

Examen Semestriel Analyse Exploratoire des Données

Partie 1

On a présenté à 16 juges 28 lots de pêches et nectarines (14 de chaque). Chaque juge a donné son avis en les classant par un rang allant de 1 (le plus apprécié) à 28 (le moins apprécié). On obtient le tableau suivant :

تم عرض 28 عينة من الخوخ والنكتارين (14 من كل نوع) على 16 حكماً. قام كل حكم بإبداء رأيه من خلال ترتيب هذه العينات برتبة تتراوح من 1 (الأكثر تفضيلاً) إلى 28 (الأقل تفضيلاً). حصلنا على الجدول التالي:

	J1	J2	J3	J4	J5	J6	J7	J8	J9	J10	J11	J12	J13	J14	J15	J16
1.nec	10	5	8	3	1	18	5	17	3	4	1	2	5	3	1	1
2.nec	3	1	9	8	6	16	8	10	2	1	8	8	9	5	6	4
3.pea	5	11	5	2	8	8	18	3	4	15	14	4	7	1	3	13
4.pea	6	12	3	4	4	7	17	2	1	16	13	7	3	8	4	14
5.pea	4	2	4	14	17	10	16	1	5	19	21	13	6	2	5	15
6.nec	2	6	16	10	13	2	11	5	13	8	10	14	10	9	15	8

.....

Question 1 (3 pts): La somme des colonnes du tableau est indiquée ci-dessous. Commentez.

J1	J2	J3	J4	J5	J6	J7	J8	J9	J10	J11	J12	J13	J14	J15	J16
406	406	406	406	406	406	406	406	406	406	406	406	406	406	406	406

Question 2 (2 pts): La matrice de corrélations des choix des juges est présentée ci-dessous. Y a-t-il deux juges qui ont fait le même classement ?

	J1	J2	J3	J4	J5	J6	J7	J8	J9	J10	J11	J12	J13	J14	J15	J16
J1	1	0.85	0.15	0.32	-0	0.32	0.41	0.53	0.71	0.68	0.52	0.26	0.46	0.72	0.37	0.45
J2		1	0.21	0.37	0.02	0.24	0.46	0.5	0.68	0.7	0.59	0.29	0.47	0.68	0.34	0.6
J3			1	0.68	0.18	-0.1	-0.2	0.32	0.25	-0.2	0.29	0.66	0.68	0.45	0.55	0.28
J4				1	0.34	-0.2	0.32	0.33	0.56	0.19	0.47	0.81	0.6	0.61	0.66	0.49
J5					1	0.03	-0	-0.1	0.15	-0.1	-0.1	0.32	-0	0.18	0.45	-0.2
J6						1	-0.2	0.59	0.15	0.08	0.06	-0.1	0.18	0.12	-0.1	0
J7							1	0.13	0.7	0.79	0.52	0.29	0.09	0.41	0.09	0.51
J8								1	0.59	0.26	0.39	0.4	0.6	0.59	0.29	0.5
J9									1	0.71	0.55	0.53	0.47	0.77	0.5	0.57
J10										1	0.64	0.3	0.15	0.52	0.17	0.59
J11											1	0.48	0.59	0.63	0.3	0.63
J12												1	0.55	0.62	0.66	0.56
J13													1	0.68	0.36	0.54
J14														1	0.69	0.6
J15															1	0.27
J16																1

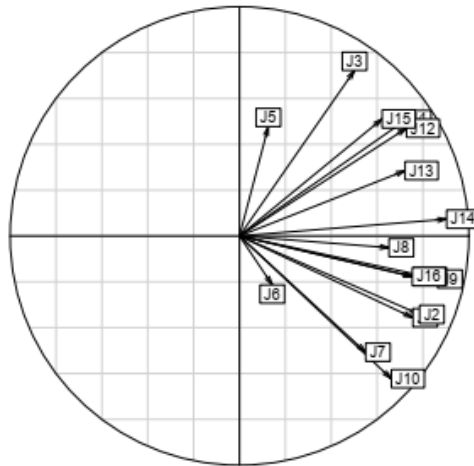
Question 3 (3 pts): On réalise une ACP sur les variables centrées réduites ; les variances des 7 premières composantes principales sont données ci-dessous.

Combien de composantes principales faut-il garder ?

Donner, sans calcul, la proportion de l'inertie expliquée.

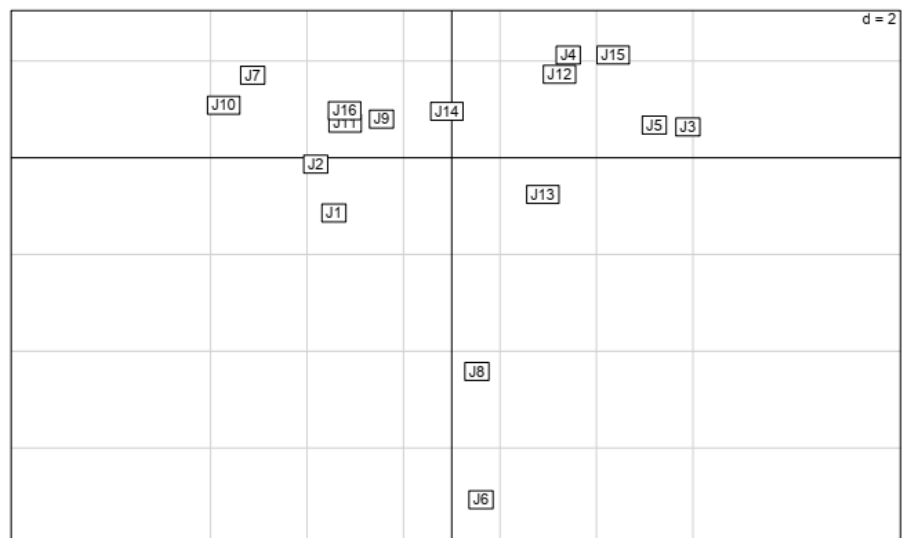
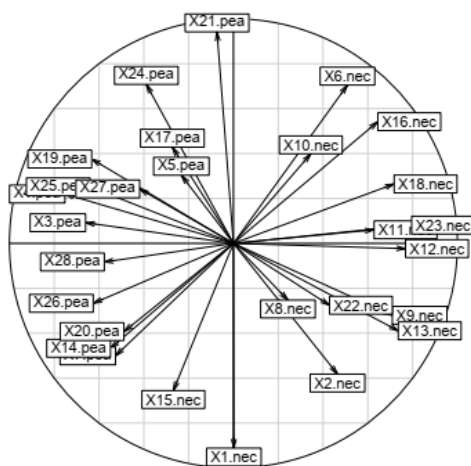
	1	2	3	4	5	6	7
Inertie	7.32	2.61	1.80	1.30	0.68	0.53	0.41
Ratio	0.46	0.62	0.73	0.81	0.86	0.89	0.92

Question 4 (3 pts): On donne ci-dessous le cercle de corrélation. Y a-t-il des juges qui se distinguent des autres ? Comment peut-on interpréter la première composante principale ? Et la deuxième ?



Partie 2

On fait une nouvelle ACP sur les variables centrées réduites du tableau transposé, c'est-à-dire que les juges sont en ligne et non en colonne. On obtient pour le premier plan principal le cercle de corrélation et le graphe des individus suivants.



Question 5 (2 pts): Quel genre d'information peut-on attendre de cette nouvelle ACP ?

Question 6 (3 pts): Quelle est l'interprétation de la première composante principale à partir du cercle de corrélation ? Quelle différence voit-on avec l'analyse précédente (Partie 1, Question 4) ?

Question 7 (2 pts): Le deuxième axe principal peut-il être interprété à l'aide du graphe des individus ?

Question 8 (2 pts): Cette analyse des données peut-elle être utilisée conjointement avec celle où les juges sont en colonne ?

Bonne réussite

Dr. T. Mehenni

Examen Semestriel Analyse Exploratoire de Données Correction

Partie 1

Question 1 (3 pts): La somme des colonnes du tableau est indiquée sur le tableau. Commentez.

Réponse : Chaque colonne contient tous les nombres de 1 à 28 dans un ordre arbitraire. La somme de la colonne vaut donc $\sum_{i=1}^{28} i = \frac{28(28+1)}{2} = 406$

Question 2 (2 pts): Y a-t-il deux juges qui ont fait le même classement ?

Réponse : Si deux juges avaient fait le même classement, leur corrélation serait égale à 1. Or, toutes les corrélations sont inférieures à 0,85. Il n'y a donc pas deux juges identiques.

Question 3 (3 pts): Combien de composantes principales faut-il garder ? Donner, sans calcul, la proportion de l'inertie qui est expliquée.

Réponse : Si on utilise la règle de Kaiser, il faut garder les 4 premières valeurs propres, qui conservent 81% de l'inertie.

Question 4 (3 pts): Y a-t-il des juges qui se distinguent des autres ? Comment peut-on interpréter la première composante principale ? Et la deuxième ?

Réponse : Sur le cercle des corrélations, on peut repérer que les juges 5 et 6 sont moins corrélés avec les composantes principales que les autres. On peut en déduire qu'ils ont une opinion originale.

Pour la première composante principale, on voit que toutes les variables sont corrélées positivement avec cette composante. Cette composante reflète donc la qualité des fruits.

On ne peut pas interpréter la seconde composante à seule vue du cercle des corrélations, puisqu'on ne sait pas ce qui différencie les juges entre eux.

Partie 2

Question 5 (2 pts): Quel genre d'information peut-on attendre de cette nouvelle ACP ?

Réponse : L'ACP avec les juges en colonne permet de travailler sur les jugements eux-mêmes (compromis, différence pêches/nectarines).

Question 6 (3 pts): Quelle est l'interprétation de la première composante principale à partir du cercle de corrélation ? Quelle différence voit-on avec l'analyse précédente (Partie 1, Question 4) ?

Réponse : La première composante principale sépare les nectarines (à droite) des pêches (à gauche). Dans l'analyse précédente, on avait une composante de compromis qui reflète la qualité globale des lots de fruits. On s'intéresse plus aux différences des lots de fruits qu'à leurs points communs.

Question 7 (2 pts): Le deuxième axe principal peut-il être interprété à l'aide du graphe des individus ?

Réponse : Le graphe des individus montre que la seconde composante principale est déterminée principalement par les juges 6 et 8.

Question 8 (2 pts): Cette analyse des données peut-elle être utilisée conjointement avec celle où les juges sont en colonne ?

Réponse : Non. Les deux approches sont opposées. L'ACP cherche les combinaisons linéaires des variables qui différencient au plus les individus. Il faut donc placer les éléments que l'on veut étudier en ligne.