Distances non-paramétriques entre profils

Raymond Van den Driessche et Ana García Gómez

Ir. Responsable de la Banque de Données Pédologiques de l'ORSTOM, 70, route d'Aulnay, F-93-Bondy.

Dr Sc. 6, rue Rabelais, F-93-Bondy.

Summary

Soil scientists familiar with storage and retrieval of field data are aware of the need to manipulate simultaneously ordinal and interval variables.
A distribution-free measure of distances between profiles using ranks is exemplified. The distances are measured on the basis of an incomplete matrix of m rows or profiles and v columns or variables. The column entries are the same variables taken in turn in all the horizons named A1, A3, B, BC, C, C1, C2.
In the numerical example from São Tomé isl. Afr. the matrix 28 x 98 consists of 28 profiles and 14 variables (hue, value, chroma, clay, silt, fine sand, coarse sand, organic matter, C/N, pH, exchangeable calcium, sum of the cations, cation exchange capacity, saturation level) in several of the 7 horizons. Therefore, the data matrix is only half filled with data, and is not restricted to field data.
Punched output in 8F9.4 of the lower half of the distance matrix is followed by a printed symbolic output in condensed form so as to allow for $m \leq 127$ ($v \leq 2000$ and $mv \leq 93000$) on the 1108 computer.

Introduction

Lors du traitement de données descriptives des horizons ou de l'environnement du profil, nous ressentons le besoin de disposer d'une mesure de distance non-paramétrique. Les hypothèses d'homoscédasticité et de normalité faites, si souvent, pour les variables de laboratoire ne sont plus de mise dans ce domaine par excellence des variables qualitatives. Bien que la méthode de rangs que nous présentons pour cette mesure n'ait pas encore fait, à notre connaissance, l'objet d'applications, nous la couplons à une mesure originale de distance entre profils entiers. Les définitions indispensables sont rappelées et la méthode est appliquée à un lot de 28 profils de St Tomé.

Rappel de définitions

Le GROUPE est un ensemble de données observées de différente nature. Il y a une seule donnée par variable. Constituent le plus souvent un groupe, toutes les données observées dans un des horizons d'un profil. Peut aussi former un groupe, la totalité des données observées dans un ensemble de profils classés de la même façon dans la classification des sols. Nous sommes d'avis de former le groupe avec toutes les données observées dans la totalité des horizons d'un même profil, mais en respectant l'identité de chaque horizon.

Opposée au groupe, la VARIABLE est un ensemble de données observables, de même nature, exprimées dans la même unité ou vocable normalisé, mais dont une seule est observée dans un même groupe. La variable "test de plasticité" est utilisée quand l'horizon est humide ou très humide; elle se matérialise par l'aptitude que possède le matériau à subir un changement continu de forme. Pour cette variable, prise à titre d'exemple, quatre données sont observables : non plastique, peu plastique, plastique, très plastique. Second exemple tiré de la description des horizons, la variable "nature des croûtes et/ou efflorescences éventuelles" offre à l'observateur quatre données : chlorurées, bicarbonatées, carbonatées, sulfatées, dont une seule est observée.

Une DONNEE, et seulement une, exprime donc l'état d'une variable dans un groupe déterminé. Les données sont exprimées en langage naturel ou en code.

Le MANQUANT est une donnée qui fait défaut et cela pour des motifs liés aussi bien à la nature même des données qu'à l'observateur.

Les IDENTIQUES sont les données d'une même variable qui ont la même expression, en langage naturel ou en code, dans deux ou plusieurs groupes. Des données ayant la même expression mais dépendant de variables différentes ne sont pas des identiques.

Le TABLEAU est un assemblage ordonné de groupes, de variables, de données et de manquants. Il y a autant de lignes que de groupes, autant de colonnes que de variables.

Exposé de la méthode

Considérons le TABLEAU ci-après. Il est constitué de données en provenance de St Tomé et nous servira d'application numérique.

L'île de St Tomé est située entre les longitudes E 6-45-0 et 6-28-19 et entre les latitudes N 0-24-30 et 0-0-0. Elle a été étudiée par les pédologues Cardoso & Garcia (1962).

Les auteurs ont donné l'identité des horizons décrits, en utilisant entre autres les symboles A1, A3, B, BC, C, C1, C2. Cela nous a permis d'extraire de leur mémoire les 28 profils nº 44A, 93, 147, 204, 125, 33A, 82, 88, 157, 212, 209, 41, 38, 111, 131, 6, 47, 56, 28A, 29, 45, 207, 23, 36A, 37, 94, 100, 69. Ces profils sont représentatifs des sols "paraferralíticos" et "fersialíticos tropicais" de l'île. L'application numérique porte donc sur 28 groupes.

Les variables retenues par nous sont, dans l'ordre, la teinte, la valeur et l'intensité Munsell de l'horizon, l'argile, le limon, le sable fin, le sable grossier, la matière organique, le rapport C/N, le pH, le calcium échangeable, la somme des cations échangeables, la capacité d'échange, le taux de saturation. L'originalité de notre approche réside dans le fait de considérer une variable déterminée (limon par exemple) dans sept horizons différents (A1, A3, B, BC, C, C1, C2) comme autant de variables différentes. Tout reste individualisé dans notre tableau : A1 et A3 ; B et BC ou BC et C ; C, C1 et C2 ne sont pas confondus.

Le GROUPE correspondant à un profil est, par conséquent, un ensemble de DONNEES provenant des VARIABLES des horizons décrits et de MANQUANTS. Dans cette optique, les manquants correspondent, à la fois, aux variables non décrites dans un horizon quelconque et aux variables des horizons n'existant pas dans le profil qui est décrit mais bien dans d'autres profils de l'île. Nous désignons les manquants par le code -1 .

L'effectif d'une variable est le nombre de DONNEES de cette variable sur l'ensemble des groupes. Si m désigne le nombre de groupes, m_i est l'effectif de la variable i . L'effectif de la première variable est 28 ; celui de la dernière est 3 .

Il nous reste à souligner quelques identiques dans ce tableau. De la variable 1, la donnée 15 est présente cinq fois, la donnée 17,5 deux fois, la donnée 20 vingt fois; ce qui fait trois lots d'identiques.

Tableau

Profil		Horizon	Teinte Munsell	Valeur Munsell	Intensité Munsell	Argile	Limon	Sable fin	Sable grossier	Matière organique	Rapport C/N	pH	Ca échangeable	Somme des cations	Capacité d'échange	Taux de saturation
44A	1	A1	15	3.5	4	72.2	18.6	4.8	4.4	4.74	8.5	4.4	4.45	7.70	14.90	51.7
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		B	15	3.5	6	86.3	5.7	5.6	2.4	1.68	6.8	4.5	2.45	5.50	10.70	51.4
		BC	15	4	6	82.9	7.4	5.2	4.5	-1	-1	4.7	2.73	4.10	8.90	46.1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
93	2	A1	15	3	2	30.5	24.1	27.0	18.4	3.82	11.1	5.9	5.12	9.48	16.98	55.8
		A3	15	3	2.5	55.1	24.4	1.4	19.1	1.49	7.8	5.9	2.55	4.75	12.95	36.7
		B	15	3	2.5	73.0	21.7	2.1	3.2	-1	-1	5.9	1.60	3.18	8.82	36.1
		BC	15	3	3.5	75.4	21.1	2.5	1.0	-1	-1	5.6	0.85	1.78	7.28	24.4
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
147	3	A1	17.5	4	4	51.2	34.9	2.4	11.5	4.67	8.5	4.7	6.53	9.37	20.87	44.9
		A3	17.5	4	4	60.5	26.8	12.5	1.1	2.00	9.7	5.0	1.21	2.30	11.60	19.8
		B	15	4	8	62.6	27.0	8.1	2.3	-1	-1	5.1	1.84	2.78	16.58	16.8
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		C1	15	4	6	51.5	27.5	17.5	3.5	-1	-1	5.1	1.16	3.17	19.42	16.3
		C2	15	4	6	36.5	26.8	31.4	5.3	-1	-1	4.9	0.94	2.22	18.02	12.3
204	4	A1	15	3.5	3	68.2	22.2	4.7	4.9	5.00	10.7	5.0	6.56	13.17	23.82	55.1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		B	12.5	3	5	73.0	24.0	2.5	0.5	-1	-1	5.6	6.06	6.56	11.71	56.0
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		C1	12.5	3.5	6	69.2	17.1	9.6	4.1	-1	-1	5.8	5.72	6.24	12.04	51.9
		C2	12.5	4	6	73.8	10.5	13.7	2.0	-1	-1	6.2	3.70	4.19	10.77	38.8
125	5	A1	12.5	3	3	9.4	19.3	58.0	13.3	5.57	10.1	6.2	16.50	21.46	26.56	80.8
		A3	12.5	3	5	7.9	10.3	63.0	18.8	4.74	9.4	5.7	9.10	10.34	19.14	54.0
		B	12.5	3	6	2.7	9.6	45.0	43.7	-1	-1	5.6	1.34	2.87	12.24	23.4
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		C	12.5	3	6	8.8	19.5	46.5	25.2	-1	-1	6.0	0.56	1.97	10.17	19.4
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
33A	6	A1	20	4	3	46.6	26.0	13.1	14.3	5.69	10.6	4.1	6.62	9.29	19.99	46.5
		A3	20	4	3	43.2	31.3	18.3	7.2	2.03	6.8	4.5	3.31	4.77	15.97	30.0
		B	20	4.5	3.5	41.9	32.8	20.4	4.9	-1	-1	4.5	3.30	4.65	14.35	32.6
		BC	20	5	4	29.5	35.0	31.4	4.1	-1	-1	4.5	3.31	4.45	15.15	29.6
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
82	7	A1	20	4.5	5	30.8	24.6	21.4	23.2	3.48	8.1	4.9	0.53	1.67	10.83	15.4
		A3	20	5	7	22.5	7.6	50.6	19.3	2.48	10.3	5.4	1.23	1.91	6.67	28.6
		B	20	5	8	40.5	9.8	31.4	18.3	-1	-1	5.6	-1	-1	-1	-1
		BC	20	5	8	60.0	12.6	2.9	24.5	-1	-1	5.4	0.50	0.90	4.28	21.0
		C	-1	-1	-1	34.5	28.6	30.8	6.1	-1	-1	5.1	0.86	1.24	4.18	29.7
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
88	8	A1	20	4	3.5	28.9	27.6	13.5	30.0	4.30	11.9	4.8	1.12	5.35	12.13	44.0
		A3	20	4	4	45.8	19.9	13.9	20.4	0.60	5.8	5.5	1.45	3.65	7.88	46.3
		B	20	4.5	5	53.4	20.9	1.3	24.4	-1	-1	5.4	1.12	3.45	7.05	49.0
		BC	20	5	8	34.9	21.6	14.8	28.7	-1	-1	5.7	3.76	6.12	10.09	60.6
		C	-1	-1	-1	29.7	40.7	21.7	7.9	-1	-1	5.8	2.87	3.30	7.54	43.8
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Tableau (suite)

Profil	Horizon	Teinte Munsell	Valeur Munsell	Intensité Munsell	Argile	Limon	Sable fin	Sable grossier	Matière organique	Rapport C/N	pH	Ca échangeable	Somme des cations	Capacité d'échange	Taux de saturation
157 9	A1	20	4	2.5	52.1	15.5	29.1	3.3	2.30	6.3	4.7	2.14	6.98	16.82	41.5
	A3	20	4	3	63.0	17.0	9.2	10.8	2.32	7.1	4.7	0.73	2.27	11.93	19.0
	B	20	4	3.5	80.0	15.5	3.3	1.2	-1	-1	5.3	1.83	2.11	11.41	18.5
	BC	20	4	4	83.5	12.7	2.1	2.7	-1	-1	5.1	0.83	1.27	8.61	14.8
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
21210	A1	20	3.5	2	19.3	18.8	12.9	49.0	3.11	9.5	5.9	7.84	11.33	18.23	62.2
	A3	17.5	5	3	30.0	20.5	14.9	34.6	1.28	5.7	5.9	2.24	4.12	7.61	54.0
	B	17.5	5	7	57.4	18.0	0.6	24.0	-1	-1	5.5	3.30	5.49	11.99	45.7
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	C	17.5	5.5	8	54.2	13.9	10.6	21.3	-1	-1	5.4	2.72	5.13	19.13	26.8
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
20911	A1	20	3.5	2	34.4	18.5	20.8	26.3	12.60	14.6	5.6	8.40	11.65	21.90	53.2
	A3	20	4	2.5	35.6	17.5	21.9	25.0	7.90	13.1	5.7	6.20	8.72	16.52	52.8
	B	17.5	5	7	76.2	13.3	6.5	4.0	-1	-1	5.8	3.42	4.44	7.87	56.4
	BC	17.5	5.5	7	79.4	12.4	3.3	4.9	-1	-1	5.4	3.34	3.64	6.29	57.9
	C	-1	-1	-1	66.2	15.8	8.6	9.4	-1	-1	5.5	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
41 12	A1	20	3	2	8.2	7.4	68.0	16.4	9.32	10.2	5.5	0.52	2.77	13.27	79.2
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	B	20	3	4	1.7	2.8	63.0	32.5	-1	-1	6.1	-1	-1	-1	-1
	BC	20	4	4	13.4	12.8	62.4	11.4	-1	-1	6.2	3.15	7.71	20.01	38.5
	C	12.5	2	4	6.4	6.6	21.2	65.8	-1	-1	6.2	1.73	1.93	5.28	36.5
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
38 13	A1	20	3	2	31.2	24.6	27.6	16.6	4.82	14.0	5.5	0.75	2.49	10.69	23.3
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	B	20	3.5	2	54.7	16.2	28.7	0.4	-1	-1	5.3	1.30	3.07	13.07	23.5
	BC	20	3.5	3	82.0	15.3	1.9	0.8	-1	-1	5.1	1.05	3.88	11.98	32.4
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
11114	A1	20	3	2	39.0	20.3	24.5	15.7	3.80	10.0	4.8	1.78	3.78	15.18	24.9
	A3	20	3	2	55.2	31.6	1.3	11.9	2.21	9.1	4.7	1.64	2.97	12.47	23.7
	B	20	3.5	2	68.4	21.1	8.6	1.9	-1	-1	5.2	1.88	3.86	13.36	28.8
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	C	20	4	2	56.0	24.6	6.0	13.4	-1	-1	5.8	1.67	3.63	13.09	27.7
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
13115	A1	20	4	2	31.4	25.4	31.9	11.3	3.20	11.6	4.9	7.82	8.69	10.72	81.1
	A3	20	4	2	44.3	21.5	31.2	3.0	2.30	8.4	5.1	0.55	0.91	6.47	14.1
	B	20	3.5	2.5	44.1	17.8	17.1	21.0	0.95	6.9	5.3	0.69	1.12	7.29	15.4
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	C1	20	4	2.5	32.3	21.2	27.5	19.0	-1	-1	5.5	1.15	1.50	6.66	22.5
	C2	20	4	3.5	16.5	22.3	39.4	21.8	-1	-1	5.7	0.61	0.94	5.89	15.9
6 16	A1	20	2.5	2	38.7	38.6	14.4	8.3	10.10	13.7	6.1	10.60	15.29	22.69	67.4
	A3	20	3.5	2	49.9	34.8	13.4	1.9	2.72	11.3	5.8	2.58	4.43	10.93	40.6
	B	20	4	2	68.0	18.7	12.9	0.4	-1	-1	5.5	0.53	2.05	7.05	29.0
	BC	20	4	3	79.7	10.1	10.1	0.1	-1	-1	4.9	0.54	1.89	7.34	25.7
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Tableau (suite)

Profil	Horizon	Teinte Munsell	Valeur Munsell	Intensité Munsell	Argile	Limon	Sable fin	Sable grossier	Matière organique	Rapport C/N	pH	Ca échangeable	Somme des cations	Capacité d'échange	Taux de saturation
47	17 A1	15	3	2	62.0	31.0	5.3	1.7	4.54	9.8	6.5	13.35	19.35	26.35	73.6
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	B	15	3	3	46.2	43.6	9.5	0.7	-1	-1	5.9	13.35	17.98	24.14	74.8
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	C	15	3	3.5	69.6	26.6	2.2	1.6	-1	-1	5.9	12.00	15.91	21.11	75.8
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
56	18 A1	15	3	3	49.5	34.4	15.7	0.41	0.90	11.9	4.7	11.40	14.36	24.76	58.0
	A3	15	3	4	62.8	25.2	11.7	0.3	4.10	9.5	4.7	2.65	4.78	12.28	38.9
	B	12.5	3	6	73.2	17.9	2.6	6.3	-1	-1	5.5	0.84	2.30	6.40	35.9
	BC	12.5	3	6	44.1	21.7	34.1	0.1	-1	-1	5.7	0.87	1.14	5.54	20.6
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
28	A19 A1	20	2.5	2	41.3	44.4	11.7	2.6	2.92	8.5	5.8	10.55	14.74	18.44	79.9
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	B	20	3	2.5	71.0	20.6	7.1	1.3	-1	-1	5.4	7.62	11.37	14.67	78.1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	C	20	3	3.5	66.0	22.0	10.5	1.5	-1	-1	5.9	4.75	8.28	12.48	66.4
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
29	20 A1	20	3	2	34.8	29.8	18.8	16.6	5.50	10.6	5.6	7.23	10.14	15.25	66.2
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	B	20	3	4	58.2	18.6	22.0	1.2	-1	-1	5.4	0.72	2.48	5.53	44.8
	BC	20	4	4	67.4	18.0	13.8	0.8	-1	-1	5.2	0.42	1.82	4.07	44.7
	C	20	4	6	70.0	15.4	13.8	0.8	-1	-1	5.5	0.41	2.59	4.79	54.2
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
45	21 A1	20	3	2	38.8	29.4	29.9	1.9	3.10	7.8	5.7	4.57	8.15	14.65	55.6
	A3	20	4	2	80.6	15.6	2.7	1.1	1.36	6.6	5.6	2.45	5.88	14.87	39.5
	B	20	4	3.5	75.0	24.0	0.6	0.4	-1	-1	5.4	1.25	4.23	11.43	37.1
	BC	20	4	4	65.0	21.8	11.9	1.3	-1	-1	5.0	0.76	1.90	8.90	21.4
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
20	722 A1	20	3	2	26.4	23.7	36.0	13.9	6.76	9.6	5.4	11.15	13.79	22.24	62.0
	A3	20	4	2	60.0	22.9	3.4	13.7	1.74	4.0	5.9	6.42	9.26	13.46	61.4
	B	20	4.5	5	74.6	19.3	0.4	5.7	-1	-1	5.6	6.30	8.21	13.35	61.2
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	C	15	4.5	6	73.0	11.8	5.8	9.7	-1	-1	5.7	5.30	6.32	10.32	61.3
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
23	23 A1	20	3	1	60.6	31.2	5.2	3.0	1.38	5.5	6.4	7.50	12.99	16.59	78.4
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	B	20	3	2	78.0	19.3	1.4	1.3	-1	-1	6.1	4.70	9.11	12.73	71.6
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	C	20	4	3	78.0	19.2	1.7	1.1	-1	-1	6.2	2.46	8.02	11.72	68.2
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
36A	24 A1	17.5	3	2	20.5	37.9	26.2	15.4	6.64	7.9	5.0	10.94	16.60	26.80	61.9
	A3	17.5	3	3	21.3	35.0	20.0	23.6	3.04	7.7	5.0	8.34	12.70	22.40	56.7
	B	17.5	3.5	4	46.3	25.6	16.2	11.9	-1	-1	5.0	8.91	13.20	20.90	63.2
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Tableau (suite)

Profil		Horizon	Teinte Munsell	Valeur Munsell	Intensité Munsell	Argile	Limon	Sable fin	Sable grossier	Matière organique	Rapport C/N	pH	Ca échangeable	Somme des cations	Capacité d'échange	Taux de saturation
37	25	A1	20	3	2	24.9	26.9	38.2	10.0	4.15	8.9	6.3	17.60	23.18	30.11	77.0
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		B	20	3	3	66.4	22.2	5.6	5.8	-1	-1	6.6	11.40	15.55	22.03	70.8
		BC	20	3.5	3	68.6	18.5	12.0	0.9	-1	-1	7.0	7.70	9.95	15.10	65.8
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
94	26	A1	20	4	2	60.2	26.6	11.2	2.0	3.41	6.4	6.1	11.40	15.13	20.13	75.0
		A3	20	4	2.5	86.2	8.5	6.5	0.8	1.36	9.9	5.9	7.60	10.63	15.63	68.0
		B	20	4	3	78.0	17.6	2.9	1.5	-1	-1	6.3	5.72	8.19	13.99	58.6
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		C	20	5	5	64.6	15.8	12.8	6.8	-1	-1	6.2	1.44	4.88	9.52	51.2
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
100	27	A1	20	3.5	1	62.0	31.2	1.1	5.7	2.29	8.3	6.7	14.15	18.44	22.10	83.5
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		B	20	3	1.5	73.5	15.5	3.2	7.8	-1	-1	6.9	-1	-1	-1	-1
		BC	20	4	2.5	61.6	14.6	4.2	19.6	0.25	7.5	6.2	7.34	16.36	20.56	79.4
		C	20	4.5	3	46.0	25.0	16.3	12.7	-1	-1	6.4	8.04	11.54	17.48	66.0
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
69	28	A1	20	3	2	19.5	29.2	19.3	32.0	8.50	10.4	5.5	8.30	11.32	20.22	56.0
		A3	20	3.5	2	37.3	28.0	19.6	15.1	3.63	9.2	5.3	1.52	3.35	10.65	31.5
		B	20	4	2.5	59.0	28.0	11.5	1.5	-1	-1	5.3	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		C	20	4	2.5	55.2	30.8	4.9	9.1	-1	-1	5.5	0.52	2.18	7.08	30.8
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
			-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Les RANGS r_{ki} sont des numéros d'ordre 1 à m_i affectés aux m_i données x_{ki} d'une variable i . Pour les e_i lots d'identiques de la variable i l'affectation des rangs suit la règle suivante : les t_{qi} identiques du même lot q sont remplacés par la moyenne arithmétique des t_{qi} rangs suivants non encore utilisés.

L'égalité entre le rang maximal et l'effectif n'est vérifiée que lorsque la donnée maximale est seule. Pour une donnée maximale en t_{qi} identiques, le rang maximal est égal à $m_i - (t_{qi} / 2) + 0,5$

Premier exemple :

Les données de la variable 98 = taux de saturation dans l'horizon C2 sont remplacées par les rangs

k	$x_{k\,98}$	$r_{k\,98}$
3	12,3	1
4	38,8	3
15	15,9	2

$e_{98} = 0$ il n'y a pas d'identiques

$m_{98} = 3$ l'effectif est de 3 données

Deuxième exemple :

Pour la variable 43 = teinte Munsell de l'horizon BC
nous remplaçons
dans les groupes k les données $x_{k\,43}$ par les rangs $r_{k\,43}$

1	15	2,5
2	15	2,5
6	20	10
7	20	10
8	20	10
9	20	10
11	17,5	4
12	20	10
13	20	10
16	20	10
18	12,5	1
20	20	10
21	20	10
25	20	10
27	20	10

L'effectif $m_{43} = 15$. La donnée minimale 12,5 est remplacée par le rang 1. Les deux identiques 15 du lot 1 sont remplacés par la moyenne arithmétique des deux rangs non encore utilisés (2+3)/2= 2,5 . La donnée suivante 17,5 est présente une seule fois, le rang 4 la remplace. Les onze identiques 20 sont remplacés par la moyenne arithmétique des onze rangs non encore utilisés, c'est-à-dire par

(5+6+7+8+9+10+11+12+13+14+15)/11=10 .

Pour mémoire, $e_{43} = 2$; $t_{1\,43} = 2$; $t_{2\,43} = 11$.

La DISTANCE est une mesure de dissemblance entre groupes. Elle est comprise entre 0 et 1 , est indépendante de la fonction de répartition des variables et est proche de celle de Kendall & Stuart (1966).

$$D_{hk} = \frac{1}{v} \sum_{i=1}^{v} \frac{(r_{hi} - r_{ki})^2}{I_i}$$

mesure la distance entre le groupe h et le groupe k .

La correction

$$I_i = \left[m_i^3 - m_i - \sum_{q=1}^{e_i} (t_{qi}^3 - t_{qi}) \right] / 12$$

est calculée pour chaque variable à partir de :

m_i effectif de la variable i

e_i nombre de lots d'identiques pour la variable i

t_{qi} nombre d'identiques dans le lot q de la variable i

Pour toute distance calculée entre deux groupes, c'est pour chacune des variables le carré de la différence entre rangs homologues, divisé par la correction, qui seul intervient. La correction prend en compte, à la fois, l'effectif de la variable, le nombre de lots d'identiques et le nombre d'identiques par lot. Elle est nulle et la variable abandonnée en conséquence lorsque $t_{qi} = m_i$ c'est-à-dire quand les données sont toutes identiques ou quand il n'y a qu'une seule donnée.

Résultats de l'application numérique

Le programme KAV écrit pour nous par A.M. Aubry de la Banque de Données Pédologiques et transposé sur 1108 par F. Savary de STAD a permis de calculer en 12 secondes les 378 distances que voici. C'est dans ce format 8F9.4 que la moitié inférieure de la matrice d'ordre 28 est disponible pour un traitement ultérieur.

Profils	1	2	3	4	5	6	7	8
	9	10	11	12	13	14	15	16
	17	18	19	20	21	22	23	24
	25	26	27					
2	.0844							
3	.0478	.0772						
4	.0371	.0435	.4425					
5	.1115	.0857	.1324	.0926				
6	.0950	.1110	.0449	.0595	.1177			
7	.1053	.1370	.0759	.1147	.1387	.1096		
8	.0964	.1013	.0666	.0791	.1309	.0574	.0761	
9	.0608	.0814	.0530	.0763	.1608	.0907	.0716	.0997
10	.0733	.0598	.0895	.0691	.1178	.0817	.1014	.0573
10	.0943							
11	.0583	.0911	.1096	.0547	.0877	.0960	.0966	.0921
11	.1053	.0663						
12	.1355	.1128	.1348	.1296	.0658	.0903	.1222	.0834
12	.1397	.1247	.1146					
13	.0895	.0542	.0807	.0919	.1103	.0813	.1126	.1046
13	.0563	.0781	.0940	.0946				
14	.0579	.0493	.0602	.0661	.1524	.0423	.1151	.0714
14	.0407	.1064	.0872	.1076	.0165			
15	.2293	.0866	.3263	.6909	.1363	.0703	.0661	.0621
15	.0539	.0874	.1066	.0455	.0451	.0506		
16	.1091	.0727	.0892	.0869	.1125	.0999	.1382	.1313
16	.0883	.1092	.0883	.1286	.0553	.0627	.0702	
17	.1115	.0790	.0944	.0441	.1624	.0947	.2235	.1603
17	.1479	.1417	.1002	.1915	.1294	.1211	.1465	.0964
18	.1048	.0669	.0582	.0449	.0790	.1119	.1618	.1338
18	.1139	.1271	.1192	.1483	.1338	.0876	.1032	.0910
18	.0953							
19	.0922	.0720	.0942	.0670	.1510	.0801	.1558	.1068
19	.1014	.1139	.0897	.1448	.0895	.0737	.1010	.0707
19	.0382	.0962						
20	.0915	.0542	.0705	.0688	.1136	.0812	.0865	.0985
20	.0727	.0996	.0735	.1181	.0540	.0795	.0344	.0330
20	.1448	.0730	.0814					
21	.0863	.0612	.0793	.0694	.1536	.0670	.1232	.0803
21	.0510	.0819	.1092	.1090	.0585	.0490	.0829	.0711
21	.0944	.0990	.0502	.0453				
22	.0796	.0578	.1110	.0615	.1368	.0709	.1634	.0899
22	.0950	.0573	.0543	.1166	.0819	.0909	.1161	.0773
22	.0863	.1037	.0690	.0860	.0438			
23	.0964	.0652	.1146	.0730	.1974	.1092	.1951	.1372
23	.0912	.1285	.1022	.1817	.1017	.0832	.1077	.0839
23	.0600	.1129	.0249	.0906	.0418	.0766		
24	.0945	.0843	.1071	.0675	.0668	.0627	.1523	.1183
24	.1516	.0954	.0793	.0950	.1004	.0973	.1501	.1024
24	.0466	.0916	.0466	.0765	.1191	.0700	.0853	
25	.1451	.0929	.1197	.0792	.0774	.1103	.1907	.1213
25	.1498	.0659	.1077	.0829	.1067	.0857	.1134	.1139
25	.0485	.1401	.0340	.1081	.0990	.0299	.0509	.0384
26	.0731	.0854	.1079	.0564	.1375	.0989	.1465	.1042
26	.0978	.0974	.0799	.1377	.0995	.1006	.1331	.0871
26	.0921	.1069	.0501	.0713	.0471	.0557	.0399	.1156
26	.0412							
27	.1441	.1199	.1253	.0977	.1591	.1322	.2169	.1168
27	.1476	.1167	.1266	.1405	.1257	.0991	.0817	.1153
27	.0796	.1787	.0495	.1530	.1140	.1037	.0521	.0977
27	.0450	.0447						
28	.1004	.0603	.0677	.0746	.1328	.0443	.0829	.0670
28	.0758	.1070	.0608	.1200	.0372	.0373	.0429	.0290
28	.1227	.0850	.0776	.0536	.0746	.0948	.0993	.0621
28	.0479	.1095	.1237					

Pour en faciliter la lecture, les mêmes distances sont remplacées, ci-après, par des symboles de classe : 0, 1, 2, ...9, A, B, ... N, P, ... Z. Les distances dépassant 0,3400 sont ainsi abandonnées.

```
          1111111111222222222
         .1234567890123456789012345678
     .............................
        1..954CABA786E96NBCDAA98AAF8FB
        2.9.859CEB96AC659887867679A9C7
        3.58. E58769BE97Y9A6A88CCBCBD7
        4.45 .A6C8876DA7 955777778786A8
        5.C9EA.CEEHC97CGECH8GCGEK78EGE
        6.AC56C.B6A9AA958AAC9978B7CAE5
        7.BE8CEB.88BADCC7ENHG9DHKGKFM9
        8.AB78E68.A6A9B87EHEBA99ECDBC7
        9.7968HA8A.ABE6569FCB86AAGFAF8
       10.8697C9B6A.7D8B9BFDCA96DA7ACB
       11.6AB69AAAB7.CA9B9BC9BB6B8B8D7
       12.ECED7AD9EDC.AB5DKFFCBCJA9EFC
       13.969AC9CB68AA.256DE9669BBBAD4
       14.6577G5C85B9B2.67D9885A9A9BA4
       15.N9Y E87769B556.8FBB49CBGCE95
       16.B899CAEE9B9D678.AA84889BC9C3
       17.C8A5HANHFFBKDDFA.A4FA9755A8D
       18.B7658CHECDCFE9BAA.A8ABCAFBI9
       19.A8A7G9GBBC9F98B84A.967354658
       20.A687C99A8A8C6844F89.59A8B8G6
       21.9787G7D969BB6598AA65.55CA5C8
       22.86C7E8H9A66C9AC89B795.8836BA
       23.A7C8KBKEADBJB9B97C3A58.9646A
       24.A9B777GCGA8ABAGB5A58C89.4CA7
       25.FAC88CKDF7J9B9CC5F4BA364.555
       26.89B6EAFBAA8EABE9AB68564C5.5B
       27.FCDAGEMCFCDFDA9C8I5GCB6A55.D
       28.B778E5978B7C44530986 8AA75BD.
```

Présentée sous cette forme matricielle, la distance est lue à l'intersection de la ligne et de la colonne qui portent, respectivement, les numéros des deux profils entre lesquels la distance est mesurée.

A titre d'exemple, les résultats sont lus ainsi : entre les profils 13 et 14 (ligne 13, colonne 14) la distance est 2 (classe 0,0101 à 0,0200), entre les profils 3 et 6 (intersection de la ligne 3 et de la colonne 6) la distance est 5 (classe 0,0401 à 0,0500), entre les profils 3 et 11 le code B représente une distance de la classe 0,1001 à 0,1100 . La distance étant une mesure relative, comprise entre 0 et 1, et le tiers inférieur méritant une attention plus particulière, la représentation ci-dessus semble suffire.

Bibliographie

Cardoso J. Carvalho, Garcia J. Sacadura, 1962 - Carta dos solos de São Tomé e Principe. Mem. Junta Invest. Ultram., 2a sér., nº 39, Lisboa, 306 p.

Kendall M.G., Stuart A., 1966 - The advanced theory of statistics. vol. 3. Design and analysis, and time series. Griffin, London, 552 p.