

Tableaux individus-variables

	variable 1	...	variable j	...	variable p
individu 1	x_1^1		x_1^j		x_1^p
individu i	x_i^1		x_i^j		x_i^p
individu n	x_n^1		x_n^j		x_n^p

Exemples des notes

	mathématiques	science	français	latin	dessin-musique
jean	6.0	6.0	5	5.5	8
aline	8.0	8.0	8	8.0	9
annie	6.0	7.0	11	9.5	11
monique	14.5	14.5	16	15.0	8
didier	14.0	14.0	12	12.5	10
andré	11.0	10.0	6	7.0	13
pierre	5.5	7.0	14	11.5	10
brigitte	13.0	12.5	8	9.5	12
evelyne	9.0	9.5	12	12.0	18

- Exemple classique en statistique multidimensionnelle
- Proposé par Fisher pour illustrer les méthodes de discrimination
- 150 iris provenant de 3 familles différentes : Virginia, Versicolor et Setosa
- Longueur et la largeur du sépale et du pétale

	<i>type</i>	<i>LoSe</i>	<i>laSe</i>	<i>LoPe</i>	<i>laPe</i>
1	1.0	5.1	3.5	1.4	0.2
2	1.0	4.9	3.0	1.4	0.2
3	1.0	4.7	3.2	1.3	0.2
4	1.0	4.6	3.1	1.5	0.2
...					
51	2.0	7.0	3.2	4.7	1.4
52	2.0	6.4	3.2	4.5	1.5
53	2.0	6.9	3.1	4.9	1.5
54	2.0	5.5	2.3	4.0	1.3
...					
101	3.0	6.3	3.3	6.0	2.5
102	3.0	5.8	2.7	5.1	1.9
103	3.0	7.1	3.0	5.9	2.1
104	3.0	6.3	2.9	5.6	1.8
...					
150	3.0	5.9	3.0	5.1	1.8

Présence d'espèces végétales dans des parcelles

variables	parcelle 1	parcelle 2	...
individus			
espèce 1	1	0	...
espèce 2	0	0	...
espèce 3	1	1	...
...
...

No	Nom	Libelle
1	C01	plaque de forme ronde
2	C14	plaque de grande taille
3	C15	fixation par bossettes de bronze
4	C16	fixation par clous de fer
5	C17	plaque dorsale carree
6	C19	ardillon a plateau rond
7	C22	damasquinure monochrome argent
8	C23	damasquinure bichrome
9	C24	damasquinure par incrustation dominante
10	C25	damasquinure par placage predominant
11	C26	decor sur toute la surface de la plaque
12	C28	trame geometrique exclusive
13	C29	tresse
14	C30	entrelacs de rubans
15	C31	entrelacs animaliers
16	C32	arabesques
17	C33	incrustations avec bande pointillee
18	C34	incrustations avec bande hachure ou nid
19	C35	incrustations avec ruban plein
20	C36	incrustations filiformes
21	C37	fond plaque d'argent
22	C38	fond hachure ou a trame geometrique
23	C39	bordures de frises geometriques ou de ha
24	C40	bordures offrant la repetition du motif
25	C41	motifs animaliers
26	C42	plaques a bords mouvementes

Données Méro (fin)

	01	14	15	16	17	19	22	23	24	25	26	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
01	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	1	1	0	0	1	0	1
02	0	0	0	1	1	0	0	1	0	1	0	0	0	0	1	0	0	1	0	1	1	0	0	1	1	1
03	0	0	1	0	0	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0
04	0	0	0	1	1	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	1	1	0	0
05	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
06	0	0	0	1	1	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	1	1	1
07	0	0	1	0	1	0	1	0	1	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	1	1
08	0	0	0	0	0	0	1	0	1	0	1	0	0	0	1	0	0	1	0	0	1	0	0	1	0	1
09	0	0	1	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	1	0	1	0
10	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	1	1	0	0	1	0	0
11	0	0	1	0	1	0	1	0	1	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0
12	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	1	1	1
13	0	0	0	1	1	0	0	1	0	1	0	0	0	0	1	0	0	0	1	1	0	0	1	1	0	0
14	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	1	0
15	1	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	1	1	0	0	1	0	1
17	0	0	1	0	0	0	1	0	1	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0
18	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	1	1	1
19	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	0	1	0	1	0	1	1	0	1	0
20	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	1	0
21	0	1	0	1	0	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0
22	0	0	0	1	1	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0
23	1	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	1	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0
25	0	0	1	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0
26	0	1	1	0	0	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0

- Nominal : codage disjonctif complet

	v
1	3
2	1
3	3
4	2
5	1

	v1	v2	v3
1	0	0	1
2	1	0	0
3	0	0	1
4	0	1	0
5	1	0	0

- Ordinale : codage additif

	v
1	3
2	1
3	3
4	2
5	1

	v1	v2	v3
1	1	1	1
2	1	0	0
3	1	1	1
4	1	1	0
5	1	0	0

Distances mesurées sur une carte

	Lond	Stoc	Lisb	Madr	Pari	Amst	Berl	Prag	Rome	Dubl
Londres	0	569	667	530	141	140	357	396	569	190
Stockholm	569	0	1212	1043	617	446	325	423	787	648
Lisbonne	667	1212	0	201	596	768	923	882	714	714
Madrid	530	1043	201	0	431	608	740	690	516	622
Paris	141	617	596	431	0	177	340	337	436	320
Amsterdam	140	446	768	608	177	0	218	272	519	302
Berlin	357	325	923	740	340	218	0	114	472	514
Prague	396	423	882	690	337	272	114	0	364	573
Rome	569	787	714	516	436	519	472	364	0	755
Dublin	190	648	714	622	320	302	514	573	755	0

- Proximités évaluées subjectivement
- Notes allant de 0 (pas de ressemblance) à 10 (forte ressemblance)

	parfums	1	2	3	4	5
1		—				
2		3	—			
3		5	8	—		
4		2	7	1	—	
5		9	3	5	7	—

Distances horaires par le train

	Bord	Gren	Lill	Limo	Lyon	Mars	Mont	Pari	Perp	Renn	Stra	Toul
Bord	0.00	7.18	5.27	2.14	7.43	5.19	3.57	2.58	4.38	5.45	7.59	2.06
Gren	10.21	0.00	4.55	8.01	1.12	3.31	3.27	2.59	4.27	6.16	5.00	5.31
Lill	4.59	5.05	0.00	4.57	3.54	7.46	6.37	1.01	10.20	4.23	5.08	7.08
Limo	2.16	8.06	5.17	0.00	5.27	9.29	6.47	3.01	5.00	8.36	10.27	3.19
Lyon	7.33	1.12	3.53	5.13	0.00	2.44	2.55	2.00	4.30	4.50	5.05	5.25
Mars	5.28	3.30	7.48	8.23	2.49	0.00	1.30	4.45	3.19	14.29	8.36	3.25
Mont	3.55	3.04	6.27	6.02	2.44	1.20	0.00	4.45	1.21	7.51	9.02	2.36
Pari	2.58	2.56	1.00	3.02	2.10	4.45	4.45	0.00	6.34	2.04	3.56	5.09
Perp	4.31	4.27	10.12	4.56	4.35	3.23	1.36	6.44	0.00	11.21	11.46	2.07
Renn	5.31	6.15	4.25	6.30	4.37	11.55	7.52	2.05	12.08	0.00	7.47	8.26
Stra	8.12	6.49	5.08	11.19	4.47	8.33	8.40	4.05	10.11	7.54	0.00	11.01
Toul	2.09	5.37	7.35	3.17	5.49	3.14	2.40	5.06	2.00	8.40	11.03	0.00

Exemples de distances

Euclidienne	$\sqrt{\sum_j (x^j - y^j)^2} = \sqrt{(x - y)'I(x - y)}$
Euclidienne pondérée	$\sqrt{(x - y)'D(x - y)}$
Mahalanobis	$\sqrt{(x - y)'S^{-1}(x - y)}$
City-block ou L_1	$\sum_j x^j - y^j $
Chebychev ou L_∞	$\max_j x^j - y^j $
Minkowski ou L_p	$(\sum_j x^j - y^j ^p)^{1/p}$
Distance du χ^2	
Distance entre variables	$d = 1 - r^2$

Exemples de distances pour données binaires

$$\begin{aligned}x &= (1 \ 0 \ 1 \ 0 \ \dots \ 0) \\y &= (0 \ 0 \ 1 \ 1 \ \dots \ 0)\end{aligned}$$

$$\begin{array}{c|cc} & 1 & 0 \\ \hline 1 & a & b \\ 0 & c & d\end{array}$$

Indice	$d(x, y)$
Csekanowski, Sorensen, Dice	$\frac{2a}{2a+b+c}$
Hamman	$\frac{(a+d)-(b+c)}{a+b+c+d}$
Jaccard	$\frac{a}{a+b+c}$
Kulezynsk	$\frac{a}{a+b}$
Ochiai	$\frac{a}{[(a+b)(a+c)]^{1/2}}$