ANALYSE DES CORRESPONDANCES SIMPLE (AFC) ET MULTIPLE (ACM)

PLAN

L'analyse des correspondances simple : AFC

Méthode d'analyse de type ACP décrivant les liaisons entre 2 variables qualitatives. Elle utilise en entrée la table de contingence croisant ces deux variables.

Tableau de contingence N

Le tableau N ventile un échantillon de taille n selon les modalités de 2 variables qualitatives : l'une en ligne avec I modalités et l'autre en colonne avec J modalités.

Notations

$$
\begin{array}{ccccc|c}
n_{11} & \dots & n_{1j} & \dots & n_{1J} & n_{1.} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
n_{i1} & \dots & n_{ij} & \dots & n_{iJ} & n_{i.} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
n_{I1} & \dots & n_{Ij} & \dots & n_{IJ} & n_{I.} \\
\hline
n_{.1} & \dots & n_{.j} & \dots & n_{.J} & n
\end{array}
$$

avec

$$
\begin{array}{l}
\mathbf{I} = \{1, \dots, I\} \\
\mathbf{J} = \{1, \dots, J\} \\
n_{.j} = \sum_{i \in I} n_{ij} \\
n_{i.} = \sum_{j \in J} n_{ij} \\
n = \sum_{i \in I} \sum_{j \in J} n_{ij}
\end{array}
$$

Exemple de tableau de contingence

	dro	sci	let	med	pha	iut
age	4814	8941	8719	1583	959	558
oua	463	1118	1224	153	33	83
ind	3610	2608	3365	1634	728	110
art	2870	5091	5966	1583	562	261
gco	2777	2608	3212	1736	794	136
pco	4444	6209	7954	2655	1125	313
lib	9813	7823	9637	8985	3143	120
ing	4444	8195	7648	3880	1307	151
adm	14071	13535	18050	7147	2134	235
mes	833	993	1071	817	182	21
mso	555	621	765	357	66	16
tec	2129	3974	4436	1021	298	235
cam	7776	10058	13308	3931	1125	407
ins	1666	5464	6425	1481	579	141
emb	5184	8071	10861	2399	529	0
con	1481	2856	3059	613	132	177
ous	2314	4098	4436	766	182	516
min	278	869	918	153	17	37
mar	185	373	306	51	0	21
man	741	1863	1989	255	33	83
ser	741	993	1377	204	33	63
arm	2592	3104	3671	817	198	136

Tableau des fréquences relatives F

On considère le tableau F des fréquences relatives obtenu en divisant l'effectif n_{ij} de chaque cellule du tableau N par l'effectif total n.

$$
\begin{array}{ccccc|c}
f_{11} & \dots & f_{1j} & \dots & f_{1J} & f_{1.} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
f_{i1} & \dots & f_{ij} & \dots & f_{iJ} & f_{i.} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
f_{I1} & \dots & f_{Ij} & \dots & f_{IJ} & f_{I.} \\
\hline
f_{.1} & \dots & f_{.j} & \dots & f_{.J} & 1
\end{array}
$$

avec

$$
\begin{array}{l}
\mathbf{I} = \{1, \dots, I\} \\
\mathbf{J} = \{1, \dots, J\} \\
f_{ij} = n_{ij}/n \\
f_{.j} = \sum_{i \in I} f_{ij} \\
f_{i.} = \sum_{j \in J} f_{ij} \\
\sum_{i \in I} \sum_{j \in J} f_{ij} = 1 \\
\sum_{i \in I} f_{i.} = \sum_{j \in J} f_{j.} = 1
\end{array}
$$

Liaison entre variables qualitatives

On dit que deux variables sont liées (ou ne sont pas indépendantes) si la connaissance de la valeur de l'une modifie la distribution attendue de l'autre.

En terme de probabilité, l'indépendance s'écrit :

$\forall i \in I$ et $j \in J$:

$$\begin{aligned}
\Pr(i \text{ et } j) &= \Pr(i)\Pr(j) \\
\frac{\Pr(i \text{ et } j)}{\Pr(i)\Pr(j)} &= 1
\end{aligned}$$

On estime

$$\begin{array}{lll}
\Pr(i \text{ et } j) & \text{par} & f_{ij} = n_{ij}/n, \\
\Pr(i) & \text{par} & f_{i.} = n_{i.}/n \\
\Pr(j) & \text{par} & f_{.j} = n_{.j}/n.
\end{array}$$

Il y aura liaison entre les 2 variables si pour certains couples (i, j),

$$f_{ij} \neq f_{i.}f_{.j} \text{ significativement}$$

ou si (ce qui est équivalent)

$$\frac{f_{ij}}{f_{i.}f_{.j}} \neq 1 \text{ significativement}$$

Liaison entre variables qualitatives

Le vecteur $f_{ij}/f_{i.}$ (avec $j \in J$) est appelé le i*ème profil ligne*.

Le vecteur $f_{ij}/f_{.j}$ (avec $i \in I$) est appelé le j*ème profil colonne*.

En cas d'indépendance tous les profils ligne (respectivement profils colonne) sont égaux au profil marginal $f_{.j}$ avec $j \in J$ (respectivement $f_{i.}$ avec $i \in I$).

$$\frac{f_{ij}}{f_{i.}} = f_{.j} \qquad\qquad \frac{f_{ij}}{f_{.j}} = f_{i.}$$

Une mesure globale de la liaison

Sous l'hypothèse d'indépendance, la quantité :

$$\chi^2 = \sum_{i,j} \left[\frac{\left(n_{ij} - \frac{n_{i.}n_{.j}}{n}\right)^2}{\frac{n_{i.}n_{.j}}{n}} \right]$$

suit une loi du χ^2 à $(I-1)(J-1)$ degrés de liberté.

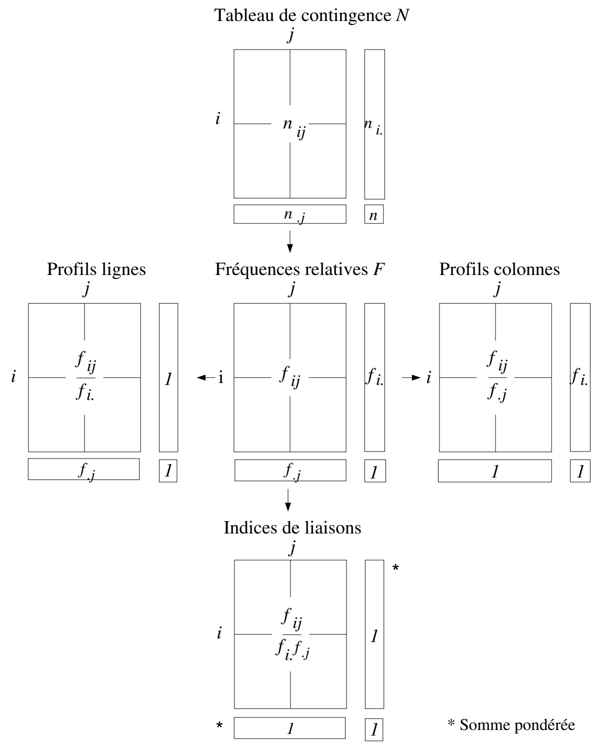
Cette quantité s'écrit encore:

$$\chi^2 = n \sum_{i,j} f_{i.} f_{.j} \left(\frac{f_{ij}}{f_{i.} f_{.j}} - 1 \right)^2$$

C'est une mesure de distance de la distribution observée à l'indépendance :

- Si le χ^2 est grand et supérieur à un seuil X_α, on rejette l'hypothèse d'indépendance : il y a liaison, due à des écarts significatifs entre $f_{ij}/(f_i f_j)$ et 1 pour au moins certains couples (i,j).
- Si le χ^2 est petit, on accepte l'hypothèse de non liaison : il n'y a rien à expliquer.

Tableaux dérivés de la table N



L'AFC définie comme une ACP

L'AFC est obtenue par l'ACP du tableau des indices de liaison $(f_{ij}/f_{i.}f_{.j})$ avec

- les poids $(f_{i.},\ i \in I)$: $D = D_I = diag(f_{i.})$,
- la métrique $M = D_J = diag(f_{.j})$.

Le tableau centré approximé par l'ACP a pour éléments $(f_{ij}/f_{i.}f_{.j} - 1)$ et a pour DVS :

$$f_{ij}/f_{i.}f_{.j} - 1 = \sum_{s=1}^{r} \lambda_s u_i^s v_j^s. \quad (1)$$

Dans cette ACP,

- Les facteurs $\mathbf{u}_s$ sont orthonormés pour la métrique D_I,
- les facteurs $\mathbf{v}_s$ sont orthonormés pour la métrique D_J,
- l'inertie totale du tableau est égale à la mesure globale de liaison :

$$\frac{\chi^2}{n} = \sum_{ij} f_{i.}f_{.j} \left(\frac{f_{ij}}{f_{i.}f_{.j}} - 1 \right)^2$$

On appelle *représentation simultanée* de l'AFC la superposition des représentations isométriques des lignes et des colonnes sur un même graphique. Dans ces représentations, la coordonnée de la ligne i (resp. colonne j) sur l'axe s est $\lambda_s u_i^s$ (resp. $\lambda_s v_j^s$).

Une définition équivalente de l'AFC

Autre façon de voir : l'AFC comme une double ACP :

- **L'ACP du tableau des profils lignes** $(f_{ij}/f_{i.},\ j \in J)$,
  - munis des poids $(f_{i.})$,
  - avec la métrique $D_J^{-1} = diag(1/f_{.j})$ (appellée *métrique du* χ^2).

  La DVS de cette ACP s'obtient en multipliant l'équation (1) de la DVS précédente par $f_{i.}$:

  $$f_{ij}/f_{i.} - f_{.j} = \sum_{s=1}^{r} \lambda_s u_i^s (f_{.j} v_j^s).$$

  Cette ACP fournit la même représentation des lignes que l'ACP des indices de liaison.

- **L'ACP du tableau des profils colonnes** $(f_{ij}/f_{.j},\ i \in I)$,
  - munis des poids $(f_{.j})$,
  - avec la métrique $D_I^{-1} = diag(1/f_{i.})$.

  La DVS de cette ACP s'obtient en multipliant l'équation (1) par $f_{.j}$:

  $$f_{ij}/f_{.j} - f_{i.} = \sum_{s=1}^{r} \lambda_s v_j^s (f_{i.} u_i^s).$$

  Cette ACP fournit la même représentation des colonnes que l'ACP des indices de liaison.

De plus ces trois ACP ont mêmes inerties totales, mêmes valeurs propres. Les contributions et indices de qualité sont aussi identiques.

Caractéristiques de l'AFC

- Le double centrage
- Les propriétés barycentriques
- l'équivalence distributionnelle

Double centrage

- L'ACP opère un centrage des variables par leurs moyennes.
- L'AFC opère un centrage sur les colonnes qui est aussi un centrage sur les lignes.

  En ACP, on est amené à étudier les proximités entre individus et les cosinus d'angles (les corrélations) entre variables centrées.

  En AFC, on n'analyse pas de corrélations mais la configuration de deux nuages centrés des lignes et des colonnes.

Propriétés barycentriques

Dans la base des vecteurs principaux $s = 1, \ldots, r$

- *le point ligne* i, de coordonnées $\lambda_s u_i^s = \sum_j \frac{f_{ij}}{f_{i.}} v_s^j$ est au barycentre des points j de coordonnées v_s^j. Chaque point j est affecté du poids $\frac{f_{ij}}{f_{i.}}$. Les poids sont les éléments du profil de la *ième* ligne.

- *le point colonne* j, de coordonnées $\lambda_s v_s^j = \sum_i \frac{f_{ij}}{f_{.j}} u_i^s$ est au barycentre des points de coordonnées u_i^s. Chaque point i est affecté du poids $\frac{f_{ij}}{f_{.j}}$. Les poids sont les éléments du profil de la *jème* colonne.

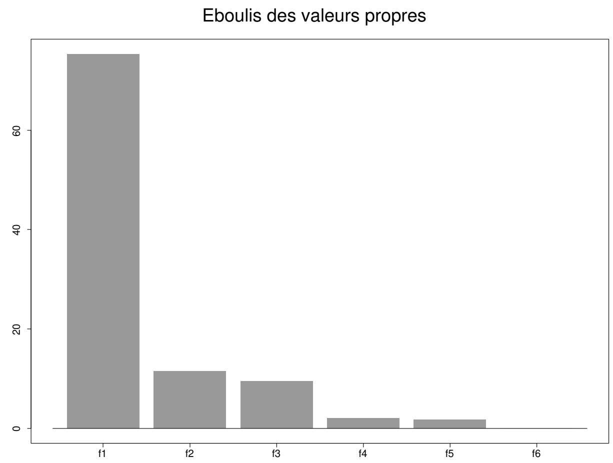
Cette propriété est à la base de l'interprétation barycentrique de l'AFC.

Équivalence distributionnelle

Cette propriété est une propriété de stabilité de l'AFC : si deux lignes de la table de contingence sont proportionnelles, on ne modifie pas l'AFC de la table en remplaçant les deux lignes par leur somme. La propriété est vraie aussi pour les colonnes.

Cette propriété permet de regrouper des catégories d'une variable qui sont homogènes vis-à-vis des catégories de l'autre variable (mêmes profils).

Pratique de l'AFC



facteurs	f1	f2	f3	f4	f5	tot
valeurs singulieres	0.2129	0.0832	0.0755	0.0346	0.0325	
valeurs propres	0.0453	0.0069	0.0057	0.0012	0.0011	0.0602
inertie exp	75.3 %	11.5 %	9.5 %	2.0 %	1.8 %	100 %
inertie cum	75.3 %	86.8 %	96.3 %	98.2 %	100 %	100 %

Cosinus carre lignes
axes principaux

	f1	f2	f3	f4	f5	TOT
age	731	10	161	96	2	1000
oua	957	8	25	9	0	1000
ind	615	264	71	49	2	1000
art	870	97	7	3	23	1000
gco	821	0	119	25	35	1000
pco	3	171	85	7	734	1000
lib	976	19	3	2	0	1000
ing	134	704	17	22	123	1000
adm	437	484	73	0	6	1000
mes	740	82	4	100	73	1000
mso	355	113	82	258	191	1000
tec	945	18	27	10	1	1000
cam	441	314	92	102	52	1000
ins	498	361	108	0	32	1000
emb	416	36	547	0	0	1000
con	941	1	27	11	19	1000
ous	606	3	341	50	0	1000
min	883	76	19	10	12	1000
mar	745	5	43	1	207	1000
man	969	9	8	1	13	1000
ser	741	223	0	27	9	1000
arm	415	526	3	4	53	1000
TOT	753	115	95	20	18	1000

Cosinus carre lignes
sous-espaces principaux

	Axe 1	Axes1a2	Axes1a3	Axes1a4	TOT
age	731	741	902	998	1000
oua	957	965	991	1000	1000
ind	615	879	950	998	1000
art	870	968	974	977	1000
gco	821	821	940	965	1000
pco	3	174	259	266	1000
lib	976	995	998	1000	1000
ing	134	838	855	877	1000
adm	437	921	994	994	1000
mes	740	822	827	927	1000
mso	355	468	551	809	1000
tec	945	962	989	999	1000
cam	441	755	847	948	1000
ins	498	859	968	968	1000
emb	416	452	1000	1000	1000
con	941	942	969	981	1000
ous	606	609	950	1000	1000
min	883	959	978	988	1000
mar	745	749	793	793	1000
man	969	978	987	987	1000
ser	741	964	964	991	1000
arm	415	941	943	947	1000
TOT	753	868	963	982	1000

Cosinus carre colonnes axes principaux

	f1	f2	f3	f4	f5	TOT
dro	478	489	23	4	5	1000
sci	830	118	2	14	36	1000
let	805	8	125	11	52	1000
med	932	46	2	18	2	1000
pha	808	85	33	46	29	1000
iut	376	2	594	23	4	1000
TOT	753	115	95	20	18	1000

Cosinus carre colonnes sous-espaces principaux

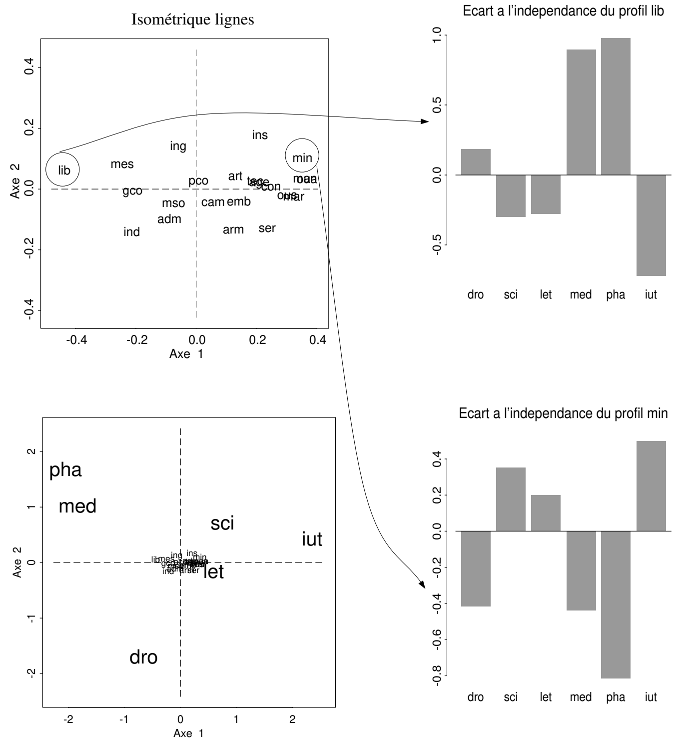
	f1	f1a2	f1a3	f1a4	TOT
dro	478	967	991	995	1000
sci	830	948	950	964	1000
let	805	812	937	948	1000
med	932	978	979	998	1000
pha	808	892	925	971	1000
iut	376	378	972	996	1000
TOT	753	868	963	982	1000

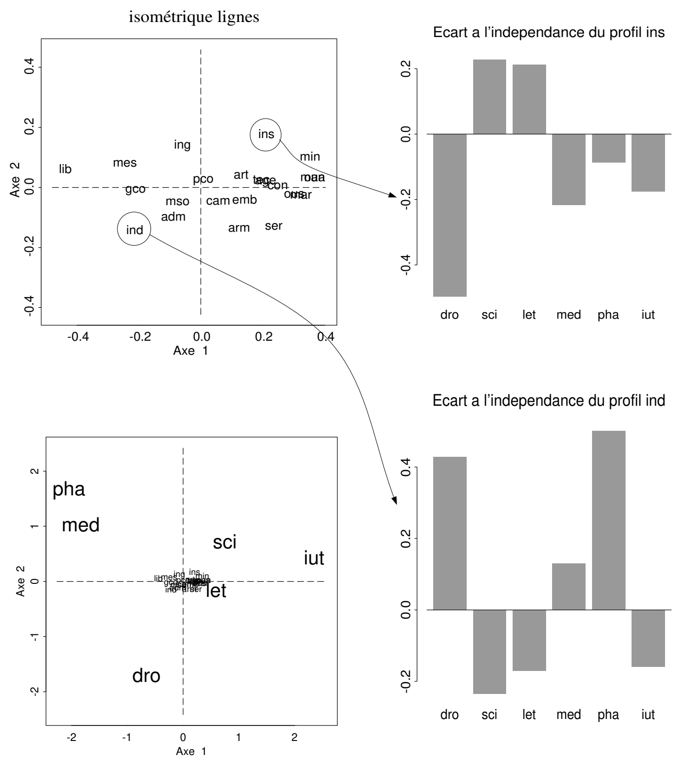
Contribution des lignes aux facteurs

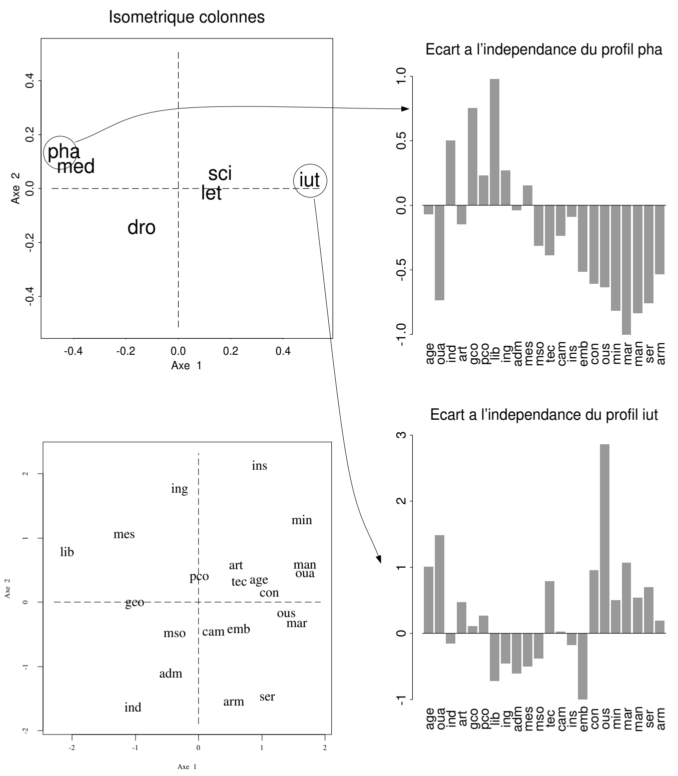
	f1	f2	f3	f4	f5	TOT	ORI
age	67	6	118	337	7	69	36
oua	25	1	5	9	0	20	86
ind	35	99	32	106	4	43	48
art	17	12	1	2	19	14	12
gco	33	0	37	38	60	30	35
pco	0	8	5	2	217	5	3
lib	479	61	12	40	1	370	124
ing	6	213	6	38	246	35	18
adm	30	217	40	1	18	52	12
mes	15	11	1	78	65	16	53
mso	1	2	2	26	22	2	11
tec	28	3	6	11	2	22	24
cam	6	28	10	53	31	10	4
ins	43	204	74	2	119	65	55
emb	32	18	329	0	1	57	28
con	31	0	7	14	28	25	40
ous	68	2	304	214	0	85	91
min	17	10	3	7	10	15	86
mar	6	0	3	0	71	6	86
man	39	2	3	1	22	31	82
ser	12	23	0	16	6	12	46
arm	9	78	1	3	51	17	21
TOT	1000	1000	1000	1000	1000	1000	1000

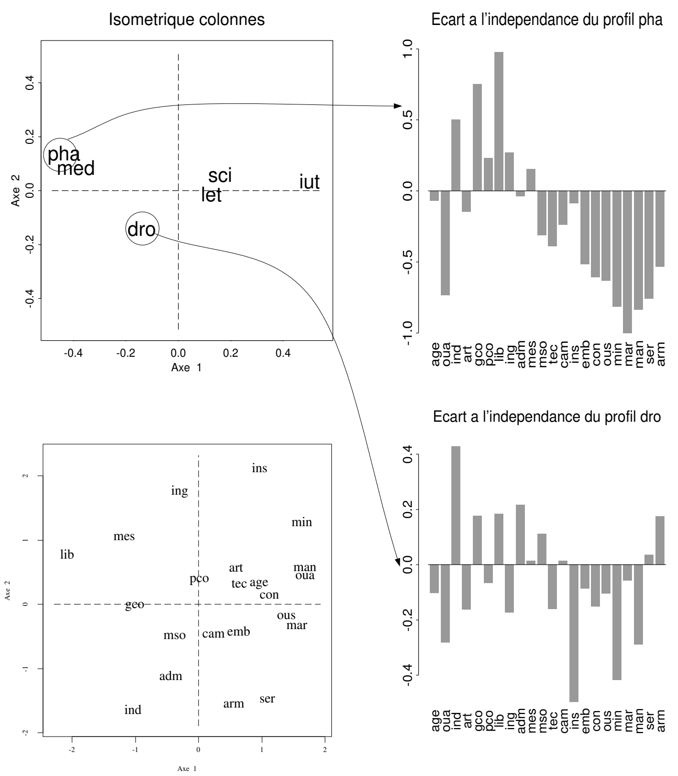
Contribution des colonnes aux facteurs

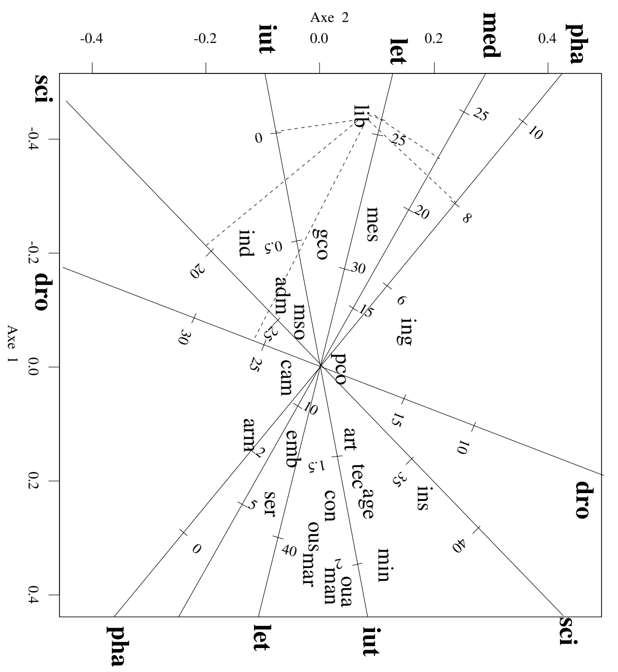
	f1	f2	f3	f4	f5	TOT	ORI
dro	89	596	35	31	40	140	35
sci	161	150	3	102	302	146	27
let	119	7	147	60	331	111	17
med	403	129	6	301	40	325	141
pha	168	115	55	365	256	157	203
iut	60	2	755	141	30	120	577
	1000	1000	1000	1000	1000	1000	1000


Isométrique lignes
Axe 1
Axe 2
lib
mes
ing
ins
min
gco
pco
art
mso
cam
emb
adm
ind
arm
ser
Ecart a l'independance du profil lib
dro
sci
let
med
pha
iut
pha
med
sci
iut
let
dro
Ecart a l'independance du profil min

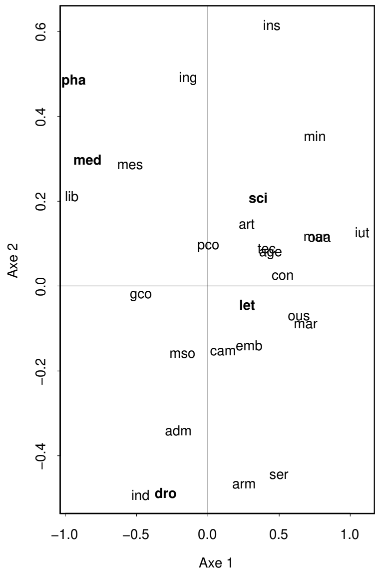

isométrique lignes
Axe 1
Axe 2
lib
mes
ing
ins
min
gco
pco
art
mso
cam
emb
adm
ind
arm
ser
Ecart a l'independance du profil ins
dro
sci
let
med
pha
iut
pha
med
sci
iut
let
dro
Ecart a l'independance du profil ind


Isometrique colonnes
pha
med
sci
let
iut
dro
Axe 1
Axe 2
Ecart a l'independance du profil pha
Ecart a l'independance du profil iut
age oua ind art gco pco lib ing adm mes mso tec cam ins emb con ous min mar man ser arm
lib mes ing ins min geo pco art tec age con man oua mso cam emb ous mar adm ind arm ser

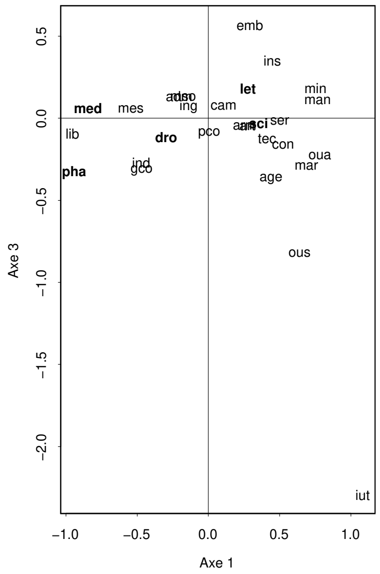

Isometrique colonnes
pha
med
sci
let
iut
dro
Axe 1
Axe 2
Ecart a l'independance du profil pha
Ecart a l'independance du profil dro
age oua ind art gco pco lib ing adm mes mso tec cam ins emb con ous min mar man ser arm
lib mes ing ins min geo pco art tec age con man oua mso cam emb ous mar adm ind arm ser



Représentation simultanée (1-2)



Représentation simultanée (1-3)



AFC - Quelques formules

Soit une table de contingence $I \times J$ à i et j modalités où n_{ij} représente les effectifs d'individus partageant la modalité i et la modalité j.

L'analyse des lignes

Profil ligne moyen : $f_{.j} = n_{.j}/n, \ j = 1 \ldots, J$

Profil de la ligne i : $f_{ij}/f_{i.} = n_{ij}/n_{i.} = \Pr(j|i), \ j = 1 \ldots, J$

Écart à l'indépendance du profil de la ligne i : $f_{ij}/f_{i.} - f_{.j}, \ j = 1 \ldots, J$

Distance du χ^2 de la ligne i au centre du nuage des lignes :

$$\sum_j \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - f_{.j} \right)^2 = \sum_j f_{.j} \left(\frac{f_{ij}}{f_{i.}f_{.j}} - 1 \right)^2 = \sum_s \left(\lambda_s u_i^s \right)^2$$

Contribution à l'inertie de la ligne i :

$$\sum_j f_{i.}f_{.j} \left(\frac{f_{ij}}{f_{i.}f_{.j}} - 1 \right)^2 = f_{i.} \sum_s \left(\lambda_s u_i^s \right)^2$$

Cosinus carré de la ligne i à l'axe s : $\dfrac{\left(\lambda_s u_i^s \right)^2}{\sum_s \left(\lambda_s u_i^s \right)^2}$

L'analyse des colonnes

Profil colonne moyen : $f_{i.} = n_{i.}/n, \ i = 1 \ldots, I$

Profil de la colonne j : $f_{ij}/f_{.j} = n_{ij}/n_{.j} = \Pr(i|j)$

Écart à l'indépendance du profil de la colonne j : $f_{ij}/f_{.j} - f_{i.}, \ i = 1 \ldots, I$

Distance du χ^2 de la colonne j au centre du nuage des lignes :

$$\sum_i \frac{1}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - f_{i.} \right)^2 = \sum_i f_{i.} \left(\frac{f_{ij}}{f_{i.}f_{.j}} - 1 \right)^2 = \sum_s \left(\lambda_s v_s^j \right)^2$$

Contribution à l'inertie de la colonne j :

$$\sum_i f_{i.}f_{.j} \left(\frac{f_{ij}}{f_{i.}f_{.j}} - 1 \right)^2 = f_{.j} \sum_s \left(\lambda_s v_s^j \right)^2$$

Cosinus carré de la colonne j à l'axe s : $\dfrac{\left(\lambda_s v_s^j \right)^2}{\sum_s \left(\lambda_s v_s^j \right)^2}$

L'analyse des correspondances multiples : ACM

- Elle permet d'étudier les liaisons entre plus de deux variables qualitatives, c'est une généralisation de l'AFC.
- Pratiquement elle se ramène à l'AFC d'un tableau particulier (le tableau de Burt ou le tableau d'indicatrices) et elle ne prend en compte que les liaisons deux à deux entre les variables qualitatives.
- Objectifs :

  . but descriptif :
  bilan des liaisons entre les variables
  bilan des ressemblances entre modalités
  bilan des ressemblances entre individus (typologie)

  . but de compression ou de recodification :
  réduction du nombre de variables ou de modalités.

Un exemple de données

57 variétés de pommes sont caractérisées par 9 variables qualitatives codées :

- *arb* = forme de l'arbre, caractérisée par 4 modalités (1 = colonnaire ; 2 = spur ; 3 = étalé ; 4 = très étalé)
- *rec* = date de récolte, caractérisée par 3 modalités (1 = précoce ; 2 = intermédiaire ; 3 = tardif ou très tardif)
- *cal* = calibre du fruit, caractérisé par 3 modalités (1 = petit ou moyen ; 2 = gros ; 3 = très gros)
- *coul* = couleur dominante du fruit ou intensité de la couleur rouge additionnelle, caractérisée par 4 modalités (1 = jaune ou vert ; 2 = rouge-orangé ; 3 = rouge; 4 = rouge-violacé)
- *pour* = proportion de coloration rouge additionnelle, caractérisée par 4 modalités (1 = absente (0-25 %) ; 2 = faible (25-50 %) ; 3 = moyenne (50-75 %) ; 4 = forte (75-100 %))
- *type* = type de coloration du fruit, caractérisé par 3 modalités (1 = lavé ; 2 = lavé-strié; 3 = strié)
- *form* = forme du fruit, caractérisée par 3 modalités (1 = allongé ; 2 = intermédiaire ; 3 = aplati)
- *ferm* = fermeté du fruit, caractérisée par 3 modalités (1 = peu ferme ; 2 = moyen ; 3 = très ferme)
- *gout* = rapport sucre / acidité, caractérisé par 3 modalités (1 = doux ; 2 = équilibré ; 3 = acidulé)

Extrait du tableau individus $\times$ variables

	arb	rec	cal	coul	pour	type	form	ferm	gout
Granny-spur	1	3	2	1	1	1	2	3	3
Granny-stand.	4	3	2	1	1	1	2	3	3
Boskoop-mut.	3	3	3	4	4	1	1	2	2
Boskoop-stand.	2	3	3	4	3	1	1	2	2
Gala-red	2	2	1	3	3	3	2	2	1
Gala	3	2	1	3	2	3	2	2	1
Elstar	3	1	1	2	2	3	2	2	3
Elista	3	1	1	3	3	1	2	2	3
Jonagold	3	2	3	3	2	2	2	2	3
Jonagored	3	2	3	4	3	3	2	2	3
New-Jonagold	3	2	3	2	3	1	2	2	3
Gloster	2	3	2	3	4	1	1	3	2
Gloster-Spur	1	3	2	3	4	1	1	3	2
Averdal	3	2	2	4	4	2	1	2	1
StarKrimson	1	2	2	4	4	3	1	2	1
Erovan	2	2	2	4	4	1	1	2	1
Goldenspur	1	2	1	1	1	1	1	2	2
Quemoni	3	2	1	1	2	1	1	2	2
Golden-972	3	2	1	1	1	1	1	2	2
Sinta	2	2	1	3	2	1	3	2	2
Fiesta	3	1	1	2	3	3	2	3	2
Arlet	3	1	1	3	3	2	2	2	2
Hilrome	4	3	3	3	4	1	2	3	2
Rome-Beauty	4	3	3	3	3	3	2	3	2
Melrose-mut.	3	3	2	2	4	1	3	3	2
Melrose	3	3	3	2	3	3	3	3	2

Tableau d'indicatrices Z

	arb				rec			cal			
	arb1	arb2	arb3	arb4	rec1	rec2	rec3	cal1	cal2	cal3	...
Granny-spur	1	0	0	0	0	0	1	0	1	0	...
Granny-stand.	0	0	0	1	0	0	1	0	1	0	...
Boskoop-mut.	0	0	1	0	0	0	1	0	0	1	...
Boskoop-stand.	0	1	0	0	0	0	1	0	0	1	...
Gala-red	0	1	0	0	0	1	0	1	0	0	...
Gala	0	0	1	0	0	1	0	1	0	0	...
Elstar	0	0	1	0	1	0	0	1	0	0	...
Elista	0	0	1	0	1	0	0	1	0	0	...
Jonagold	0	0	1	0	0	1	0	0	0	1	...
Jonagored	0	0	1	0	0	1	0	0	0	1	...
New-Jonagold	0	0	1	0	0	1	0	0	0	1	...
...	...	...	...	...	...	...	...	...	...	...	...

Notations

Questions	1					k			K					Total
Réponses	1	.	j	.	m_1	.	j	.	1	.	j	.	m_K	$m = \sum_{k=1}^{K} m_k$
1	1	.	0	.	0				0	.	0	.	1	K
.	.	.	.	.	.				.	.	.	.	.	.
Individus i	.	.	.	.	.				.	.	.	.	.	K
.	.	.	.	.	.				.	.	.	.	.	.
n	0	.	1	.	0				1	.	0	.	0	K
Total	n_{11}	.	n_{j1}	.	$n_{m_1 1}$	.	n_{jk}	.	n_{1K}	.	n_{jK}	.	$n_{m_K K}$	$n \times K$

$$\sum_{j=1}^{m_k} n_{jk} = n$$

Tableau de Burt B

		arb				rec			cal			
		arb1	arb2	arb3	arb4	rec1	rec2	rec3	cal1	cal2	cal3	...
arb	arb1	4	0	0	0	0	2	2	1	3	0	...
	arb2	0	10	0	0	0	8	2	5	2	3	...
	arb3	0	0	34	0	12	14	8	22	7	5	...
	arb4	0	0	0	9	1	3	5	2	3	4	...
rec	rec1	0	0	12	1	13	0	0	9	4	0	...
	rec2	2	8	14	3	0	27	0	15	6	6	...
	rec3	2	2	8	5	0	0	17	6	5	6	...
cal	cal1	1	5	22	2	9	15	6	30	0	0	...
	cal2	3	2	7	3	4	6	5	0	15	0	...
	cal3	0	3	5	4	0	6	6	0	0	12	...
coul	coul1	2	1	8	2	1	8	4	8	4	1	...
	coul2	0	1	8	1	5	2	3	5	3	2	...
	coul3	1	6	14	3	6	11	7	16	4	4	...
	coul4	1	2	4	3	1	6	3	1	4	5	...
pour	pour1	2	0	3	2	0	3	4	4	3	0	...
	pour2	0	3	10	0	3	9	1	9	2	2	...
	pour3	0	4	14	2	7	8	5	12	2	6	...
	pour4	2	3	7	5	3	7	7	5	8	4	...
type	type1	3	7	16	6	5	14	13	16	8	8	...
	type2	0	1	8	1	4	5	1	6	3	1	...
	type3	1	2	10	2	4	8	3	8	4	3	...
form	form1	3	4	11	1	3	11	5	7	8	4	...
	form2	1	4	18	7	7	13	10	18	6	6	...
	form3	0	2	5	1	3	3	2	5	1	2	...
ferm	ferm1	0	1	2	0	1	2	0	1	1	1	...
	ferm2	2	8	25	3	11	24	3	23	8	7	...
	ferm3	2	1	7	6	1	1	14	6	6	4	...
gout	gout1	1	2	6	1	1	6	3	6	3	1	...
	gout2	2	8	19	3	9	15	8	17	8	7	...
	gout3	1	0	9	5	3	6	6	7	4	4	...

ACM : AFC du tableau des indicatrices Z
Distance entre individus

Un individu est caractérisé par l'ensemble des valeurs des réponses qu'il a fournies aux K différentes questions.

Exemples :

– variétés proches : Baujade et Granny-standard

	arb	rec	cal	coul	pour	type	form	ferm	gout
Codage	1234	123	123	1234	1234	123	123	123	123
Baujade	0001	001	100	1000	1000	100	010	001	001
Granny-stand.	0001	001	010	1000	1000	100	010	001	001

– variétés éloignées : Granny-spur et Vista-bella

	arb	rec	cal	coul	pour	type	form	ferm	gout
Codage	1234	123	123	1234	1234	123	123	123	123
Granny-spur	1000	001	010	1000	1000	100	010	001	001
Vista-Bella	0010	100	100	0010	0001	010	001	100	010

Matrice $D_J = diag(\frac{n_{jk}}{n \times K})$

Carré de la distance entre 2 individus i et i' :

$$d^2(i,i') = \sum_{k \in \Delta_{ii'}} \frac{n}{K} \left(\frac{1}{n_{jk}} + \frac{1}{n_{j'k}} \right)$$

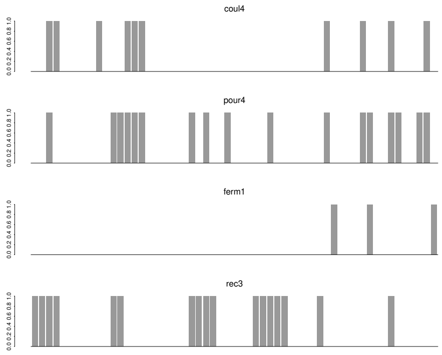
avec $\Delta_{ii'} \subset \{1, \ldots, K\}$ ensemble des questions auxquelles i et i' ont donné des réponses différentes ($jk \neq j'k$)

Distance entre réponses

La réponse à une question est caractérisée par l'ensemble des valeurs fournies par les individus.

Exemples :

profils proches coul4 et pour4 (couleur rouge-violacé et fort pourcentage de coloration additionnelle)
profils éloignés ferm1 et rec3 (fruit peu ferme et récolte tardive)



Matrice $D_I = \frac{1}{n} I_n$ car le poids des lignes est constant et égal à K.

Carré de la distance entre la réponse j à la question k et la réponse j' à la question k' :

$$d^2(j, j') = \frac{n}{n_{jk} n_{j'k'}} \delta_{kk'}$$

$\delta_{kk'}$ = nbre d'individus ayant répondu j à la question k mais pas j' à la question k', ou ayant répondu j' à la question k' mais pas j à la question k.

AFC du tableau Z

L'AFC de Z s'obtient par la DVS de :

$$X = nF_z D_J^{-1} - E = \frac{1}{K} Z D_J^{-1} - E = \sum_s \lambda_s u_s {}^t v_s$$

pour les métriques D_I et D_J ; les vecteurs u_s sont D_I orthonormés, le vecteurs v_s sont D_J orthonormés.

n est le nombre d'individus

$F_z = \frac{1}{nK} Z$ est le tableau de fréquences associé à Z

D_J est la matrice diagonale des poids des colonnes de Z

K est le nombre de questions

$D_I = \frac{1}{n} I_n$ est la matrice diagonale des poids des lignes de Z

E est la matrice de dimensions $m \times n$ dont tous les éléments sont égaux à 1

Résultats de l'analyse

– Inertie globale de l'AFC du tableau Z : $\dfrac{m-K}{K}$

– Contribution d'une modalité à l'inertie :

$\dfrac{1}{K}\left(1-\dfrac{n_{jk}}{n}\right)$ = contribution de la réponse j à la question k

K = nombre de questions (variables)

m = nombre total de réponses

n = nombre d'individus

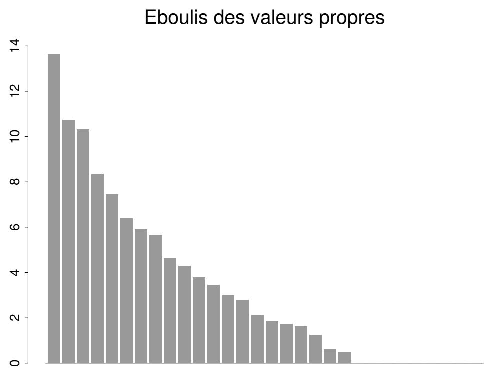
n_{jk} = nombre d'individus ayant répondu j à la question k

– Contribution d'une question à l'inertie :

$\dfrac{1}{K}(m_k - 1)$ = contribution de la question k

Résultats de l'exemple

```
Inertie totale:  2.33

Valeurs propres

 f1   f2   f3   f4   f5   f6   f7   f8   f9   f10  f11  f12  f13  f14   f15
0.32 0.25 0.24 0.19 0.17 0.15 0.14 0.13 0.11 0.10 0.09 0.08 0.07 0.065 0.05

 f16  f17  f18  f19  f20  f21   f22     f23     f24     f25     f26     f27
0.04 0.04 0.04 0.03 0.01 0.01 1.5e-16 8.9e-17 7.7e-17 5.6e-17 4.4e-17 3.4e-17

  f28     f29     f30
2.9e-17 1.9e-18 -2.3e-17
```



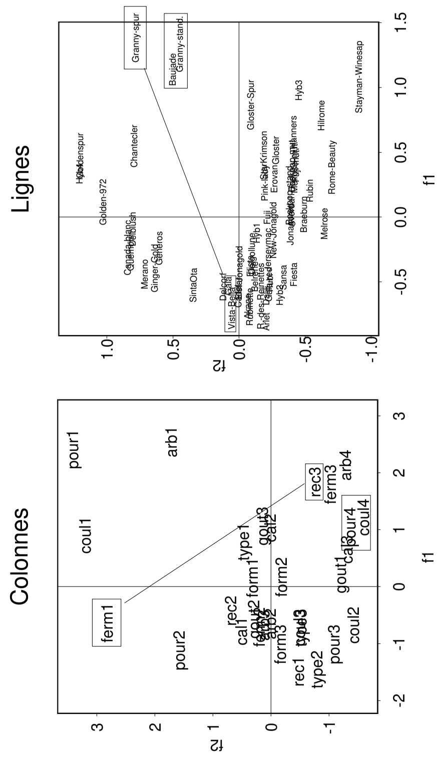

```
Pourcentage d'inertie expliquee par les facteurs:

                      f1 f2 f3 f4 f5 tot
     inertie expliquee  14 11 10  8  7  51
     inertie cumulee    14 24 35 43 51  51
```

Contribution des modalités et des variables aux facteurs et à l'inertie totale:

	f1	f2	f3	f4	f5	f>5	TOT	ORI
arb1	51	23	15	12	106	49	44	116
arb2	8	0	82	18	82	45	39	41
arb3	30	1	21	5	17	23	19	6
arb4	80	29	11	4	29	45	40	47
arb	169	53	129	39	234	162	142	210
rec1	56	5	36	2	123	31	37	30
rec2	9	25	65	21	41	19	25	10
rec3	108	18	24	22	3	25	33	21
rec	173	48	125	45	167	75	95	61
cal1	35	15	19	3	4	27	23	8
cal2	31	0	12	3	183	32	35	25
cal3	10	40	8	23	147	37	38	33
cal	76	55	39	29	334	96	96	66
coul1	21	261	7	0	1	10	37	30
coul2	9	39	91	2	18	47	39	41
coul3	24	11	1	11	4	44	28	12
coul4	29	49	121	14	0	33	39	41
coul	83	360	220	27	23	134	143	124

pour1	80	159	26	0	5	22	42	63
pour2	32	64	1	15	46	42	37	30
pour3	40	45	28	1	1	35	31	16
pour4	37	60	71	5	30	24	33	21
pour	189	328	126	21	82	123	143	130
type1	39	14	6	16	7	23	21	7
type2	41	12	1	16	57	54	39	41
type3	15	7	8	83	6	49	35	25
type	95	33	15	115	70	126	95	73
form1	1	4	96	13	43	34	32	18
form2	2	2	68	28	32	21	23	8
form3	16	0	1	247	1	36	41	54
form	19	6	165	288	76	91	96	80
ferm1	2	46	3	183	0	49	45	158
ferm2	37	3	17	40	0	11	16	4
ferm3	101	32	51	15	0	21	34	23
ferm	140	81	71	238	0	81	95	185
gout1	1	28	34	89	0	51	39	41
gout2	20	6	6	81	5	20	21	7
gout3	33	1	70	30	9	40	35	25
gout	54	35	110	200	14	111	95	73
	1000	1000	1000	1000	1000	1000	1000	1000

Biplot isométrique ligne



Variables illustratives

On enrichit l'interprétation d'une ACM en rajoutant des variables supplémentaires ou illustratives.

Une réponse étant au barycentre des individus qui l'ont donnée, il est facile de représenter par ce moyen des réponses supplémentaires. Si ces réponses sont suffisamment éloignées du barycentre, alors cette question est liée aux variables actives, ou plus précisément, au résumé qui en est fait par les premiers facteurs.

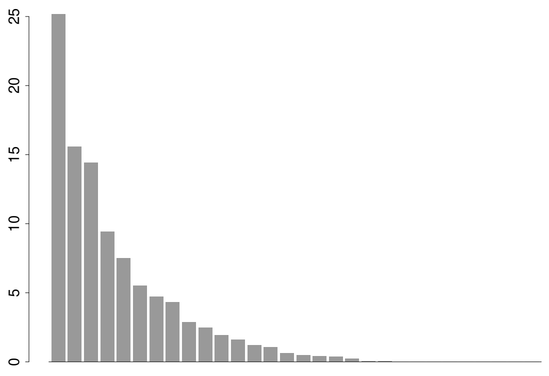
Analyse du tableau de Burt B

- Profils : le tableau étant symétrique, les profils lignes sont identiques aux profils colonnes. Une seule ACP sera nécessaire pour l'analyse.
- Distances : les distances sont déformées régulièrement et faiblement si λ_1 est proche de λ_2.
- Métrique : D_B matrice diagonale du poids des lignes et des colonnes
- Les analyses (AFC) de B et de Z sont telles que les valeurs propres de l'AFC de B sont celles de l'AFC de Z au carré.
- Les pourcentages d'inertie expliquée sont plus forts que dans l'ACM de Z (nuage plus allongé sur le $1er$ axe.

Résultats de l'ACM sur B

```
Inertie totale:  0.4

Valeurs propres :

       f1         f2         f3        f4         f5         f6         f7
0.1013181 0.06278225 0.05808789 0.0379203 0.03026723 0.02215653 0.01903697
       f8         f9        f10         f11        f12         f13
0.0173106 0.01164472 0.01000652 0.007828113 0.00646734 0.004877728
        f14         f15         f16         f17       f18          f19
0.004236716 0.002454416 0.001881952 0.001618166 0.0014098 0.0008408889
         f20          f21          f22          f23          f24          f25
0.0001971365 0.0001175645 6.301526e-18 4.152271e-18 3.240682e-18 1.861229e-18
         f26           f27           f28         f29          f30
4.874816e-19 -1.654147e-18 -3.902081e-18 -8.2177e-18 -9.546523e-18
```

Eboulis des valeurs propres



Pourcentage d'inertie expliquée par les facteurs :

	f1	f2	f3	f4	f5	tot
inertie expliquee	25	16	14	9	8	72
inertie cumulee	25	41	55	65	72	72

Contribution des modalités aux facteurs

	f1	f2	f3	f4	f5	f>5	TOT	ORI
arb1	51	23	15	12	106	45	40	108
arb2	8	0	82	18	82	43	34	36
arb3	30	1	21	5	17	21	18	6
arb4	80	29	11	4	29	43	41	49
rec1	56	5	36	2	123	27	37	31
rec2	9	25	65	21	41	16	25	10
rec3	108	18	24	22	3	25	43	27
cal1	35	15	19	3	4	29	23	8
cal2	31	0	12	3	183	28	32	23
cal3	10	40	8	23	147	40	34	31
coul1	21	261	7	0	1	6	48	40
coul2	9	39	91	2	18	48	36	39
coul3	24	11	1	11	4	54	24	11
coul4	29	49	121	14	0	33	43	46
pour1	80	159	26	0	5	16	53	82
pour2	32	64	1	15	46	43	35	29
pour3	40	45	28	1	1	39	32	17
pour4	37	60	71	5	30	22	38	24
type1	39	14	6	16	7	22	21	7
type2	41	12	1	16	57	58	34	37
type3	15	7	8	83	6	54	30	21
form1	1	4	96	13	43	36	29	17
form2	2	2	68	28	32	20	21	8
form3	16	0	1	247	1	31	36	49
ferm1	2	46	3	183	0	51	40	142
ferm2	37	3	17	40	0	11	19	5
ferm3	101	32	51	15	0	22	45	31
gout1	1	28	34	89	0	59	34	37
gout2	20	6	6	81	5	19	20	7
gout3	33	1	70	30	9	41	34	24
	1000	1000	1000	1000	1000	1000	1000	1000

Représentation Isométrique ligne (ou colonne)