

Devoir surveillé d'Analyse des données

1 ère année du cycle de formation d'ingénieurs

Durée de l'épreuve : 1 heure 30 - Documents non autorisés

Nombre de pages : 2 - Date de l'épreuve : 6 mars 2018

Exercice 1 : On considère le tableau de données, noté X , qui est défini par :

$$X = \begin{array}{c|ccccc} & j_1 & j_2 & j_3 & j_4 & j_5 \\ \hline i_1 & 1 & 6 & 11 & 16 & 21 \\ i_2 & 2 & 7 & 12 & 17 & 22 \\ i_3 & 3 & 8 & 13 & 18 & 23 \\ i_4 & 4 & 9 & 14 & 19 & 24 \\ i_5 & 5 & 10 & 15 & 20 & 25 \end{array}$$

où la $j^{\text{ème}}$ colonne désigne la variable x^j et la $i^{\text{ème}}$ ligne désigne l'individu x_i . Par la suite, on considère les résultats de l'ACP sur matrice variance du tableau X (i.e., ACP de X lorsque \mathbb{R}^5 est muni de la métrique identité et chaque individu possède un poids égal à $1/5$).

1. Soit g le centre de gravité du nuage des cinq individus associés au tableau X . Vérifier que $g = x_{i_3}$ puis calculer le tableau centré, noté Y , qui est associé à X .
2. Combien existe-t-il d'axes factoriels non triviaux ? (justifier votre réponse)

3. Montrer que le vecteur $\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ est un vecteur directeur d'un axe factoriel non trivial

que l'on déterminera (on choisira sa première coordonnée de façon à ce qu'elle soit positive).

4. Déterminer l'inertie totale du nuage étudié.
5. Calculer la première composante principale de l'individu i_5 .
6. Déterminer la représentation des variables dans le nouveau système d'axes factoriels de l'espace des variables.

Exercice 2 : Afin d'étudier l'image de 12 assurances auprès de leurs clients un questionnaire a été administré auprès des clients des 12 assurances. Dans ce questionnaire on a demandé aux répondants d'indiquer, pour l'assurance dont ils sont clients, leur degré d'accord ((1) si pas du tout d'accord, (2) si plutôt pas d'accord, (3) si neutre, (4) si plutôt d'accord, (5) si tout à fait d'accord) pour chacune des 10 affirmations suivantes :

Item 1.	Des amis ou des membres de votre famille y sont déjà clients
Item 2.	Bonne relation avec la clientèle
Item 3.	Elle a une bonne réputation
Item 4.	Vous avez une connaissance qui travaille dans cette assurance
Item 5.	Bonne information sur ses produits et services
Item 6.	Sa communication et ses publicités vous plaisent
Item 7.	Les services qu'elle offre sont meilleurs que ceux des autres assurances
Item 8.	Les services qu'elle offre sont moins chers que ceux des autres assurances
Item 9.	Elle propose les meilleures offres
Item 10.	Vous avez confiance en cette assurance

A la fin de l'étude on disposait d'un tableau, que nous avons appelé Assurances, donnant les scores moyens des 12 assurances sur les 10 affirmations. On a réalisé une ACP normée sur les variables de ce tableau. Les résultats de cette ACP sont les suivants :

```
> library(FactoMineR)
> res.pca = PCA(Assurances, graph=F)
> round(res.pca$eig[1:6,1],2)# Les 6 premières valeurs propres
[1] 4.25 2.14 1.63 0.71 0.48 0.36
> round(cbind(res.pca$var$coord[,1:3]),2)
      Dim.1 Dim.2 Dim.3
Item 1  0.86 -0.03 -0.38
Item 2   0.50  0.50 -0.01
Item 3  0.88  0.08 -0.39
Item 4   0.79 -0.51 -0.04
Item 5   0.15  0.83  0.15
Item 6   0.43  0.77 -0.09
Item 7   0.64 -0.37  0.47
Item 8   0.27  0.36  0.76
Item 9   0.59 -0.25  0.68
Item 10  0.90 -0.07 -0.22
> round(cbind(res.pca$ind$coord[,1:3]),2)
      Dim.1 Dim.2 Dim.3
1      3.37  2.07  1.02
2      3.51 -0.40  0.20
3      0.16 -0.12  1.00
4     -2.16 -1.02  0.01
5     -1.53 -1.18  1.28
6     -1.51 -1.28  1.36
7      3.43 -1.87 -1.63
8     -0.82 -1.13  0.22
9     -1.39  0.44 -2.62
10    -1.58  2.21 -1.66
11    -0.65 -0.39 -0.38
12    -0.83  2.67  1.21
```

1. Déterminer le nombre d'axes à retenir.
2. Donner une interprétation des 3 premiers axes.
3. Représenter la première carte des individus puis donner une interprétation de cette carte.