

Master Biologie Intégrative et Physiologie 2011-2012

Biostatistiques avancées

Responsable du cours : Yves Desdevises

Travaux Dirigés n° 7

Corrigé

Exercice 1

La matrice suivante présente des distances génétiques entre plusieurs espèces de poissons de la famille des Sparidae, appartenant aux genres *Oblada* et *Diplodus*. Ces distances sont calculées à partir de la comparaison de séquences d'ADN mitochondrial 16S.

	<i>D. sargus</i>	<i>D. annularis</i>	<i>O. melanura</i>	<i>D. vulgaris</i>	<i>D. cervinus</i>	<i>D. puntazzo</i>
<i>D. sargus</i>	-					
<i>D. annularis</i>	0,048	-				
<i>O. melanura</i>	0,031	0,054	-			
<i>D. vulgaris</i>	0,015	0,038	0,038	-		
<i>D. cervinus</i>	0,025	0,061	0,044	0,036	-	
<i>D. puntazzo</i>	0,023	0,050	0,019	0,021	0,036	-

1. Construisez un arbre phylogénétique de ces espèces à l'aide de la méthode UPGMA.

Étapes du groupement :

Une nouvelle matrice doit être recalculée à chaque étape.

Premier groupe formé *D. sargus* – *D. vulgaris* (Ds/Dv) à 0,015.

	Ds/Dv	<i>D. annularis</i>	<i>O. melanura</i>	<i>D. cervinus</i>	<i>D. puntazzo</i>
Ds/Dv	-				
<i>D. annularis</i>	0,043	-			
<i>O. melanura</i>	0,035	0,054	-		
<i>D. cervinus</i>	0,031	0,061	0,044	-	
<i>D. puntazzo</i>	0,022	0,050	0,019	0,036	-

Deuxième groupe formé: *D. puntazzo* – *O. melanura* (Dp/Om) à 0,019.

Exemple de calcul : $0,043 = (0,048 + 0,038)/2$

	Ds/Dv	<i>D. annularis</i>	Dp/Om	<i>D. cervinus</i>
Ds/Dv	-			
<i>D. annularis</i>	0,043	-		
Dp/Om	0,028	0,052	-	
<i>D. cervinus</i>	0,031	0,061	0,040	-

Troisième groupe formé: Ds/Dv – Dp/Om (Ds/Dv/Dp/Om) à 0,028.
Exemple : 0,048 = (0,043 + 0,052)/2

	Ds/Dv/Dp/Om	<i>D. annularis</i>	<i>D. cervinus</i>
Ds/Dv/Dp/Om	-		
<i>D. annularis</i>	0,048	-	
<i>D. cervinus</i>	0,035	0,061	-

Quatrième groupe formé: Ds/Dv/Dp/Om – *D. cervinus* (Ds/Dv/Dp/Om/Dc) à 0,035

	Ds/Dv/Dp/Om/Dc	<i>D. annularis</i>
Ds/Dv/Dp/Om/Dc	-	
<i>D. annularis</i>	0,051	-

Dernier groupe formé Ds/Dv/Dp/Om/Dc – *D. annularis* à 0,051. Attention, ici la pondération des groupes entre en jeu. Chaque objet de départ "pèse" toujours 1, et ici un groupe de 4 fusionne avec un groupe de 1 (jusque là, seuls des groupes de tailles égales fusionnaient et la nouvelle distance était égale à la moyenne des distances originales). Dans ce cas le calcul donne :
 $((0,048 \times 4) + 0,061)/5 = 0,051$

Dendrogramme :

Distance	0,015	0,019	0,028	0,035	0,050
Similarité	0,985	0,981	0,972	0,965	0,950

```

D.sargus      !<-----
                !<-----
D.vulgaris    !<-----          !
                !<-----
O.melanura    !<-----          !
                !<-----          !
D.puntazzo    !<-----          !
                !<-----          !
D.cervinus    !<-----          !
                !<-----          !
D.annularis  !<-----          !

```

2. D'après ce résultat, que pensez-vous du statut taxonomique du genre *Oblada*?

Ce genre est niché au sein du genre *Diplodus*. Son statut taxonomique n'est apparemment pas correct. Le genre *Oblada* devrait être inclus dans le genre *Diplodus*.

Exercice 2

Lors d'une campagne océanographique, des chercheurs ont prélevé 20 espèces de crustacés planctoniques (notées 1-20) à 10 stations du golfe du Lion (notées A-J). A chaque station, les abondances de chacune des espèces ont été relevées (données non montrées), et plusieurs descripteurs environnementaux ont été mesurés. On veut savoir comment se structure la communauté, et comprendre quels sont les facteurs écologiques responsables de cette structure.

Tableau des caractéristiques des stations

	Profondeur (m)	[Chl. a] ($\mu\text{g/l}$)	Température (°C)
Station A	7	3,21	18
Station B	61	2,12	16
Station C	32	0,32	17
Station D	18	4,76	18
Station E	73	1,73	16
Station F	5	1,12	15
Station G	27	4,34	14
Station H	20	0,56	17
Station I	16	0,86	16
Station J	25	2,56	16

1. Quel type d'analyse peut permettre d'appréhender ces questions ?

On peut faire une ordination en espace réduit. Comme il s'agit de la distribution d'espèces (relations unimodales avec l'environnement), il faut utiliser une analyse factorielle des correspondances.

Voici un diagramme d'ordination issu de l'analyse choisie (la distance entre les stations est préservée) :

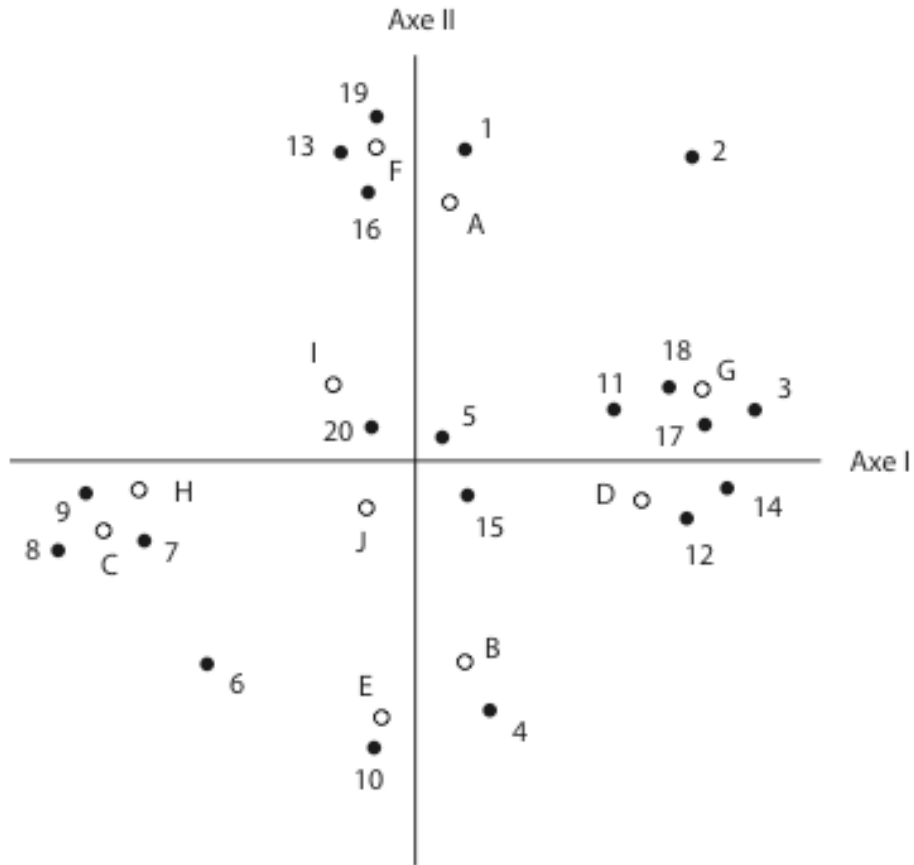


Tableau issu de l'analyse correspondant au diagramme (données partielles)

	$\lambda_1 = 0,42$	$\lambda_2 = 0,17$	$\lambda_3 = 0,14$	$\lambda_4 = 0,06$	$\Sigma\lambda = 1,18$
	Axe I	Axe II	Axe III	Axe IV	
Espèce 3	23	3	16	12	
Espèce 11	15	3	2	21	
Espèce 12	19	-4	20	16	
Espèce 14	21	-2	22	15	
Espèce 17	20	2	-1	19	
Espèce 18	18	4	-3	18	
Station A	2	12	12	7	
Station C	-15	-4	-4	-8	
Station E	-2	-10	7	15	
Station G	14	4	8	13	
Station J	-3	-2	1	0	

2. Que peut-on dire du plan représenté sur le diagramme ?

% de variation représentée par l'axe I : $0,42/(1,18) = 0,356$, soit 35,6 %

% de variation représentée par l'axe II : $0,17/(1,18) = 0,144$, soit 14,4 %

Le plan représenté contient donc $35,6 + 14,4 = 50$ % de la variance du nuage de points.

3. Pensez-vous que les espèces 3, 11, 12, 14, 17, et 18 forment bien un groupe dans l'espace multidimensionnel ?

Non, le plan I X II, dans lequel ces espèces forment un groupe, est une projection du nuage de points. Ces espèces sont clairement séparées sur l'axe III, comme en témoignent leurs coordonnées.

4. Quelles sont les caractéristiques majeures des axes I et II ? Comment les interpréter en référence aux variables environnementales mesurées ?

L'axe I représente la variation en chlorophylle a : les espèces vivant dans les milieux pauvres en chl a sont situées sur sa gauche, celles en nécessitant beaucoup se placent vers la droite.

L'axe II est principalement lié aux variations de profondeur: les stations de faible profondeur et les espèces associées sont situées en haut de l'axe, les espèces profondes étant en bas.

5. Que peut-on dire des préférences écologiques de chacune des espèces 1, 2, 3, 9, et 5 ?

1 : surface

2 : beaucoup de chl a et eau peu profonde (surface)

3 : beaucoup de chl a

5 : préférence pour des valeurs moyennes. Espèce ubiquiste ou exigeant des valeurs moyennes. Il faudrait connaître ses valeurs de tolérance sur les axes pour se prononcer.

9 : peu de chl a

6. Qu'est-ce qui caractérise la station J ?

Elle est au milieu du diagramme, donc elle possède des valeurs moyennes pour les variables considérées.

Exercice 3

Voici deux matrices de distances entre 5 stations (1 - 5). L'une est calculée sur la base des abondances de différentes espèces végétales, l'autre à partir de descripteurs environnementaux. Afin de savoir si les caractéristiques de l'environnement structurent les communautés de plantes, on désire calculer la corrélation entre ces matrices.

	2	3	4	5		2	3	4	5
1	0,17	0,38	0,42	0,39	1	0,24	0,32	0,39	0,21
2		0,72	0,44	0,12	2		0,41	0,82	0,21
3			0,35	0,41	3			0,33	0,35
4				0,03	4				0,09

1. Quel test statistique doit-on effectuer ?

Il s'agit d'un test de corrélation entre matrices de distances : test de Mantel.

2. Le tableau ci-dessous donne le résultat de l'analyse, réalisée par un logiciel. Interprétez-le.

$$\frac{\text{per} < |Z|}{974} \quad | \quad \frac{\text{per} = |Z|}{1} \quad | \quad \frac{\text{per} > |Z|}{25}$$

Le tableau donne la position de la statistique observée Z par rapport à la distribution aléatoire de la statistique sous H_0 . Cette distribution est générée par 999 (= 974 + 25) permutations. Le nombre de valeurs aléatoires de Z supérieures ou égales à la valeur observée est 25 + 1 = 26. La probabilité que H_0 soit vraie est donc : 26/1000 = 0,026. On peut donc rejeter H_0 (car 0,026 < 0,05) et supporter l'hypothèse qu'il y a une corrélation linéaire entre les deux matrices de distance.