

Analyse des correspondances binaires

Tableau de fréquence et indépendance

Véronique Tremblay

Dans cette capsule

- Tableau de fréquences (vocabulaire et notation)
- Indépendance



X : Une variable catégorielle à n modalités

Y : Une variable catégorielle à p modalités

\mathbf{K} : Le tableau de fréquence (contingence)

k_{ij} : Le nombre d'individus appartenant à la catégorie i de X et à la catégorie j de Y

$k_{i\bullet}$: La somme des observations dans la modalité i de X

$k_{\bullet j}$: La somme des observations dans la modalité j de Y

$k_{\bullet\bullet}$: Le nombre total d'observations

Tableau de fréquence K

	Chatains	Roux	Blonds	$k_{i\bullet}$
Marrons	119	26	7	152
Noisette	54	14	10	78
Verts	29	14	16	59
Bleus	84	17	94	195
$k_{\bullet j}$	286	71	127	484

Fréquences relatives



F: Tableau des fréquences relatives

f_{ij} : Proportion d'individus appartenant à la catégorie i de X et à la catégorie j de Y

Exemple de tableau de fréquences relatives

$$f_{ij} = \frac{k_{ij}}{k_{\bullet\bullet}}$$

Fréquences

	Chatains	Roux	Blonds	$k_{i\bullet}$
Marrons	119	26	7	152
Noisette	54	14	10	78
Verts	29	14	16	59
Bleus	84	17	94	195
$k_{\bullet j}$	286	71	127	484

Fréquences relatives

	Chatains	Roux	Blonds
Marrons	0.25	0.05	0.01
Noisette	0.11	0.03	0.02
Verts	0.06	0.03	0.03
Bleus	0.17	0.04	0.19

Marges



$f_{i\bullet}$: La proportion des individus dans la modalité i de X

$f_{\bullet j}$: La proportion des individus dans la modalité j de X

$$f_{i\bullet} = \sum_{j=1}^p f_{ij} = k_{i\bullet}/k_{\bullet\bullet}$$

Fréquences

	Chatains	Roux	Blonds	$k_{i\bullet}$
Marrons	119	26	7	152
Noisette	54	14	10	78
Verts	29	14	16	59
Bleus	84	17	94	195
$k_{\bullet j}$	286	71	127	484

Fréquences relatives et marge ligne

	Chatains	Roux	Blonds	$f_{i\bullet}$
Marrons	0.25	0.05	0.01	0.31
Noisette	0.11	0.03	0.02	0.16
Verts	0.06	0.03	0.03	0.12
Bleus	0.17	0.04	0.19	0.40

Marge colonne

$$f_{\bullet j} = \sum_{i=1}^n f_{ij} = k_{\bullet j} / k_{\bullet \bullet}$$

Fréquences

	Chatains	Roux	Blonds	$k_{i\bullet}$
Marrons	119	26	7	152
Noisette	54	14	10	78
Verts	29	14	16	59
Bleus	84	17	94	195
$k_{\bullet j}$	286	71	127	484

Fréquences relatives et marge colonne

	Chatains	Roux	Blonds
Marrons	0.25	0.05	0.01
Noisette	0.11	0.03	0.02
Verts	0.06	0.03	0.03
Bleus	0.17	0.04	0.19
$f_{\bullet j}$	0.59	0.15	0.26

Marges

	Chatains	Roux	Blonds	$f_{i\bullet}$
Marrons	0.25	0.05	0.01	0.31
Noisette	0.11	0.03	0.02	0.16
Verts	0.06	0.03	0.03	0.12
Bleus	0.17	0.04	0.19	0.40
$f_{\bullet j}$	0.59	0.15	0.26	1.00

Profils lignes et colonnes

Profils colonnes

$$C_j = \left(\frac{k_{1j}}{k_{\bullet j}}, \dots, \frac{k_{nj}}{k_{\bullet j}} \right) = \left(\frac{f_{1j}}{f_{\bullet j}}, \dots, \frac{f_{nj}}{f_{\bullet j}} \right)$$

Fréquences

	Chatains	Roux	Blonds	$k_{i\bullet}$
Marrons	119	26	7	152
Noisette	54	14	10	78
Verts	29	14	16	59
Bleus	84	17	94	195
$k_{\bullet j}$	286	71	127	484

Profils colonnes

	Marrons	Noisette	Verts	Bleus
Chatains	0.42	0.19	0.10	0.29
Roux	0.37	0.20	0.20	0.24
Blonds	0.06	0.08	0.13	0.74

En termes matriciels on pose $D_p = \text{diag}(f_{\bullet j})$, de sorte que

$$C = D_p^{-1} F$$

Profil colonne moyen

$$\left(\sum_{j=1}^p f_{\bullet j} \frac{f_{1j}}{f_{\bullet j}}, \dots, \sum_{j=1}^p f_{\bullet j} \frac{f_{nj}}{f_{\bullet j}} \right) = (f_{1\bullet}, \dots, f_{n\bullet}).$$

Profil colonne moyen

Marrons	Noisette	Verts	Bleus
0.31	0.16	0.12	0.4

Profil colonne et indépendance

Profils colonnes et profil colonne moyen

	Marrons	Noisette	Verts	Bleus
Chatains	0.42	0.19	0.10	0.29
Roux	0.37	0.20	0.20	0.24
Blonds	0.06	0.08	0.13	0.74
$f_{i\bullet}$	0.31	0.16	0.12	0.40

Corrélation entre les profils colonnes

	Chatains	Roux	Blonds
Chatains	1.00	0.92	0.13
Roux	0.92	1.00	-0.16
Blonds	0.13	-0.16	1.00

Profils lignes

$$L_i = \left(\frac{k_{i1}}{k_{i\bullet}}, \dots, \frac{k_{ip}}{k_{i\bullet}} \right) = \left(\frac{f_{i1}}{f_{i\bullet}}, \dots, \frac{f_{ip}}{f_{i\bullet}} \right)$$

Fréquences

	Chatains	Roux	Blonds	$k_{i\bullet}$
Marrons	119	26	7	152
Noisette	54	14	10	78
Verts	29	14	16	59
Bleus	84	17	94	195
$k_{\bullet j}$	286	71	127	484

Profils lignes

	Chatains	Roux	Blonds
Marrons	0.78	0.17	0.05
Noisette	0.69	0.18	0.13
Verts	0.49	0.24	0.27
Bleus	0.43	0.09	0.48

En termes matriciels, on posera $D_n = \text{diag}(f_{i\bullet})$ de sorte que

$$L = D_n^{-1} F$$

Profil ligne moyen

$$\left(\sum_{i=1}^n f_{i\bullet} \frac{f_{i1}}{f_{i\bullet}}, \dots, \sum_{i=1}^n f_{i\bullet} \frac{f_{ip}}{f_{i\bullet}} \right) = (f_{\bullet 1}, \dots, f_{\bullet p})$$

Profil ligne moyen

Chatains	Roux	Blonds
0.59	0.15	0.26

Profils lignes et indépendance

Profils lignes et profil ligne moyen

	Marrons	Noisette	Verts	Bleus
Chatains	0.42	0.19	0.10	0.29
Roux	0.37	0.20	0.20	0.24
Blonds	0.06	0.08	0.13	0.74
$f_{\bullet j}$	0.59	0.15	0.26	0.59

Corrélation entre les profils lignes

	Marrons	Noisette	Verts	Bleus
Marrons	1.00	1.00	0.96	0.24
Noisette	1.00	1.00	0.98	0.32
Verts	0.96	0.98	1.00	0.50
Bleus	0.24	0.32	0.50	1.00

- Fréquences et fréquences relatives
- Marges
- Profils lignes et profils colonnes