

Problème Pratique de Statistique - 31

Similarités le long de la rivière Willamette



<http://zebu.uoregon.edu/1996/es202/willamette.gif>

J. Van Sickle (1997), dans une étude sur un problème complexe, donne à titre d'illustrations une matrice de similarité. La liste des espèces de poissons est connue par une étude de 1992 en 19 stations de la rivière Willamette. Ces sites sont approximativement représentés sur la carte ci-dessus. On connaît la distance entre la station et la confluence avec la Columbia River :

$$d_0 = c(2, 6, 10, 14, 27, 35, 40, 47, 63, 77, 93, 124, 150, 182, 206, 232, 240, 283, 296)$$

On ne conserve du tableau faunistique que la matrice de similarité entre des sites, sous la forme de l'indice de Jaccard. Si C_{ij} est le nombre d'espèces présentes dans les deux sites i et j , si U_i (respectivement U_j) est le nombre d'espèces présentes seulement dans le site i (respectivement j), la similarité entre le site i et le site j vaut :

$$s_{ij} = \frac{C_{ij}}{C_{ij} + U_i + U_j}$$

(1 si les deux listes sont identiques, 0 si elles sont disjointes).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	0.55	0.50	0.40	0.40	0.27	0.31	0.25	0.25	0.27	0.25	0.14	0.20	0.27	0.21	0.07	0.11	0.12	0.19
2	0.55	1.00	0.78	0.67	0.50	0.50	0.50	0.45	0.33	0.43	0.33	0.21	0.27	0.33	0.29	0.21	0.29	0.25	0.30
3	0.50	0.78	1.00	0.86	0.62	0.62	0.45	0.40	0.40	0.38	0.40	0.25	0.31	0.38	0.33	0.25	0.25	0.29	0.26
4	0.40	0.67	0.86	1.00	0.71	0.71	0.50	0.44	0.44	0.42	0.44	0.27	0.33	0.31	0.25	0.27	0.27	0.21	0.21
5	0.40	0.50	0.62	0.71	1.00	0.71	0.67	0.30	0.30	0.42	0.30	0.17	0.23	0.21	0.15	0.27	0.27	0.21	0.21
6	0.27	0.50	0.62	0.71	0.71	1.00	0.50	0.44	0.44	0.42	0.44	0.27	0.33	0.21	0.25	0.40	0.36	0.31	0.28
7	0.31	0.50	0.45	0.50	0.67	0.50	1.00	0.45	0.45	0.67	0.45	0.21	0.36	0.33	0.29	0.31	0.47	0.33	0.37
8	0.25	0.45	0.40	0.44	0.30	0.44	0.45	1.00	0.56	0.50	0.75	0.50	0.42	0.38	0.45	0.25	0.33	0.29	0.33
9	0.25	0.33	0.40	0.44	0.30	0.44	0.45	0.56	1.00	0.64	0.75	0.36	0.55	0.38	0.45	0.25	0.33	0.29	0.40
10	0.27	0.43	0.38	0.42	0.42	0.42	0.67	0.50	0.64	1.00	0.50	0.27	0.40	0.38	0.33	0.27	0.41	0.29	0.40
11	0.25	0.33	0.40	0.44	0.30	0.44	0.45	0.75	0.75	0.50	1.00	0.50	0.55	0.50	0.60	0.36	0.33	0.38	0.33
12	0.14	0.21	0.25	0.27	0.17	0.27	0.21	0.50	0.36	0.27	0.50	1.00	0.64	0.73	0.70	0.33	0.50	0.46	0.39
13	0.20	0.27	0.31	0.33	0.23	0.33	0.36	0.42	0.55	0.40	0.55	0.64	1.00	0.62	0.73	0.39	0.53	0.50	0.50
14	0.27	0.33	0.38	0.31	0.21	0.21	0.33	0.38	0.38	0.38	0.50	0.73	0.62	1.00	0.82	0.36	0.50	0.57	0.47
15	0.21	0.29	0.33	0.25	0.15	0.25	0.29	0.45	0.45	0.33	0.60	0.70	0.73	0.82	1.00	0.42	0.47	0.67	0.53
16	0.07	0.21	0.25	0.27	0.27	0.40	0.31	0.25	0.25	0.27	0.36	0.33	0.39	0.36	0.42	1.00	0.62	0.58	0.47
17	0.11	0.29	0.25	0.27	0.27	0.36	0.47	0.33	0.33	0.41	0.33	0.50	0.53	0.50	0.47	0.62	1.00	0.60	0.67
18	0.12	0.25	0.29	0.21	0.21	0.31	0.33	0.29	0.29	0.29	0.38	0.46	0.50	0.57	0.67	0.58	0.60	1.00	0.65
19	0.19	0.30	0.26	0.21	0.21	0.28	0.37	0.33	0.40	0.40	0.33	0.39	0.50	0.47	0.53	0.47	0.67	0.65	1.00

Cette information suffit-elle à caractériser l'évolution du peuplement de poissons le long de la rivière ?

Van Sickle, J. 1997. Using mean similarity dendrograms to evaluate classifications. *Journal of Agricultural, Biological, and Environmental Statistics* 4:370-388.

Pour saisir les données :

```
aa=c(0.55,0.50,0.40,0.40,0.27,0.31,0.25,0.25,0.27,0.25,0.14,0.20,0.27,0.21,0.07,0.11)
aa=c(aa,0.12,0.19,0.78,0.67,0.50,0.50,0.50,0.45,0.33,0.43,0.33,0.21,0.27,0.33,0.29,0.21)
aa=c(aa,0.29,0.25,0.30,0.86,0.62,0.62,0.45,0.40,0.40,0.38,0.40,0.25,0.31,0.38,0.33,0.25)
aa=c(aa,0.25,0.29,0.26,0.71,0.71,0.50,0.44,0.44,0.42,0.44,0.27,0.33,0.31,0.25,0.27,0.27)
aa=c(aa,0.21,0.21,0.71,0.67,0.30,0.30,0.42,0.30,0.17,0.23,0.21,0.15,0.27,0.27,0.21,0.21)
aa=c(aa,0.50,0.44,0.44,0.42,0.44,0.27,0.33,0.21,0.25,0.40,0.36,0.31,0.28,0.45,0.45,0.67)
aa=c(aa,0.45,0.21,0.36,0.33,0.29,0.31,0.47,0.33,0.37,0.56,0.50,0.75,0.50,0.42,0.38,0.45)
aa=c(aa,0.25,0.33,0.29,0.33,0.64,0.75,0.36,0.55,0.38,0.45,0.25,0.33,0.29,0.40,0.50,0.27)
aa=c(aa,0.40,0.38,0.33,0.27,0.41,0.29,0.40,0.50,0.55,0.50,0.60,0.36,0.33,0.38,0.33,0.64)
aa=c(aa,0.73,0.70,0.33,0.50,0.46,0.39,0.62,0.73,0.39,0.53,0.50,0.50,0.82,0.36,0.50,0.57)
aa=c(aa,0.47,0.42,0.47,0.67,0.53,0.62,0.58,0.47,0.60,0.67,0.65)
```