

Examen : Traitement statistique des données sensorielles

Durée 2h00. Le sujet comporte 5 pages ainsi que 2 pages pour les annexes. Les documents ainsi que les téléphones portables, tablettes, ordinateurs ne sont pas autorisés. La calculatrice est autorisée. La qualité de la rédaction sera prise en compte.

Exercice 1. Comparaison de deux produits sur une échelle d'acidité On a demandé à 33 juges de noter deux yaourts A et B sur une échelle d'acidité. On note x_j^A (respectivement x_j^B) la note du j -ème juge pour le produit A (resp. B). Voici les résultats :

j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
x_j^A	6.69	6.38	4.08	5.42	7.72	6.75	6.92	6.77	4.93	5.37	5.02	5.85	6.38	6.22	4.58	7.63	6.83	6.08
x_j^B	6.76	6.46	5.41	4.30	6.36	6.71	6.19	5.53	8.69	5.75	6.75	6.65	6.98	4.39	5.81	4.76	7.62	5.48
j	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33			
x_j^A	6.67	5.75	7.76	5.44	7.41	5.38	5.08	7.90	5.47	4.73	6.43	7.44	4.14	5.76	4.57			
x_j^B	6.04	4.66	5.66	4.90	6.38	6.31	6.03	8.34	7.05	5.56	6.21	7.57	6.25	5.94	6.84			

On donne $\sum_{j=1}^{33} x_j^A = 199.55$, $\sum_{j=1}^{33} x_j^B = 204.34$, $\sum_{j=1}^{33} (x_j^A)^2 = 1244.755$, $\sum_{j=1}^{33} (x_j^B)^2 = 1298.719$, $\sum_{j=1}^{33} x_j^A x_j^B = 1240.982$.

- Dans le cadre d'un modèle de notations à perturbations gaussiennes :
 - Démontrer que $\sum_{j=1}^{33} (x_j^A - x_j^B)^2 = 61.51$.
 - Calculer un intervalle de confiance à 95% pour la différence d'acidité entre A et B . *Indication : on donne les quantiles de la loi $\mathcal{N}(0; 1)$ suivant : $q_{0.95} = 1.645$ et $q_{0.975} = 1.96$.*
 - Tester au niveau 0.05 la significativité statistique de la différence d'acidité constatée en moyenne pour A et B .
- Dans le cas où d'un modèle de notations à perturbations quelconques, quel test peut-on utiliser ? Décrire succinctement le principe de ce test (5 lignes maximum).

Exercice 2. Analyse de jus d'orange On a mesuré 7 variables sensorielles pour 6 jus d'orange dégustés à température ambiante ("amb.") ou frais ("fr."). Voici les notes relevées :

	Odeur (Intensité)	Odeur (spécificité)	Pulpe	Goût	Acide	Amer	Sucré
Pampril amb.	2.82	2.53	1.66	3.46	3.15	2.97	2.60
Tropicana amb.	2.76	2.82	1.91	3.23	2.55	2.08	3.32
Fruvita fr.	2.83	2.88	4.00	3.45	2.42	1.76	3.38
Joker amb.	2.76	2.59	1.66	3.37	3.05	2.56	2.80
Tropicana fr.	3.20	3.02	3.69	3.12	2.33	1.97	3.34
Pampril fr.	3.07	2.73	3.34	3.54	3.31	2.63	2.90

On procède alors à une Analyse en Composantes Principales (ACP) normée. Toutes les grandeurs présentées dans la suite sont arrondies à 10^{-3} près, on se méfiera donc des erreurs d'arrondis.

- Quelques généralités :**
 - Qu'est ce qu'un individu statistique dans ce cas ? Quelles sont les dimensions en jeu ?
 - Rappeler brièvement le principe de l'ACP normée (5 lignes maximum).

(c) Voici les valeurs propres $\lambda_1, \dots, \lambda_6$

	valeurs propres	% de variance	% cumulés de variance
λ_1	4.744	67.767	67.767
λ_2	1.333	19.047	86.814
λ_3	0.820	11.712	98.526
λ_4	0.084	1.200	99.726
λ_5	0.019	0.274	100.000
λ_6	0	0	100

Justifier le nombre d'axes factoriels qu'il est pertinent de considérer pour analyser ce jeu de données.

(d) La valeur propre λ_6 est nulle. Expliquer ce que cela révèle de la structure des données et ce que cela implique sur les axes factoriels.

2. À propos des variables

(a) Voici le tableau des "cos²" dans l'espace dual :

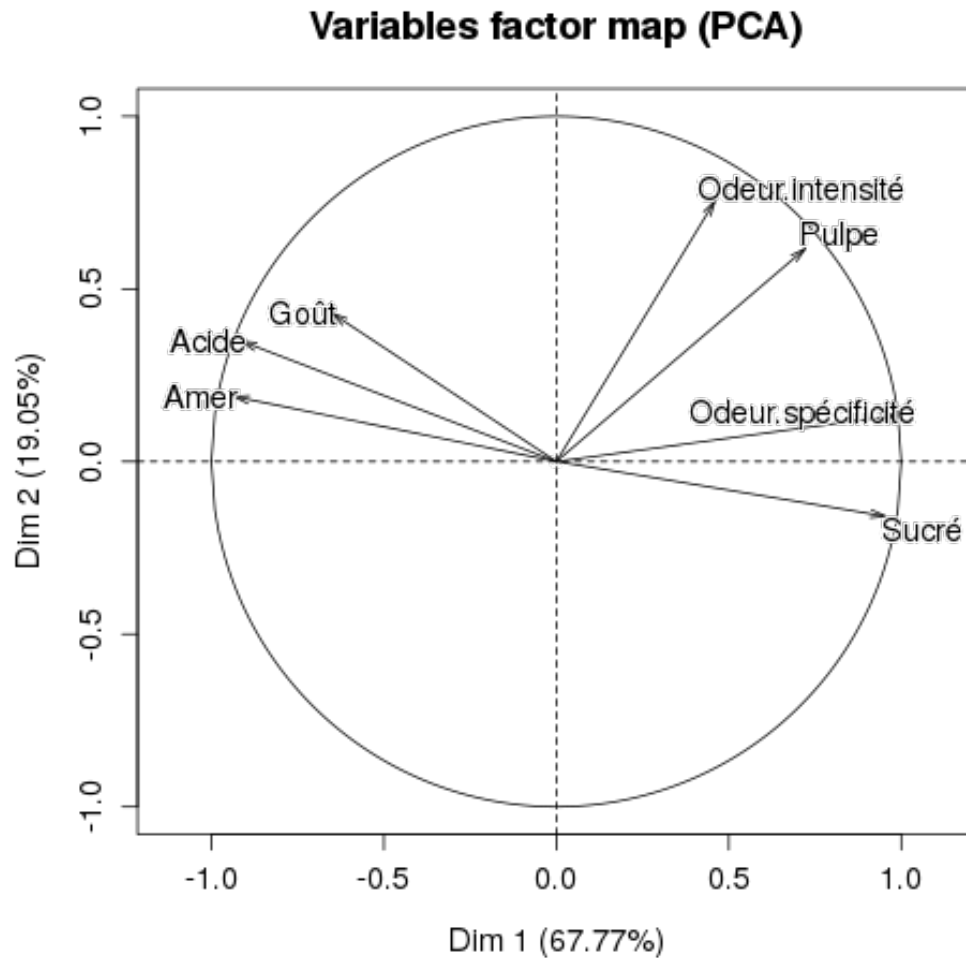
	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Odeur (intensité)	0.211	0.569	0.219	0.000	0.000
Odeur (spécificité)	0.971	0.018	0.003	0.006	0.002
Pulpe	0.521	0.380	0.089	0.009	0.001
Goût	0.422	0.184	0.391	0.000	0.002
Acide	0.833	0.121	0.000	0.042	0.003
Amer	0.874	0.035	0.081	0.001	0.009
Sucré	0.912	0.025	0.035	0.026	0.002

... et le tableau des corrélations :

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Odeur (intensité)	0.460	0.754	-0.468	0.008	0.004
Odeur (spécificité)	0.985	0.134	-0.058	0.077	0.041
Pulpe	0.722	0.617	0.298	-0.096	-0.031
Goût	-0.650	0.429	0.626	0.005	0.048
Acide	-0.913	0.348	-0.021	0.205	-0.057
Amer	-0.935	0.188	-0.285	-0.028	0.093
Sucré	0.955	-0.159	0.187	0.161	0.048

- i. Quelle interprétation géométrique des "cos²" peut-on faire ?
 - ii. Quelle interprétation géométrique de la corrélation peut-on faire ?
 - iii. Quel est la relation entre ces 2 tableaux? *Indication : on rappelle que l'on a procédé à une ACP normée.*
- (b) Quelles sont les trois variables les moins bien représentées sur le premier axe factoriel dual? Quelle est la variable la moins bien représentée sur le premier plan factoriel dual?
- (c) Combien d'axes faut-il considérer pour visualiser l'essentiel des disparités en termes de goût ?

(d) Voici le cercle de corrélation pour les 2 premiers axes :



Compléter l'annexe 1 : le cercle de corrélations pour les axes 1 et 3.

- (e) Qualifier (en justifiant) les couples de variables suivants de positivement corrélés, négativement corrélés, décorrélés : (Odeur intensité, sucré),(sucré,amer), (Goût,Pulpe).
- (f) Identifier 3 faisceaux de variables dans le premier plan factoriel et interpréter le premier axe factoriel.

3. À propos des individus

(a) Voici le tableau des “cos²” dans l'espace factoriel direct :

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Pampryl amb.	0.970	0.001	0.012	0.014	0.003
Tropicana amb.	0.201	0.753	0.002	0.039	0.004
Fruvita fr.	0.557	0.000	0.434	0.008	0.000
Joker amb.	0.820	0.158	0.005	0.000	0.016
Tropicana fr.	0.823	0.028	0.147	0.002	0.000
Pampryl fr.	0.233	0.733	0.008	0.026	0.000

... et le tableau des coordonnées des jus d'orange dans l'espace factoriel direct :

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Pampryl amb.	-2.984	-0.082	-0.333	-0.356	0.168
Tropicana amb.	0.886	-1.715	-0.087	0.393	0.123
Fruvita fr.	1.937	0.040	1.710	-0.232	-0.007
Joker amb.	-1.896	-0.834	-0.154	-0.007	-0.266
Tropicana fr.	3.186	0.589	-1.345	-0.173	-0.027
Pampryl fr.	-1.129	2.002	0.209	0.374	0.009

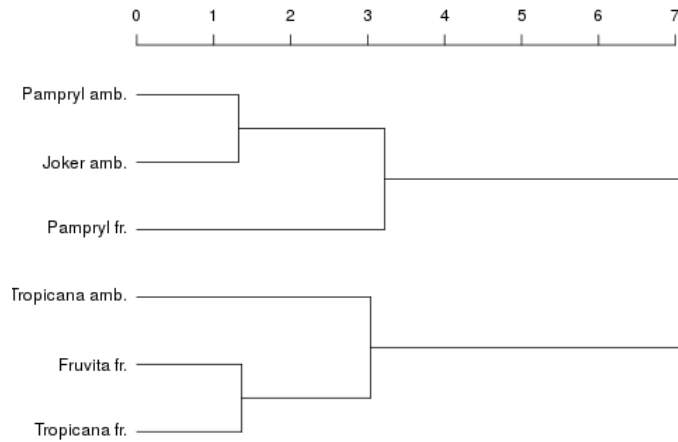
- i. Interpréter géométriquement le tableau des "cos²".
 - ii. (Question bonus) Quel est la relation entre les deux tableaux? *Indication : penser à normaliser les individus.*
- (b) Quels sont les individus les moins bien représentés sur le premier axe factoriel? Même question pour le premier plan factoriel?
 - (c) Dans quel espace à 2 dimensions le "Fruvita fr." est-il correctement représenté?
 - (d) Voici le tableau des contributions des individus pour chacun des axes factoriels directs (c'est le pourcentage d'inertie le long de chaque axe factoriel qui est expliqué par un individu) :

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Pampryl amb.	31.288	0.085	2.254	25.073	24.634
Tropicana amb.	2.761	36.771	0.154	30.563	13.083
Fruvita fr.	13.182	0.020	59.450	10.635	0.046
Joker amb.	12.631	8.686	0.481	0.009	61.525
Tropicana fr.	35.660	4.335	36.774	5.930	0.635
Pampryl fr.	4.479	50.102	0.887	27.790	0.076

Justifier quels sont les produits qui contribuent le plus au troisième axe factoriel direct?

- (e) En annexe 2 est donné le premier plan factoriel direct. Tracer sur ce graphique les ou l'axe(s) correspondant à la ou aux variable(s) qui permette(nt) de séparer au mieux les jus d'orange selon la température de dégustation (ambient : amb. ou frais : fr.).
- (f) Les marques Tropicana et Fruvita sont des jus d'orange de Floride (les autres marques ont une origine différente). Toujours en annexe 2, tracer la ou les variable(s) qui permet(tent) d'expliquer au mieux la différence de provenance?

- (g) On procède à une Classification Ascendante Hiérarchique (CAH) sur les projections des individus dans le premier plan factoriel. Voici le dendrogramme associé :



À partir du dendrogramme :

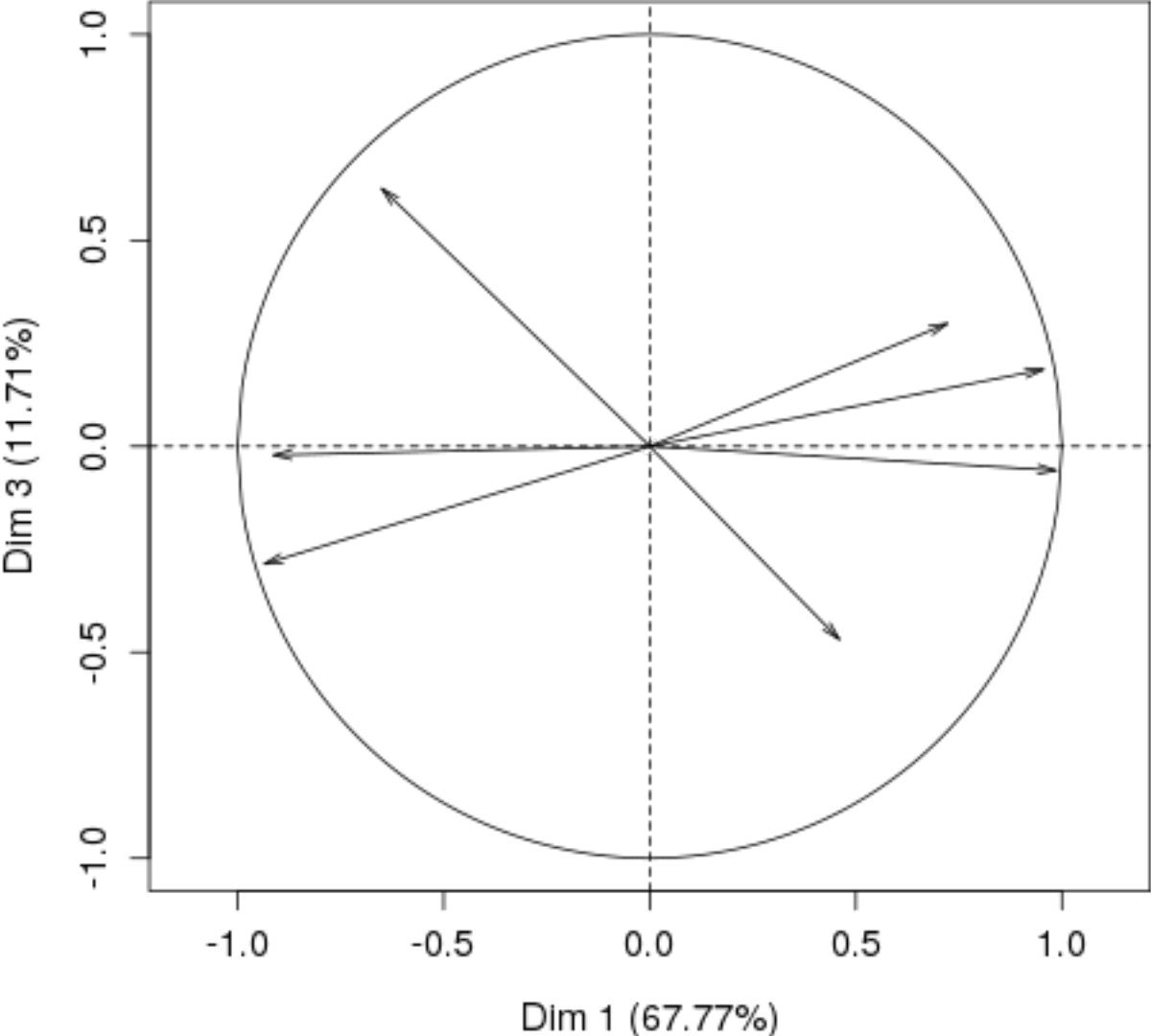
- On propose de classer les jus d'orange en 2 classes : identifier les classes et les interpréter.
- Proposer (en justifiant) un autre découpage pertinent. Identifier alors les classes.

Exercice 3. Loi du χ^2 (Bonus) Soit X et Y deux v.a indépendantes et identiquement distribuées qui suivent une loi du $\chi^2(5)$:

- Déterminer la loi de la v.a $Z = X + Y$.
- Calculer l'espérance et la variance de Z . *Indication : si T suit un $\chi^2(1)$ alors $\mathbb{E}(T) = 1$ et $\text{Var}(T) = 2$.*
- Tracer qualitativement la densité de la loi de X et la densité de la loi de Z .

Annexe 1

Variables factor map (PCA)



Annexe 2

Individuals factor map (PCA)

