

# *Exercices, chapitre 7*

## *Coefficients de similarité*

Octobre 2011

1) Pour les deux objets du tableau

	Descripteurs							
Objet 1	1	1	1	1	0	0	1	0
Objet 2	0	1	0	0	1	0	0	1

calculer les similarités suivantes:

Coefficient de simple concordance,  $S_1$

Coefficient de Jaccard,  $S_7$

Coefficient de Sørensen,  $S_8$

Vérifier vos calculs à l'aide de la fonction **dist.binary** de la bibliothèque **ade4** de **R**. Attention: la numérotation des coefficients dans l'article de Gower & Legendre (1986) n'est pas la même que dans le manuel. Pouvez-vous identifier ces 3 coefficients par leur formule et trouver comment les calculer? Information sur la fonction: `?dist.binary`

2) Pour les deux objets du tableau

	Descripteurs					
Objet 1	4	3	0	0	1	2
Objet 2	2	3	0	2	0	1

calculer les similarités suivantes:

Coefficient d'Estabrook-Rogers,  $S_{16}$ , avec  $k = 1$  pour tous les descripteurs.

Coefficient de Steinhaus,  $S_{17}$

3) 200 sites ont été étudiés par Alain Leduc dans la MRC du Haut Saint-Laurent. Les sites furent décrits à l'aide de plusieurs variables physiques et topographiques. Cinq de ces variables se trouvent dans le tableau ci-dessous; la valeur NA désigne l'absence d'information.

	Descripteurs				
Site 1	1	4	6	3	4
Site 2	NA	3	6	4	2
Site 3	5	4	6	6	8
Site 4	3	0	10	NA	1

Les cinq variables du tableau sont les suivantes :

1. Drainage: variable semi-quantitative. Traitez cette variable comme si elle était *quantitative*.
2. Pierrosité: variable semi-quantitative. Traitez cette variable comme si elle était *quantitative*.
3. Topographie: variable *qualitative*, classes 0 à 10 non ordonnées.
4. Exposition: variable partiellement quantitative. Les classes 1 à 8 correspondent aux 8 secteurs de la rose des vents; 9 signifie 'terrain plat'. Traitez cette variable comme si elle était *qualitative*.
5. Texture de l'horizon 1 du sol: variable *qualitative* à 8 classes.

- Calculer la matrice de similarités  $S_{15}$  entre les 4 objets (sites). Il y a 6 coefficients à calculer à la main.
- Vérifiez vos calculs en R à l'aide de *daisy()* (bibliothèque CLUSTER) ou de *gowdis()* (bibliothèque FD). Attention: les variables qualitatives doivent recevoir la classe "factor"; voir la commande *as.factor()* ainsi que la démonstration en classe. Présentez vos résultats sous la forme d'une matrice **S** ou **D**.

4) Calculer la similarité  $S_{16}$  entre les 4 objets (sites) de la matrice présentée au problème numéro 3. Il y a 6 coefficients à calculer à la main. Les variables de A. Leduc étaient structurées comme suit :

1. Drainage: variable *semi-quantitative* à 7 classes; utiliser  $k = 3$  dans la formule de  $S_{16}$ .
2. Pierrosité: variable *semi-quantitative*, classes 0 à 6; utiliser  $k = 2$  dans la formule de  $S_{16}$ .
3. Topographie: variable *qualitative*, classes 0 à 10 non ordonnées.
4. Exposition: variable *partiellement quantitative*. Les classes 1 à 8 correspondent aux 8 secteurs de la rose des vents; la classe 9 correspond à 'terrain plat'. Utiliser la matrice de similarités partielles suivante dans la formule de  $S_{16}$  :

	CI-1	CI-2	CI-3	CI-4	CI-5	CI-6	CI-7	CI-8	CI-9
<b>Classe 1</b>	----								
<b>Classe 2</b>	0.5	----							
<b>Classe 3</b>	0	0.5	----						
<b>Classe 4</b>	0	0	0.5	----					
<b>Classe 5</b>	0	0	0	0.5	----				
<b>Classe 6</b>	0	0	0	0	0.5	----			
<b>Classe 7</b>	0	0	0	0	0	0.5	----		
<b>Classe 8</b>	0.5	0	0	0	0	0	0.5	----	
<b>Classe 9</b>	0	0	0	0	0	0	0	0	----

5. Texture de l'horizon 1 du sol: variable *qualitative* à 8 classes. Utiliser la matrice de similarités partielles suivante dans la formule de  $S_{16}$  :

	CI-1	CI-2	CI-3	CI-4	CI-5	CI-6	CI-7	CI-8
<b>Classe 1</b>	----							
<b>Classe 2</b>	0.5	----						
<b>Classe 3</b>	0	0.5	----					
<b>Classe 4</b>	0	0.5	0.5	----				
<b>Classe 5</b>	0	0	0	0.5	----			
<b>Classe 6</b>	0	0	0.5	0.5	0.5	----		
<b>Classe 7</b>	0	0	0.5	0	0	0.5	----	
<b>Classe 8</b>	0	0	0	0	0	0	0.5	----

- Présentez vos résultats sous la forme d'une matrice de similarités  $S$ .