

Exercice, chapitre 9

Analyse en composantes principales

Mars 2019

1. Écrire une fonction en langage R qui réalise l'analyse en composantes principales. Inclure les options suivantes: (a) calcul à partir des données brutes centrées ou des données centrées réduites; (b) cadrage de type 1 (préserver les distances) ou de type 2 (préserver les corrélations). **Fournissez une copie de votre fonction dans votre rapport.**

Réalisez l'analyse en composantes principales de la matrice suivante, à partir de sa matrice de dispersion (données centrées mais **non réduites**):

$$\text{mat} = \begin{bmatrix} 2 & 2 & 5 \\ 6 & 2 & 2 \\ 7 & 4 & 3 \\ 10 & 4 & 4 \end{bmatrix}$$

2. Réalisez l'ACP de mat à l'aide de votre fonction en langage R. **Fournissez une copie de vos résultats dans votre rapport.**

3. Pour vérifier vos résultats, réalisez l'ACP à l'aide de la fonction `rda()` de `{vegan}`. Y a-t-il des différences par rapport à vos résultats? Expliquez pourquoi après avoir consulté le document « `decision-vegan.pdf` », page 5 ainsi que le tableau 2 de la page 6. Vous trouverez ce document dans le dossier de `vegan` sur votre ordinateur. Vous pouvez aussi comparer vos résultats à ceux de la fonction `PCA.newr()` du livre *Numerical ecology with R*.

4. Utilisez les résultats obtenus de votre fonction pour répondre aux questions suivantes. Ces questions pourront vous demander des calculs additionnels, à l'ordinateur ou à la main.

4.1. Matrice de covariance, valeurs propres et vecteurs propres (normés à 1).

4.2. Cadrage de type 1: représentez les objets dans le plan des deux premières composantes principales (matrice de résultats numériques et dessin).

4.3. Cadrage de type 1: représentez les descripteurs dans le plan des deux premières composantes principales (matrice de résultats numériques et dessin).

4.4. Pour le graphique avec cadrage de type 1 (question 4.3), calculez la contribution équilibrée des descripteurs et tracez le **cercle de contribution équilibrée** sur le graphique.

4.5. Cadrage de type 2: représentez les descripteurs dans le plan des deux premiers axes. Dans ce cadrage, les vecteurs propres sont normés à $\sqrt{\lambda}$. Résultats numériques et dessin.

4.6. Quel(s) descripteur(s) contribue(nt) fortement à la formation de chacun des trois axes principaux ? Pour répondre à cette question, examinez les valeurs contenues dans les matrices \mathbf{U} et $\mathbf{U}\mathbf{\Lambda}^{1/2}$.

4.7. Calculez de deux façons la corrélation entre les descripteurs et les composantes principales:

(a) en calculant l'équation 9.12 du manuel (édition 2012) et

(b) à l'aide de la fonction `cor()` de R: `cor(mat, res$F)` ou `cor(mat, res$G)` où mat est votre matrice de données et res est le fichier de résultats de votre fonction d'ACP.

Les deux méthodes de calcul devraient produire les mêmes résultats. Comparez vos résultats numériques avec l'interprétation graphique.

4.8. Comparez les représentations obtenues aux points 4.3 (cadrage de type 1) et 4.4 (cadrage de type 2) ci-dessus, en calculant l'angle entre les descripteurs et les axes principaux pour chacune de ces représentations. La façon de faire est décrite aux pages 436 et 440 du chapitre 9 (édition 2012).

Notez que, comme l'indique le fichier de documentation, la fonction `acos()` de R produit des angles en radians.

Conservez suffisamment de décimales au cours de vos calculs pour assurer une bonne précision jusqu'à la fin des calculs.