

Analyse des Correspondances Simples

Anne B Dufour

Octobre 2010

Analyse Factorielle des correspondances

Nishisato, dans son livre *Analysis of categorical data : dual scaling and its applications* (1980), l'appelle *dual scaling* mais cite (p. 11) les noms de :

- *the method of reciprocal averages*
- *additive scoring*
- *appropriate scoring*
- *canonical scoring*
- *Guttman weighting*
- *principal component analysis of qualitative data*
- *optimal scaling*
- *Hayashi's theory of quantification*
- *simultaneous linear regression*
- *correspondence factor analysis*
- *biplot*

Analyse Factorielle des correspondances

- Les extensions, généralisations, utilisations particulières dans chaque discipline, sont tellement nombreuses qu'établir la liste des approches n'est plus un objectif raisonnable.
- Le schéma de dualité permet de faire le lien entre toutes les approches.



- 1 Croisement de deux variables qualitatives
- 2 Exploration des modalités croisées

Introduction

Soient A et B , deux variables qualitatives ayant respectivement I et J modalités. Soit n , le nombre d'individus sur lesquels A et B ont été observées. On note n_{ij} le nombre d'individus possédant à la fois la modalité i de la variable A et la modalité j de la variable B .

	B_1	\dots	B_j	\dots	B_J	total
A_1	n_{11}	\dots	n_{1j}	\dots	n_{1J}	$n_{1.}$
\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
A_i	n_{i1}	\dots	n_{ij}	\dots	n_{iJ}	$n_{i.}$
\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
A_I	n_{I1}	\dots	n_{Ij}	\dots	n_{IJ}	$n_{I.}$
total	$n_{.1}$	\dots	$n_{.j}$	\dots	$n_{.J}$	$n_{..}$

Introduction

La table de contingence observée est un tableau croisé où les colonnes correspondent aux J modalités de la variable B et les lignes aux I modalités de la variable A .

Il est important de rappeler que :

- i) tout individu présente 1 modalité et 1 seule de chaque variable ;
- ii) chaque modalité doit avoir été observée une fois, sinon elle est supprimée.

Exemple

Le tableau donne la répartition de 5694 couples formés de deux conjoints actifs en fonction de la catégorie socio-professionnelle de la femme (F en lignes) et du mari (H en colonnes) en 1982.

Le code des lignes et des colonnes est commun :

- 1 Agriculteur exploitant agri et salarié agricole ouva
- 2 Patron (commerce et industrie) pat, profession libérale et cadre supérieur sup
- 3 Cadre moyen moy
- 4 Employé emp
- 5 Ouvrier ouv et personnel de service serv
- 6 autres aut

Il est extrait d'un article de Vallet (1986) et simplifié pour des raisons pédagogiques.

Exemple

```
library(ade4)
data(mariages)
mariages
```

	Hagri	Houva	Hpat	Hsup	Hmoy	Hemp	Houv	Hserv	Haut
Fagri	420	3	8	2	2	4	20	0	0
Fouva	2	12	0	0	0	1	10	0	0
Fpat	9	1	333	38	24	22	48	7	3
Fsup	3	1	18	158	53	18	24	2	5
Fmoy	19	4	61	201	309	113	225	14	29
Femp	20	9	111	159	323	348	756	34	65
Fouv	13	14	41	24	79	118	795	22	19
Fserv	8	11	30	21	61	82	382	44	20
Faut	0	0	1	3	2	1	2	0	6

Pour 53 couples sur les 5694 de l'étude, la femme appartient à une profession libérale ou est cadre supérieur tandis que son mari est cadre moyen.

Théorie des tests

Dans la théorie des tests, l'objectif est de définir s'il existe un lien ou non entre les deux variables qualitatives A et B .

L'hypothèse nulle H_0 est l'indépendance entre les deux variables (absence d'effet).

L'hypothèse alternative H_1 est la présence d'un lien entre les deux variables.

La réponse à la question est le test du Chi-Deux de contingence.

Table de contingence théorique

Pour ce faire, on construit la table de contingence dite théorique : répartition des n individus sous l'hypothèse H_0 .

Soit i une modalité de la variable A et j une modalité de la variable B . La fréquence théorique est définie par :

$$\begin{aligned}P(A = i \cap B = j) &= P(A = i) \times P(B = j) \\ &= \frac{n_{i.}}{n} \times \frac{n_{.j}}{n}\end{aligned}$$

L'effectif théorique est donc :

$$n \times \frac{n_{i.}}{n} \times \frac{n_{.j}}{n} = \frac{n_{i.} \cdot n_{.j}}{n}$$

Valeur du Chi-Deux de contingence

On compare les valeurs de la table de contingence observée avec les valeurs de la table de contingence théorique, construite sous l'hypothèse nulle.

$$\chi_{obs}^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{\left(n_{ij} - \frac{n_{i.} \cdot n_{.j}}{n} \right)^2}{\frac{n_{i.} \cdot n_{.j}}{n}}$$

La simple différence observée / théorique ne suffit pas.

Pourquoi diviser par l'effectif théorique ?

Exemple.

On joue à pile ou face 10 fois. On gagne 1 fois, l'écart est de 4.

On joue à pile ou face 100 fois. On gagne 46 fois, l'écart est le même.

théorique	observé	écart	écart au carré	écart au carré/théorique
5	1	4	16	3.2
50	46	4	16	0.32

Mais gagner 1 fois sur 10 n'est pas la même chose que gagner 46 fois sur 100.

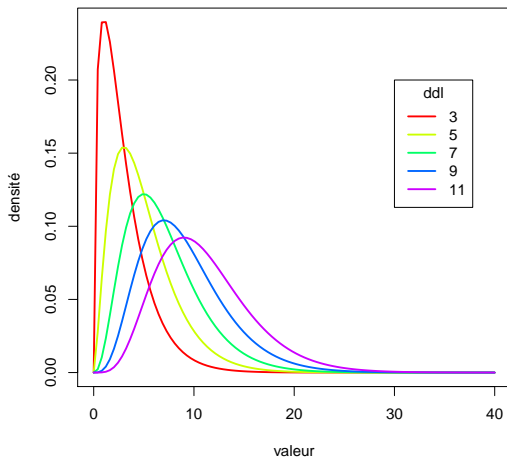
Lien entre deux variables qualitatives

- ① Point de vue descriptif : coefficient de Cramer

$$V = \sqrt{\frac{\chi_{obs}^2}{n \inf(I-1, J-1)}}$$

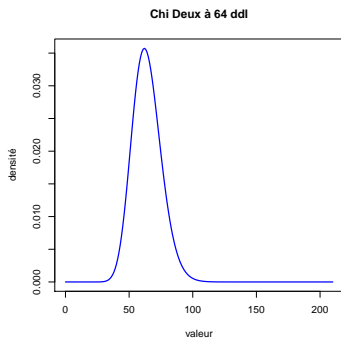
- ② Point de vue inférentiel. Sous l'hypothèse H_0 , la théorie nous dit que la statistique du test suit une loi du Chi-Deux à $(I-1)(J-1)$ degrés de liberté.

Lois du Chi Deux



Exemple

La valeur de la statistique du Chi-Deux calculée sur les données du mariage est **8722.15** et le coefficient de Cramer vaut **0.41**.



Les deux variables sont liées. Peut-on en savoir plus ?

Visualisation d'une table de contingence

- ❶ Les effectifs observés sont visualisés par :
 - ❶ des carrés
 - ❷ des couleurs.
- ❷ La table est visualisée sous forme de mosaïque à l'aide :
 - ❶ des distributions marginales
 - ❷ et des surfaces de résidus standardisés.

Tables de contingence autour d'un petit exemple :

observée

	Hsup	Hmoy	Hemp
Fsup	158	53	18
Fmoy	201	309	113
Femp	159	323	348

théorique

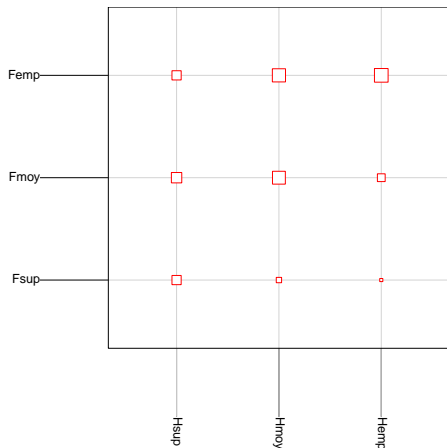
	Hsup	Hmoy	Hemp
Fsup	70.52	93.26	65.21
Fmoy	191.86	253.72	177.42
Femp	255.61	338.02	236.37

standardisé

	Hsup	Hmoy	Hemp
Fsup	10.42	-4.17	-5.85
Fmoy	0.66	3.47	-4.84
Femp	-6.04	-0.82	7.26

Visualisation des observations

Carrés proportionnels :



Couleurs proportionnelles :

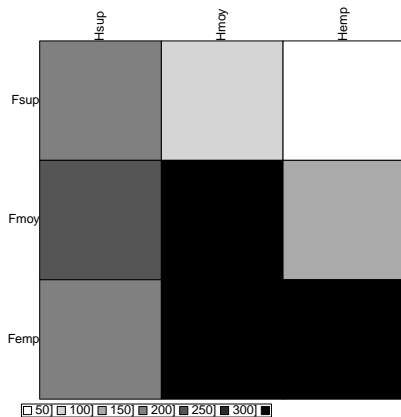


Diagramme en mosaïque - version simple

La surface d'un élément de la mosaïque est proportionnelle aux effectifs contenus dans les cellules de la table de contingence.

- 1 Sur l'axe horizontal, la distribution marginale de la variable B est utilisée pour déterminer les largeurs des mosaïques : $p_{.j} = \frac{n_{.j}}{n_{..}}$

Hsup	Hmoy	Hemp
0.31	0.41	0.28

- 2 Pour chaque modalité de B , on répartit les modalités de la variable A . La hauteur dans chaque bande est donc proportionnelle à : $p_{i/j} = \frac{n_{ij}}{n_{.j}}$

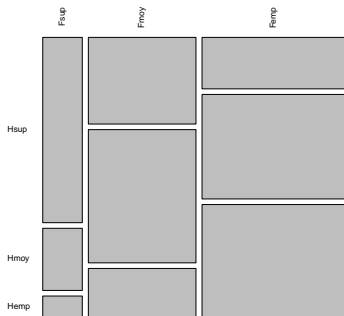
	Hsup	Hmoy	Hemp
Fsup	0.31	0.10	0.03
Fmoy	0.29	0.45	0.16
Femp	0.33	0.67	0.73

La surface d'un élément de la mosaïque est le produit de la largeur par la hauteur et représente bien la fréquence relative observée : $n_{ij}/n_{..}$.

La fréquence marginale en ligne sert de référence visuelle. Toutes les comparaisons sont alors possibles.

Diagramme en mosaïque - exemple

Observations



Hypothèse d'indépendance (baseline)

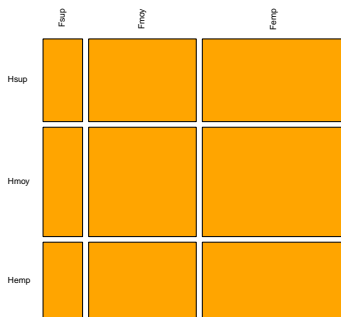


Diagramme en mosaïque - résidus standardisés

S'il y avait indépendance entre les deux variables, les distributions conditionnelles seraient égales aux distributions marginales. Les mosaïques seraient toutes alignées.

Une autre représentation se construit autour des résidus standardisés :

$$d_{ij} = \frac{n_{ij} - \frac{n_{i.}n_{.j}}{n}}{\sqrt{\frac{n_{i.}n_{.j}}{n}}}$$

Deux couleurs sont superposées sur la mosaïque.

- Le bleu signifie une sur représentativité des effectifs observés par rapport aux effectifs théoriques.
- Le rouge signifie une sous représentativité des effectifs observés par rapport aux effectifs théoriques.

```
mosaicplot(ssmariages, main = "Diagramme en\n mosaïque - Exemple",
  las = 2, shade = T)
```

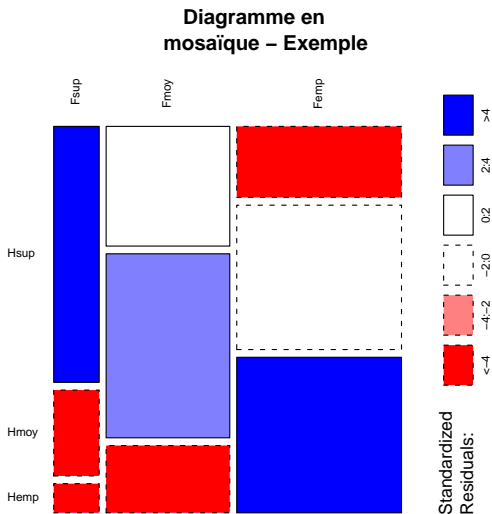
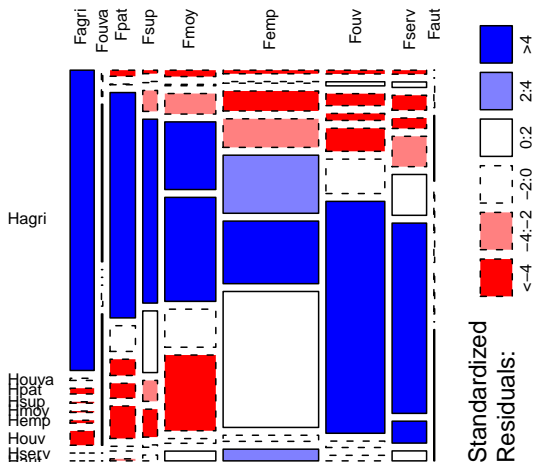


Diagramme en mosaïque – Exemple complet



Rappel sur les données de départ

On considère un tableau \mathbf{T} de nombres positifs ou nuls, comportant I lignes et J colonnes. On note n_{ij} son terme général, $n_{i.}$ et $n_{.j}$ les sommes marginales, n la somme de tous les éléments du tableau :

$$n_{i.} = \sum_{j=1}^J n_{ij} \quad n_{.j} = \sum_{i=1}^I n_{ij} \quad n = \sum_{i=1}^I n_{i.} = \sum_{j=1}^J n_{.j}$$

On calcule les fréquences conjointes p_{ij} , les fréquences marginales $p_{i.}$ et $p_{.j}$:

$$p_{ij} = \frac{n_{ij}}{n} \quad p_{i.} = \frac{n_{i.}}{n} \quad p_{.j} = \frac{n_{.j}}{n}$$

Objectif

Rechercher la meilleure représentation simultanée de deux ensembles constituant les lignes et les colonnes d'une table de contingence. Les lignes et les colonnes sont deux partitions des mêmes n individus. Elles jouent donc des rôles parfaitement symétriques.

On raisonne en terme de profils



Les modalités d'une même variable sont alors comparables entre elles. Les proximités entre les points s'interprètent comme des similitudes.

Nuage des I lignes dans \mathbb{R}^J

Chaque point a pour coordonnées :

$$\begin{bmatrix} p_{ij} \\ p_{i.} \end{bmatrix}$$

et est affecté du poids $p_{i.}$, sa fréquence relative.

Son centre de gravité est la fréquence marginale des colonnes :

$$\sum_{i=1}^I p_{i.} \frac{p_{ij}}{p_{i.}} = p_{.j}$$

Nuage des J colonnes dans \mathbb{R}^I

Chaque point a pour coordonnées :

$$\begin{bmatrix} p_{ij} \\ p_{.j} \end{bmatrix}$$

et est affecté du poids $p_{.j}$, sa fréquence relative.

Son centre de gravité est la fréquence marginale des lignes :

$$\sum_{j=1}^J p_{.j} \frac{p_{ij}}{p_{.j}} = p_{i.}$$

Critère d'ajustement

maximiser la somme pondérée des carrés des distances entre les points et le centre de gravité du nuage.

Distance entre deux points lignes

La distance euclidienne entre deux points i et i' exprimée sur le tableau des effectifs traduit la différence d'effectifs entre les deux modalités de la variable A .

La distance euclidienne entre deux points i et i' exprimée sur le tableau des profils ligne traduit la ressemblance ou la différence entre les deux modalités de la variable A .

$$d^2(i, i') = \sum_{j=1}^J \left(\frac{p_{ij}}{p_{i.}} - \frac{p_{i'j}}{p_{i'.}} \right)^2$$

Cette distance favorise les colonnes qui ont un poids $p_{.j}$ important. On pondère donc chaque écart par l'inverse du poids de la colonne et on calcule une nouvelle distance.

$$d^2(i, i') = \sum_{j=1}^J \frac{1}{p_{.j}} \left(\frac{p_{ij}}{p_{i.}} - \frac{p_{i'j}}{p_{i'.}} \right)^2$$

Schéma de base

On note \mathbf{P} le tableau des p_{ij} , \mathbf{D}_I et \mathbf{D}_J les matrices diagonales :

$$\mathbf{D}_I = \text{Diag}(p_{1.}, \dots, p_{I.}) \quad \mathbf{D}_J = \text{Diag}(p_{.1}, \dots, p_{.J})$$

Soit \mathbf{L} le tableau :

$$\mathbf{L} = \mathbf{D}_I^{-1} \mathbf{P} = \begin{bmatrix} p_{ij} \\ p_{i.} \end{bmatrix}$$

On a vu que le centrage d'une ligne i était $p_{.j}$ donc le centrage du nuage des I lignes est :

$$\mathbf{1}_{IJ} \mathbf{D}_J$$

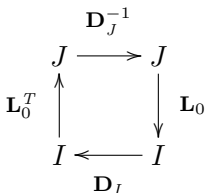
Le tableau centré est donc :

$$\mathbf{L}_0 = \mathbf{L} - \mathbf{1}_{IJ} \mathbf{D}_J$$

A chaque point, on a associé sa fréquence relative comme poids $p_{i.}$ (\mathbf{D}_I). Pour ne pas favoriser les colonnes avec un poids $p_{.j}$ important, on a pondéré l'écart entre deux points i et i' par l'inverse de ce poids $\frac{1}{p_{.j}}$ (\mathbf{D}_J^{-1}).

Schéma de base

L'Analyse des Correspondances est l'analyse du triplet $(\mathbf{L}_0, \mathbf{D}_J^{-1}, \mathbf{D}_I)$ visualisée par le schéma de dualité ci-dessous :



Éléments propres du schéma

- 1 Calcul de $\mathbf{H} = \mathbf{D}_J^{-1/2} \mathbf{L}_0^T \mathbf{D}_I \mathbf{L}_0 \mathbf{D}_J^{-1/2}$
- 2 Diagonalisation de \mathbf{H} , matrice symétrique réelle
- 3 Conservation des K premières valeurs propres non nulles dans $\Lambda_K = \text{Diag}(\lambda_1, \dots, \lambda_K)$ et des K premiers vecteurs propres associés, orthonormés pour la métrique canonique, en colonne dans \mathbf{U}_K . \mathbf{U}_K a J lignes et K colonnes et vérifie $\mathbf{U}_K^T \mathbf{U}_K = \mathbf{I}_K$.

Axes principaux et composantes principales

- Dans \mathbb{R}^I , on définit les axes principaux :

$$\tilde{\mathbf{A}} = \mathbf{D}_J^{-1/2} \mathbf{U}^{-1/2}$$

- Dans \mathbb{R}^J , on définit les composantes principales :

$$\tilde{\mathbf{C}} = \mathbf{L}_0 \mathbf{D}_J^{-1/2} \mathbf{U}$$

Les axes principaux et les composantes principales sont centrés, de variance λ et de covariance nulle deux à deux.

On les note \mathbf{A} et \mathbf{C} lorsqu'ils sont de variances unités.

Distributions conditionnelles par ligne

Dans les notations du cas général, nous pouvons d'abord considérer le tableau des distributions conditionnelles par lignes, $\mathbf{L} = \mathbf{D}_I^{-1}\mathbf{P}$, de terme général :

$$p_{j/i} = \frac{p_{ij}}{p_i}$$

```
P <- mariages/sum(mariages)
round(P, digits = 4)
```

	Hagri	Houva	Hpat	Hsup	Hmoy	Hemp	Houv	Hserv	Haut
Fagri	0.0718	0.0005	0.0014	0.0003	0.0003	0.0007	0.0034	0.0000	0.0000
Fouva	0.0003	0.0021	0.0000	0.0000	0.0000	0.0002	0.0017	0.0000	0.0000
Fpat	0.0015	0.0002	0.0569	0.0065	0.0041	0.0038	0.0082	0.0012	0.0005
Fsup	0.0005	0.0002	0.0031	0.0270	0.0091	0.0031	0.0041	0.0003	0.0009
Fmoy	0.0032	0.0007	0.0104	0.0344	0.0528	0.0193	0.0385	0.0024	0.0050
Femp	0.0034	0.0015	0.0190	0.0272	0.0552	0.0595	0.1292	0.0058	0.0111
Fouv	0.0022	0.0024	0.0070	0.0041	0.0135	0.0202	0.1359	0.0038	0.0032
Fserv	0.0014	0.0019	0.0051	0.0036	0.0104	0.0140	0.0653	0.0075	0.0034
Faut	0.0000	0.0000	0.0002	0.0005	0.0003	0.0002	0.0003	0.0000	0.0010

Distributions conditionnelles

Distribution conditionnelle par ligne :

```
DI <- apply(P, 1, sum)
print(DI, digits = 3)
```

Fagri	Fouva	Fpat	Fsup	Fmoy	Femp	Fouv	Fserv	Faut
0.07846	0.00427	0.08291	0.04821	0.16667	0.31197	0.19231	0.11265	0.00256

Madame est employée dans 31.2% des couples.

Distribution conditionnelle par colonne :

```
DJ <- apply(P, 2, sum)
print(DJ, digits = 3)
```

Hagri	Houva	Hpat	Hsup	Hmoy	Hemp	Houv	Hserv	Haut
0.0844	0.0094	0.1031	0.1036	0.1458	0.1209	0.3867	0.0210	0.0251

Monsieur est employé dans 12.9% des couples.

Distributions conditionnelles par ligne

```
L <- P/DI
print(L, digits = 3)
```

	Hagri	Houva	Hpat	Hsup	Hmoy	Hemp	Houv	Hserv	Haut
Fagri	0.9150	0.00654	0.0174	0.00436	0.00436	0.00871	0.0436	0.00000	0.00000
Fouva	0.0800	0.48000	0.0000	0.00000	0.00000	0.04000	0.4000	0.00000	0.00000
Fpat	0.0186	0.00206	0.6866	0.07835	0.04948	0.04536	0.0990	0.01443	0.00619
Fsup	0.0106	0.00355	0.0638	0.56028	0.18794	0.06383	0.0851	0.00709	0.01773
Fmoy	0.0195	0.00410	0.0626	0.20615	0.31692	0.11590	0.2308	0.01436	0.02974
Femp	0.0110	0.00493	0.0608	0.08712	0.17699	0.19068	0.4142	0.01863	0.03562
Fouv	0.0116	0.01244	0.0364	0.02133	0.07022	0.10489	0.7067	0.01956	0.01689
Fserv	0.0121	0.01669	0.0455	0.03187	0.09256	0.12443	0.5797	0.06677	0.03035
Faut	0.0000	0.00000	0.0667	0.20000	0.13333	0.06667	0.1333	0.00000	0.40000

```
apply(L, 1, sum)
```

Fagri	Fouva	Fpat	Fsup	Fmoy	Femp	Fouv	Fserv	Faut
1	1	1	1	1	1	1	1	1

Quand Madame est cadre supérieur, dans 56% des cas Monsieur l'est aussi.

Distributions conditionnelles par colonne

```
C <- P/rep(DJ, rep(9, 9))
print(C, digits = 3)
```

	Hagri	Houva	Hpat	Hsup	Hmoy	Hemp	Houv	Hserv	Haut
Fagri	0.85020	0.0545	0.01327	0.00330	0.00234	0.00566	0.008842	0.0000	0.0000
Fouva	0.00405	0.2182	0.00000	0.00000	0.00000	0.00141	0.004421	0.0000	0.0000
Fpat	0.01822	0.0182	0.55224	0.06271	0.02814	0.03112	0.021220	0.0569	0.0204
Fsup	0.00607	0.0182	0.02985	0.26073	0.06213	0.02546	0.010610	0.0163	0.0340
Fmoy	0.03846	0.0727	0.10116	0.33168	0.36225	0.15983	0.099469	0.1138	0.1973
Femp	0.04049	0.1636	0.18408	0.26238	0.37866	0.49222	0.334218	0.2764	0.4422
Fouv	0.02632	0.2545	0.06799	0.03960	0.09261	0.16690	0.351459	0.1789	0.1293
Fserv	0.01619	0.2000	0.04975	0.03465	0.07151	0.11598	0.168877	0.3577	0.1361
Faut	0.00000	0.0000	0.00166	0.00495	0.00234	0.00141	0.000884	0.0000	0.0408

```
apply(C, 2, sum)
```

Hagri	Houva	Hpat	Hsup	Hmoy	Hemp	Houv	Hserv	Haut
1	1	1	1	1	1	1	1	1

Quand Monsieur est cadre supérieur, dans 26% des cas Madame l'est aussi.

Analyse d'inertie

Les coordonnées des lignes fournies par la procédure générale sont donc celles de l'analyse d'inertie du nuage des I distributions conditionnelles par ligne, pondéré et centré par la distribution marginale des lignes utilisant la métrique des inverses des poids des colonnes. Il s'en suit par symétrie que

les coordonnées des colonnes fournies par la procédure générale sont celles de l'analyse d'inertie du nuage des J distributions conditionnelles par colonne, pondéré et centré par la distribution marginale des colonnes utilisant la métrique des inverses des poids des lignes.

Analyse sous ade4

```
aco1 <- dudi.coa(mariages, scann = F, nf = 3)
aco1$li
```

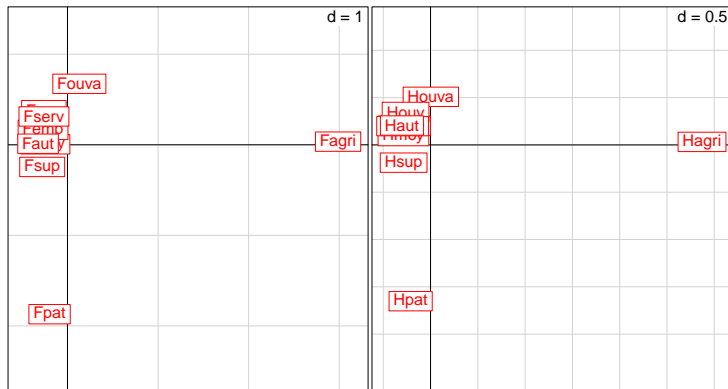
	Axis1	Axis2	Axis3
Fagri	2.9876578	0.034732908	0.01512670
Fouva	0.1310457	0.673421974	-1.11021775
Fpat	-0.1976602	-1.870622482	-0.41111744
Fsup	-0.2794160	-0.244576471	1.34651789
Fmoy	-0.2442007	0.005415925	0.56367956
Femp	-0.2698755	0.165143802	0.01656019
Fouv	-0.2605504	0.382632268	-0.45557477
Fserv	-0.2599946	0.312806843	-0.35717337
Faut	-0.3250941	0.011702601	0.57167553

```
aco1$co
```

	Comp1	Comp2	Comp3
Hagri	2.870290805	0.03197752	0.029405977
Houva	0.003024503	0.50859530	-0.794099946
Hpat	-0.210267176	-1.65236518	-0.376169437
Hsup	-0.284430764	-0.18751537	1.055072754
Hmoy	-0.286640500	0.09323807	0.466421889
Hemp	-0.277572109	0.20331653	0.003440163
Houv	-0.264128842	0.34854196	-0.339861174
Hserv	-0.295511528	0.19843526	-0.306386757
Haut	-0.301895712	0.20513307	0.154874185

Analyse sous ade4

```
par(mfrow = c(1, 2))
s.label(aco1$li)
s.label(aco1$co)
```



Relation entre les deux cartes

Analyse Factorielle des Correspondances



deux analyses d'inertie, deux nuages dans deux espaces
deux métriques et deux pondérations

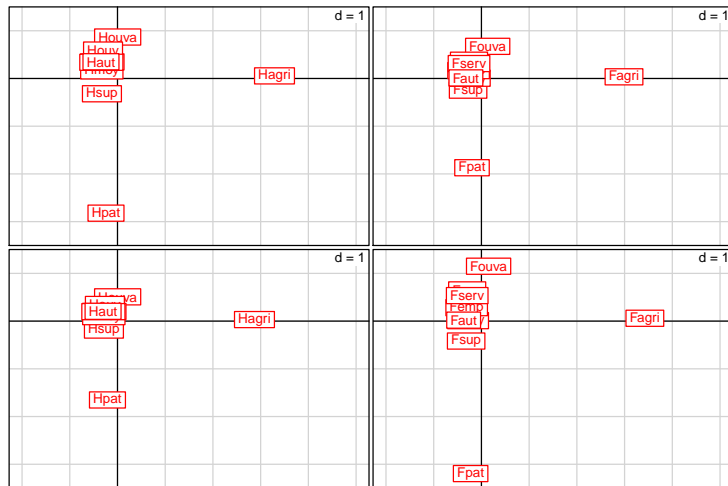
La procédure d'analyse des correspondances conduit à deux analyses d'inertie de deux nuages dans deux espaces avec deux métriques et deux pondérations, tous les paramètres utilisés dérivant du tableau initial. Les deux cartes des deux analyses retenues entretiennent des relations très fortes qui sont une des caractéristiques de l'AFC.

Moyennes conditionnelles

Les lignes positionnées par les coordonnées de variances λ_k (colonnes de $\tilde{\mathbf{C}}_K$) sont à la moyenne conditionnelle des colonnes positionnées par les coordonnées de variances 1 (colonnes de \mathbf{A}_k).

Réciproquement, les colonnes positionnées par les coordonnées de variances λ_k (colonnes de $\tilde{\mathbf{A}}_K$) sont à la moyenne conditionnelle des lignes positionnées par les coordonnées de variances 1 (colonnes de \mathbf{C}_K).

Relation entre les deux cartes



Dilatation (1)

On ne peut placer simultanément les lignes à la moyenne des colonnes et les colonnes à la moyenne des lignes.

Si on superpose les cartes construites avec les coordonnées de $\tilde{\mathbf{C}}_K$ et $\tilde{\mathbf{A}}_K$, on a une double relation de transition qui s'écrit :

$$\begin{cases} \tilde{C}_K(i, k) = \frac{1}{\sqrt{\lambda_k}} \sum_{j=1}^J p_{j/i} \tilde{A}_K(j, k) \\ \tilde{A}_K(j, k) = \frac{1}{\sqrt{\lambda_k}} \sum_{i=1}^I p_{i/j} \tilde{C}_K(i, k) \end{cases}$$

Dilatation (2)

Cette propriété est optimale, en ce sens que, si on a un code numérique \mathbf{x} des lignes \mathbf{D}_I -centré et un code numérique des colonnes \mathbf{y} \mathbf{D}_J -centré présentant une double relation en moyenne conditionnelle dilatée, alors :

$$\begin{cases} \mathbf{x}_i = a \sum_{j=1}^J p_{j/i} \mathbf{y}_j \\ \mathbf{y}_j = a \sum_{i=1}^I p_{i/j} \mathbf{x}_i \end{cases} \Rightarrow \begin{cases} \mathbf{x} = a \mathbf{P} \mathbf{D}_J^{-1} \mathbf{y} \\ \mathbf{y} = a \mathbf{P}^T \mathbf{D}_I^{-1} \mathbf{x} \end{cases} \Rightarrow \mathbf{x} = a^2 \mathbf{P} \mathbf{D}_J^{-1} \mathbf{P}^T \mathbf{D}_I^{-1} \mathbf{x}$$

La dilatation est la moins sensible si a est le plus proche possible de l'unité, ce qui est réalisé pour $a = \sqrt{\lambda_1}$ et les deux premiers scores de norme 1.

Représentation finale

