

Chapitre 9

ANALYSE FACTORIELLE DES CORRESPONDANCES

1. NATURE DES DONNÉES ÉTUDIÉES.

1.1 Données initiales.

- Tableaux de contingence obtenus par tri croisé d'un échantillon ou assimilés
- Tableaux de mesures positives
- Tableaux disjonctifs complets (analyse des correspondances multiples)

Exemple de tableau de contingence:

X : 3 modalités brunes : fumeur de brunes blondes : fumeur de blondes non fumeur		Y : 4 modalités mm : mineur masculin mf : mineur féminin MF majeur féminin MM : majeur masculin			
		Y			
		mm	mf	MF	MM
X	brunes	63	37	41	47
	blondes	36	55	39	38
	non fumeur	34	27	72	38

Notations :

Dans le cas général, le tableau des proportions est noté P_{IJ} :

		Y					
		y_1	y_2	...	y_i	...	y_q
X	x_1	p_{11}	p_{12}	...	p_{1i}	...	p_{1q}
	x_2	p_{21}	p_{22}	...	p_{2j}	...	p_{2q}

	x_i	p_{i1}	p_{ii}	...	p_{ij}	...	p_{iq}

	x_p	p_{p1}	p_{p2}	...	p_{pi}	...	p_{pq}

On a évidemment : $p_{ij} = n_{ij} / N$ en notant n_{ij} l'effectif figurant à la ligne i (modalité x_i) et la colonne j (modalité y_j), et N l'effectif total.

1.2 unités statistiques.

Définition : les marges du tableau sont les répartitions suivantes :

- $P_i = (p_{i\bullet})$ $i = 1, \dots, p$: répartition des réponses à la question X suivant les modalités x_1, x_2, \dots (en proportions)
 $p_{i\bullet}$ = somme des p_{ij} pour i fixé
- $P_j = (p_{\bullet j})$ $j = 1, \dots, q$: répartition des réponses à la question Y suivant les modalités y_1, y_2, \dots (en proportions)
 $p_{\bullet j}$ = somme des p_{ij} pour j fixé

Définition des profils

Les profils lignes sont les répartitions suivant les modalités en colonnes des individus donnés en lignes :

	mm	mf	MF	MM	total
profil brunes	0.335	0.197	0.218	0.250	1
profil blondes	0.214	0.327	0.232	0.226	1
profil non fumeur	0.199	0.158	0.421	0.222	1
centre de gravité P_j	0.252	0.226	0.288	0.233	1

profils lignes P_j^i

La notation est la suivante :

		Y					
		y_1	y_2	...	y_i	...	y_q
X	P_j^1	p_1^1	p_2^1	...	p_i^1	...	p_q^1
	P_j^2	p_1^2	p_2^2	...	p_i^2	...	p_q^2

	P_j^i	p_1^i	p_2^i	...	p_i^i	...	p_q^i

	P_j^p	p_1^p	p_2^p	...	p_i^p	...	p_q^p
	P_j	$p_{\bullet 2}$	$p_{\bullet 3}$...	$p_{\bullet 4}$...	$p_{\bullet 5}$

Les profils colonnes sont les répartitions suivant les modalités en lignes des individus donnés en colonnes.

	profil mm	profil mf	profil MF	profil MM	centre de gravité P_i
brunes	0.474	0.311	0.270	0.382	0.357
blondes	0.271	0.462	0.257	0.309	0.319
non fumeur	0.256	0.227	0.474	0.309	0.324
total	1	1	1	1	1

profils colonnes P_i^j

La notation est la suivante :

		Y					centre de gravité P_i	
		P_i^1	P_i^2	...	P_i^j	...		P_i^q
X	x_1	p_1^1	p_1^2	...	p_1^j	...	p_1^q	$p_{1\bullet}$
	x_2	p_2^1	p_2^2	...	p_2^j	...	p_2^q	$p_{2\bullet}$

	x_i	p_i^1	p_i^2	...	p_i^j	...	p_i^q	$p_{i\bullet}$

	x_p	p_p^1	p_p^2	...	p_p^j	...	p_p^q	$p_{p\bullet}$
		1	1		1		1	

2. PRÉSENTATION DE LA MÉTHODE.

2.1 Notion de distance entre deux profils.

Définitions :

- La distance du χ^2 entre deux profils lignes P_j^i et $P_j^{i'}$ est définie par son carré de la façon suivante :

$$d^2(i,i') = \sum_{j=1}^q [p_j^i - p_j^{i'}]^2 / p_{\bullet j}$$

- La distance du χ^2 entre deux profils colonnes P_i^j et $P_i^{j'}$ est définie par son carré de la façon suivante :

$$d^2(i,i') = \sum_{j=1}^p [p_i^j - p_i^{j'}]^2 / p_{i\bullet}$$

Exemples :

	x ₁	x ₂	x ₃
x ₁	0.000		
x ₂	0.136	0.000	
x ₃	0.226	0.252	0.000

Distances entre les profils lignes

	y ₁	y ₂	y ₃	y ₄
y ₁	0.000			
y ₂	0.192	0.000		
y ₃	0.264	0.325	0.000	
y ₄	0.037	0.109	0.128	0.000

Distances entre les profils colonnes

2.2 Description d'un ensemble de profils. Définitions.

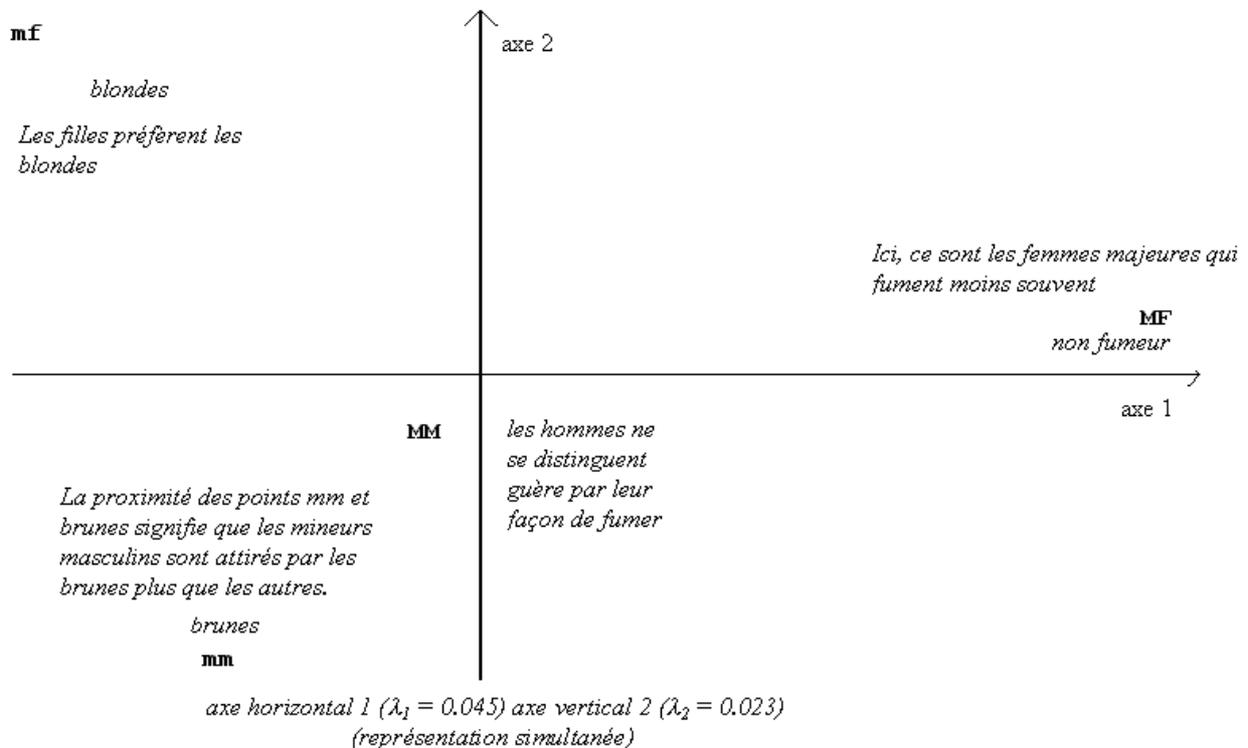
- Recherche des axes les plus proches des points (critère des moindres carrés)
- Axes principaux et composantes principales : facteurs principaux.
- Inertie = variance des facteurs
- somme des inertie : χ^2/N (N : somme du tableau)
- Valeur propre « triviale » égale à 0 ou 1 suivant le logiciel.
- $\chi^2 = 35.726$, $P(X^2 > x^2) = 0.0000$
- $\lambda_1 = 0.04495$, $\lambda_2 = 0.02284$, $\lambda_3 = 0.00000$, $\lambda_4 = 0$

2.3 Propriété de dualité.

- Cette propriété relie les profils colonnes et les profils lignes.
- Les ensembles de profils ont les mêmes valeurs propres.
- Il existe une relation entre les facteurs sur les profils lignes et les facteurs sur les profils colonnes.

3. INTERPRÉTATION DES RÉSULTATS.

3.1 Représentation graphique. Plans principaux.



3.2 Paramètres d'aide à l'interprétation.

Poids : chaque profil intervient d'autant plus dans l'analyse que son poids, égal à la somme p_i (profil ligne) ou p_j (profil colonne) est élevé.

contributions : les contributions mesurent l'influence des profils dans le calcul des axes principaux.

Moyennes et variances des facteurs : comme en ACP, les moyennes des facteurs sont nulles, les variances égales aux valeurs propres, et les covariances à 0. On n'oubliera pas de pondérer les profils dans les calculs.

Coordonnées des profils lignes :

	poids	C(i)	Cos ² (i)	Contrib(i)	C(i)	Cos ² (i)	Contrib(i)
brunes	0.357	-0.128	0.331	12.9	-0.181	0.669	51.4
blondes	0.319	-0.168	0.449	20.0	0.186	0.551	48.2
ne fume pas	0.324	0.305	0.997	67.1	0.017	0.003	0.4

Coordonnées des profils colonnes :

	poids	C(j)	Cos ² (j)	Contrib(j)	C(j)	Cos ² (j)	Contrib(j)
mm	0.252	-0.131	0.286	9.7	-0.207	0.714	47.5
mf	0.226	-0.226	0.513	25.7	0.220	0.487	48.0
MF	0.288	0.316	0.980	64.1	0.045	0.020	2.5
MM	0.233	-0.030	0.312	0.5	-0.044	0.688	2.0