

Exercice, chapitre 8b

Associations d'espèces

Février 2005, fév. 2006, oct. 2006, mars 2007, oct. 2007, oct. 2008, fév. 2009, fév. 2010

Cet exercice consiste à rechercher des associations significatives d'espèces. Vous utiliserez le fichier de données "Spiders_28x12.spe" (Aart and Smeenk-Enserink, 1975) fourni dans le dossier de travaux pratiques en R. Ce fichier contient les abondances de 12 espèces d'araignées récoltées à 28 sites dans une zone de dunes au nord de La Haye aux Pays-Bas.

De la page <http://www.bio.umontreal.ca/legendre/reprints/>, téléchargez l'article suivant qui décrit la méthode :

Legendre, P. 2005. Species associations: the Kendall coefficient of concordance revisited. *Journal of Agricultural, Biological, and Environmental Statistics* 10: 226-245.

1. Recherche des associations d'espèces

Voici les étapes à réaliser à l'aide du langage R.

1.1. Groupement des espèces significatives

- Importer le fichier de données des araignées dans votre console R.
- Produisez une transformation de Hellinger de cette matrice à l'aide de la fonction *decostand()* de VEGAN. Voir la section 4 du document *Travaux_pratiques_en_R.pdf*.
- Après la transformation de Hellinger, centrez et réduisez les données par espèce (colonne) à l'aide de la fonction *decostand()* de VEGAN.
- Recherche des groupes d'espèces par partitionnement en K centroïdes (méthode (c) à la p. 239 de l'article): transposez la matrice , puis calculez un partitionnement des espèces par la méthode des K centroïdes, à l'aide de la fonction *cascadeKM()* de la bibliothèque VEGAN, pour $K = 2$ à $(p/2)$ groupes; p est le nombre d'espèces. Quelle partition maximise le critère de Calinski-Harabasz? Quelle partition maximise le critère SSI? Doit-on à toute force suivre la recommandation de ce critère ou ne s'agit-il que d'une suggestion?
- Groupement agglomératif (méthode (d) à la p. 239 de l'article): à l'aide de la fonction *cor()*, calculez une matrice de corrélations de Pearson ou de Spearman entre les espèces ayant subi la transformation de Hellinger. Convertissez la matrice des corrélations entre espèces, dont le nom est par exemple 'araignees.cor', en une matrice de distances de type 'dist' par la commande *as.dist(1-araignees.cor)*. À l'aide de la fonction *agnes()*, réalisez un groupement agglomératif de Ward sur cette matrice de distances. Des exemples sont fournis à la section 12.2 des *Travaux_pratiques_en_R.pdf*. Combien de groupes d'espèces trouvez-vous? Utilisez la fonction *cutree()* pour extraire le vecteur des groupes d'espèces.

-
- Les partitions en 2 et 3 groupes obtenus à l'aide des deux méthodes (K centroïdes et groupement agglomératif de Ward) sont-elles les mêmes?

1.2. Identification des espèces significatives: bibliothèque KENDALL.W

- Utilisez la fonction *kendall.global()* de la bibliothèque VEGAN pour réaliser les tests globaux sur les groupes d'espèces que vous aurez identifiés à la section précédente pour les données ayant subi la transformation de Hellinger. Faites le travail pour les partitions en 2 groupes et en 3 groupes. Les détails d'utilisation se trouvent dans le fichier d'aide **?kendall.global**. Demandez 9999 permutations pour les tests statistiques. Les tests globaux sont-ils significatifs au niveau $\alpha = 0.05$ (probabilité corrigée) ?
- Utilisez la fonction *kendall.post()* pour réaliser les tests *a posteriori* de la contribution des espèces à la concordance de leur groupe. Les détails d'utilisation se trouvent dans le fichier d'aide **?kendall.post**. Demandez d'abord 999 permutations pour les tests statistiques; vous connaîtrez ainsi le temps d'attente. Y a-t-il des valeurs de la statistique *Spearman.mean* qui sont négatives? Selon l'article, que devez-vous faire lorsque cela se produit ? Y a-t-il des espèces qui ne sont pas significatives dans une partition et qui le deviennent dans l'autre?

Avant de rédiger votre rapport, répétez les analyses avec 9999 permutations; cela vous permettra d'obtenir encore des résultats significatifs après la correction pour tests multiples.

2. Rapport

- Présentez vos résultats dans votre rapport et discutez-les. Quelle partition (en 2 ou 3 groupes) est, écologiquement, la plus informative?