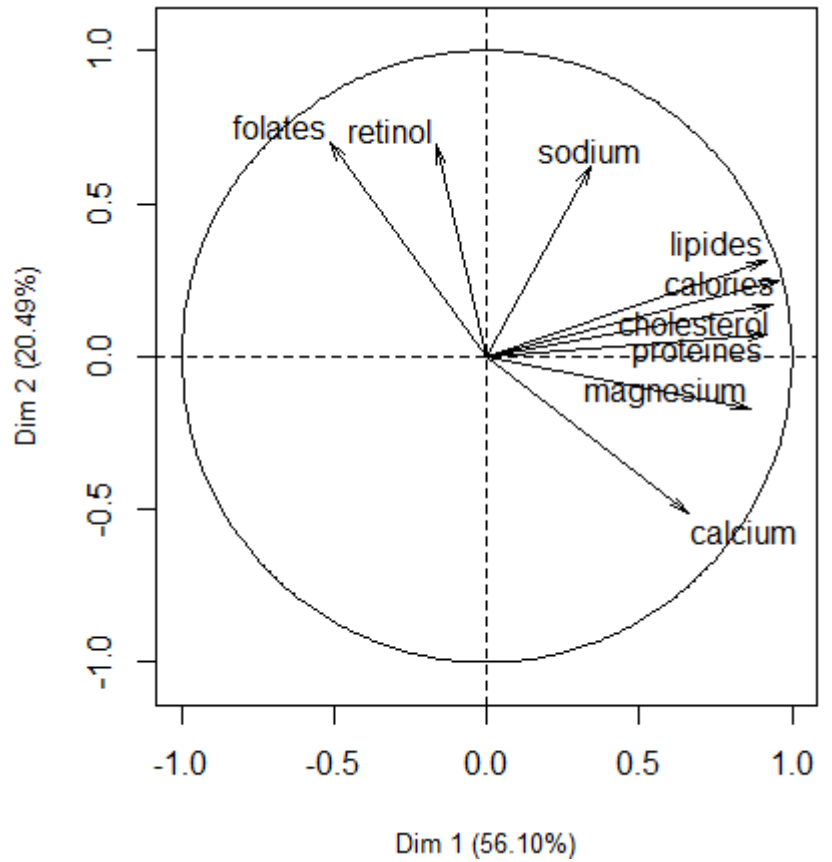


ING1-GI : DATA EXPLORATION - EXAMEN 2016-2017

Résultats de l'ACP pour les variables

	§cos2		
	Dim.1	Dim.2	Dim.3
calories	0,91	0,06	0,00
sodium	0,11	0,39	0,29
calcium	0,44	0,26	0,03
lipides	0,84	0,10	0,01
retinol	0,03	0,48	0,36
folates	0,26	0,49	0,03
proteines	0,84	0,01	0,02
cholesterol	0,88	0,03	0,00
magnesium	0,74	0,03	0,12

	§contrib		
	Dim.1	Dim.2	Dim.3
calories	18,04	18,04	18,04
sodium	2,27	2,27	2,27
calcium	8,65	8,65	8,65
lipides	16,57	16,57	16,57
retinol	0,54	0,54	0,54
folates	5,19	5,19	5,19
proteines	16,66	16,66	16,66
cholesterol	17,34	17,34	17,34
magnesium	14,75	14,75	14,75



Résultats de l'ACP pour les individus

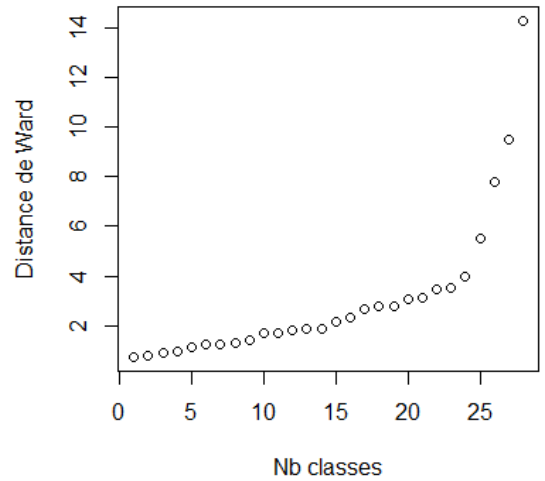
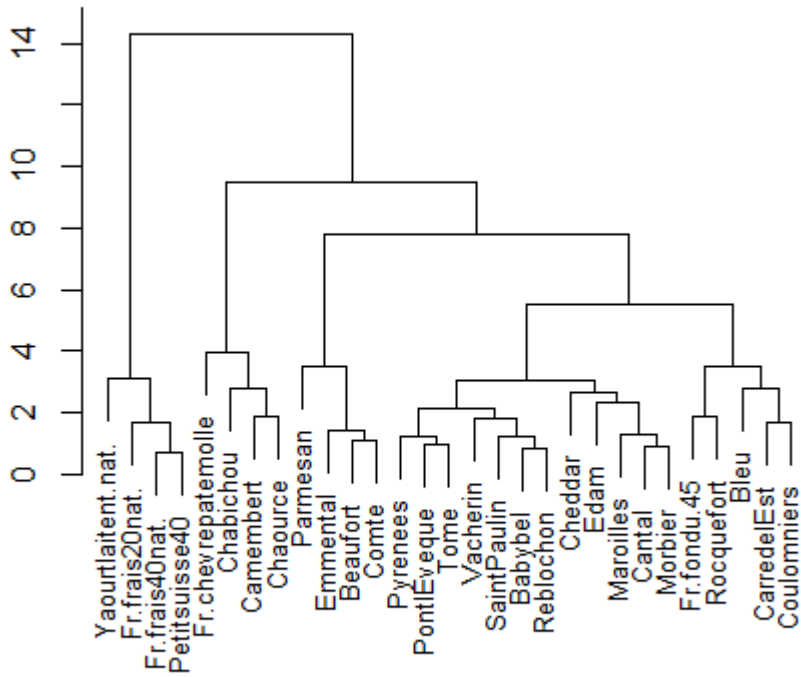
Attention : Le graphique des individus se trouve sur une feuille à part, à rendre avec votre copie

	§cos2			§contrib		
	Dim 1	Dim 2	Dim 3	Dim 1	Dim 2	Dim 3
CarredelEst	0,05	0,43	0,30	0,28	6,08	9,03
Babybel	0,31	0,15	0,08	0,15	0,21	0,24
Beaufort	0,78	0,10	0,03	5,86	2,06	1,23
Bleu	0,10	0,09	0,32	0,40	0,90	7,17
Camembert	0,11	0,42	0,07	0,65	6,94	2,57
Cantal	0,70	0,10	0,10	1,85	0,72	1,54
Chabichou	0,01	0,54	0,12	0,03	8,15	3,77
Chaource	0,10	0,75	0,12	0,57	11,57	3,93
Cheddar	0,65	0,02	0,01	2,47	0,20	0,24
Comte	0,72	0,07	0,07	6,91	1,96	3,90
Coulomniens	0,16	0,29	0,16	0,48	2,41	2,73
Edam	0,45	0,26	0,21	1,57	2,53	4,32
Emmental	0,62	0,21	0,13	5,65	5,24	6,86
Fr.chevrepate molle	0,45	0,29	0,20	6,86	12,24	17,49
Fr.fondu.45	0,02	0,13	0,32	0,04	1,05	5,32

ING1-GI : DATA EXPLORATION - EXAMEN 2016-2017

	\$cos2			\$contrib		
	Dim 1	Dim 2	Dim 3	Dim 1	Dim 2	Dim 3
Fr.frais20nat.	0,86	0,12	0,00	15,74	6,23	0,12
Fr.frais40nat.	0,93	0,04	0,01	14,20	1,55	0,53
Maroilles	0,59	0,06	0,13	2,04	0,59	2,56
Morbier	0,72	0,01	0,17	1,15	0,03	1,58
Parmesan	0,59	0,01	0,18	5,98	0,21	10,52
Petitsuisse40	0,92	0,04	0,00	12,83	1,35	0,34
PontlEveque	0,01	0,17	0,46	0,01	0,44	2,57
Pyrenees	0,46	0,01	0,34	0,59	0,05	2,52
Reblochon	0,28	0,01	0,03	0,14	0,02	0,08
Rocquefort	0,16	0,50	0,17	0,79	6,77	4,92
SaintPaulin	0,13	0,43	0,00	0,16	1,42	0,00
Tome	0,01	0,06	0,41	0,02	0,20	3,19
Vacherin	0,17	0,48	0,02	0,31	2,35	0,22
Yaourtlaitent.nat.	0,63	0,31	0,00	12,27	16,53	0,50

Résultats de CAH



NOM :

PRENOM :

GROUPE :

Graphique de l'ACP pour les individus

