

Analyse Factorielle des Correspondances

Tableau de contingence

Analyse des profils - Analyse des Associations

Décomposition orthogonale du χ^2

Ricco RAKOTOMALALA

Université Lumière Lyon 2



Origine sociale (CSP parents) vs. Choix d'études à l'Université

F.G. Carpentier

<http://geai.univ-brest.fr/~carpentier/>

CSP\Filière	Droit	Sciences	Médecine	IUT
Exp.agri	80	99	65	58
Patron	168	137	208	62
Cadre.sup	470	400	876	79
Employé	145	133	135	54
Ouvrier	166	193	127	129

Pourquoi croiser ? L'une des variables peut emmener de l'information sur l'autre (Ex. influence de la CSP sur les choix de filières).

Quelques questions simples pour caractériser la relation :

1. Structure des filières choisies selon la CSP (**profils** lignes)
2. Différences/ressemblances des structures de choix selon les CSP (**distance**)
3. Idem – Structure des CSP des parents selon les filières (profils colonnes)
4. **Associations** entre CSP et FILIERES : Qui choisit quoi ? Quoi est choisi par qui ?



Notations – Tableau de contingence

CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	80	99	65	58	302
Patron	168	137	208	62	575
Cadre.sup	470	400	876	79	1825
Employé	145	133	135	54	467
Ouvrier	166	193	127	129	615
Total	1029	962	1411	382	3784

Y / X	x_1	x_l	x_L	Σ
y_1		\vdots		
y_k	\dots	n_{kl}	\dots	$n_{k.}$
		\vdots		
y_K				
Σ		$n_{.l}$		n

$P(\text{Cadre.Sup}) = 1825 / 3784 = 48.2\% \rightarrow$ Proportion des enfants de « Cadre Sup. » parmi les enquêtés

$P(\text{Médecine}) = 1411 / 3784 = 37.3\% \rightarrow$ Proportion des enquêtés ayant choisi la filière « Médecine »

$P(\text{Médecine \& Cadre.Sup}) = 876 / 3784 = 23.2\% \rightarrow$ Proportion des personnes « ayant choisi 'Médecine' ET qui sont enfants de 'Cadre.sup' » (idée de concomitance, association)

$P(\text{Médecine / Cadre.Sup}) = 876 / 1825 = 48.0\% \rightarrow$ Proportion des personnes « ayant choisi le 'Médecine' SACHANT QU'ils sont enfants de 'Cadre.Sup' » (idée de causalité)

PLAN

1. Analyse des profils lignes - Distance entre profils – Distance à l'origine - Inertie
2. Analyse des profils colonnes
3. Analyse des associations lignes / colonnes- KHI-2 d'écart à l'indépendance
4. Décomposition du KHI-2 - Contributions, Résidus
5. Analyse factorielle des correspondances
6. Les logiciels (SPAD, SAS, Tanagra, R)
7. Conclusion
8. Bibliographie



Profils lignes

Définition - Distance entre profils - Distance à l'origine



Profils lignes – Distance entre profils (Distance du KHI-2)

$$P(X = l / Y = k) = \frac{n_{kl}}{n_{k.}}$$

CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	0.265	0.328	0.215	0.192	1
Patron	0.292	0.238	0.362	0.108	1
Cadre.sup	0.258	0.219	0.480	0.043	1
Employé	0.310	0.285	0.289	0.116	1
Ouvrier	0.270	0.314	0.207	0.210	1
Total	0.272	0.254	0.373	0.101	1

P(Sciences) = 962 / 3784 = 25.4%

P(Sciences / Cadre sup) = 400 / 1825 = 21.9%

P(Sciences / Ouvrier) = 193 / 615 = 31.4%

Question : Globalement, les enfants de « cadre sup. » et « d’ouvrier » font-ils les mêmes choix ? Et les enfants de « patron » et de « cadre sup. » ?

Distance entre profils :
$$d^2(k, k') = \sum_{l=1}^L \frac{n_{.l}}{n} \left(\frac{n_{kl}}{n_{k.}} - \frac{n_{k'l}}{n_{k' .}} \right)^2$$

Les écarts entre modalités rares sont exacerbés.

$$d^2(cadre, ouvrier) = \frac{1}{0.272} (0.258 - 0.270)^2 + \frac{1}{0.254} (0.219 - 0.314)^2 + \frac{1}{0.373} (0.480 - 0.207)^2 + \frac{1}{0.101} (0.043 - 0.210)^2 = 0.5109$$

$$d^2(cadre, patron) = \frac{1}{0.272} (0.258 - 0.292)^2 + \frac{1}{0.254} (0.219 - 0.238)^2 + \frac{1}{0.373} (0.480 - 0.362)^2 + \frac{1}{0.101} (0.043 - 0.108)^2 = 0.084611$$

Les choix de filières des enfants de patrons sont plus proches (similaires) de ceux des cadres sup. qu’ils ne le sont de ceux des enfants d’ouvriers.

Distance à l'origine (distance au profil « moyen »)

CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	80	99	65	58	302
Patron	168	137	208	62	575
Cadre.sup	470	400	876	79	1825
Employé	145	133	135	54	467
Ouvrier	166	193	127	129	615
Total	1029	962	1411	382	3784

CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	0.265	0.328	0.215	0.192	1
Patron	0.292	0.238	0.362	0.108	1
Cadre.sup	0.258	0.219	0.480	0.043	1
Employé	0.310	0.285	0.289	0.116	1
Ouvrier	0.270	0.314	0.207	0.210	1
Total	0.272	0.254	0.373	0.101	1

Profil marginal = structure de choix sans distinction de CSP = Profil moyen c.à-d. moyenne pondérée des profils lignes

Ex.
$$0.272 = \frac{1}{3784} \times \left(302 \times 0.265 + 575 \times 0.292 + 1825 \times 0.258 + 467 \times 0.310 + 615 \times 0.270 \right)$$

Distance à l'origine =
Distance au profil moyen

$$\begin{aligned}
 d^2(\text{cadre}) &= \frac{1}{0.272} (0.258 - 0.272)^2 + \frac{1}{0.254} (0.219 - 0.254)^2 \\
 &+ \frac{1}{0.373} (0.480 - 0.373)^2 + \frac{1}{0.101} (0.043 - 0.101)^2 \\
 &= 0.0693
 \end{aligned}$$

CSP\Filière	DISTO ²
Exp.agri	0.1703
Patron	0.0033
Cadre.sup	0.0693
Employé	0.0301
Ouvrier	0.2055

Les enfants d'ouvriers et d'exploitants agricoles présentent les structures de choix (les profils) les plus « différents » de l'ensemble des étudiants.

Distance à l'origine et distance entre profils

CSP\Filière	DISTO ²
Exp.agri	0.1703
Patron	0.0033
Cadre.sup	0.0693
Employé	0.0301
Ouvrier	0.2055

Analyse globale : positionnement par rapport à la moyenne.

Les enfants d'ouvriers et d'exploitants agricoles présentent les structures de choix (les profils) les plus « différents » de l'ensemble des étudiants.

CSP	Exp.agri	Patron	Cadre.sup	Employé	Ouvrier
Exp.agri	0	0.1621	0.4538	0.0874	0.0042
Patron		0	0.0846	0.0245	0.1918
Cadre.sup			0	0.1768	0.5109
Employé				0	0.1154
Ouvrier					0

Analyse approfondie : positionnement (distance) deux à deux.

Les enfants d'ouvriers et d'exploitants agricoles présentent des structures de choix très proches.

En revanche, cadre sup se démarque fortement de ces deux CSP.

L' AFC fournit une représentation graphique du positionnement relatif des modalités !

C'est très pratique quand leur nombre est élevé.



INERTIE = DISTO x Poids de la modalité

CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	80	99	65	58	302
Patron	168	137	208	62	575
Cadre.sup	470	400	876	79	1825
Employé	145	133	135	54	467
Ouvrier	166	193	127	129	615
Total	1029	962	1411	382	3784

CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	0.265	0.328	0.215	0.192	1
Patron	0.292	0.238	0.362	0.108	1
Cadre.sup	0.258	0.219	0.480	0.043	1
Employé	0.310	0.285	0.289	0.116	1
Ouvrier	0.270	0.314	0.207	0.210	1
Total	0.272	0.254	0.373	0.101	1

Quantité d'information portée par la modalité « Cadre » = son poids x son écartement par rapport au profil moyen

$$\begin{aligned}
 \text{Inertie(cadre)} &= \text{poids(cadre)} \times d^2(\text{cadre}) \\
 &= \frac{1825}{3784} \times 0.0693 \\
 &= 0.4832 \times 0.0693 \\
 &= 0.0334
 \end{aligned}$$

- Ex. « Exploitant Agricole » est différent, mais concerne peu d'individus → son inertie ne se démarque pas.
- Ex. A contrario, cf. « Cadre.sup ».

CSP\Filière	DISTO ²	Poids	INERTIE
Exp.agri	0.1703	0.0798	0.0136
Patron	0.0033	0.1520	0.0005
Cadre.sup	0.0693	0.4823	0.0334
Employé	0.0301	0.1234	0.0037
Ouvrier	0.2055	0.1625	0.0334
		Total	0.0846

Somme(INERTIE) = Quantité d'information disponible dans les données ! C'est un indicateur fondamental.

Profils colonnes

Définition - Distance entre profils - Distance à l'origine



$$P(Y = k / X = l) = \frac{n_{kl}}{n_{.l}}$$

P(Cadre sup) = 48.2%

P(Cadre sup / Droit) = 45.7%

P(Cadre sup / IUT) = 20.7%

Question : Dans les filières, a-t-on les mêmes structures de CSP ?

CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	0.078	0.103	0.046	0.152	0.080
Patron	0.163	0.142	0.147	0.162	0.152
Cadre.sup	0.457	0.416	0.621	0.207	0.482
Employé	0.141	0.138	0.096	0.141	0.123
Ouvrier	0.161	0.201	0.090	0.338	0.163
Total	1	1	1	1	1

Distance entre profils :

$$d^2(l, l') = \sum_{k=1}^K \frac{n}{n_{k.}} \left(\frac{n_{kl}}{n_{.l}} - \frac{n_{kl'}}{n_{.l'}} \right)^2$$

Les écarts entre modalités rares sont exacerbés.

$$d^2(\text{droit}, \text{sciences}) = \frac{1}{0.080} (0.078 - 0.103)^2 + \frac{1}{0.152} (0.163 - 0.142)^2 + \dots = 0.024$$

$$d^2(\text{droit}, \text{médecine}) = \frac{1}{0.080} (0.078 - 0.046)^2 + \frac{1}{0.152} (0.163 - 0.147)^2 + \dots = 0.118$$

Le profil sociologique des étudiants en « droit » est plus proche de ceux en « sciences » qu'en « médecine ».

Distance à l'origine et inertie

CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	0.078	0.103	0.046	0.152	0.080
Patron	0.163	0.142	0.147	0.162	0.152
Cadre.sup	0.457	0.416	0.621	0.207	0.482
Employé	0.141	0.138	0.096	0.141	0.123
Ouvrier	0.161	0.201	0.090	0.338	0.163
Total	1	1	1	1	1

CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	80	99	65	58	302
Patron	168	137	208	62	575
Cadre.sup	470	400	876	79	1825
Employé	145	133	135	54	467
Ouvrier	166	193	127	129	615
Total	1029	962	1411	382	3784

Le profil sociologique des étudiants en « IUT » est le plus différent de la globalité.

Filière	Droit	Sciences	Médecine	IUT	
DISTO ²	0.0047	0.0272	0.0928	0.4145	
Poids	0.2719	0.2542	0.3729	0.1010	Total
Inertie	0.0013	0.0069	0.0346	0.0418	0.0846

Somme(INERTIE) = Identique à celle issue de l'étude des profils lignes. Normal, on analyse le même tableau, mais avec un point de vue différent !

« IUT » et « Médecine » sont les modalités les porteuses d'information.

Association lignes - colonnes

Test d'indépendance du KHI-2

Décomposition du KHI-2



H0 : X et Y sont indépendants → $P(Y=k \ \& \ X=l) = P(Y=k) \times P(X=l)$

↑ Probabilité conjointe

Effectifs observés					
CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	80	99	65	58	302
Patron	168	137	208	62	575
Cadre.sup	470	400	876	79	1825
Employé	145	133	135	54	467
Ouvrier	166	193	127	129	615
Total	1029	962	1411	382	3784

Les effectifs sont calculés à partir des marges.

Effectifs sous H0					
CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	82.1	76.8	112.6	30.5	302
Patron	156.4	146.2	214.4	58.0	575
Cadre.sup	496.3	464.0	680.5	184.2	1825
Employé	127.0	118.7	174.1	47.1	467
Ouvrier	167.2	156.4	229.3	62.1	615
Total	1029	962	1411	382	3784

$$e_{kl} = \frac{n_{k.} \times n_{.l}}{n}$$

Ex. $82.1 = \frac{302 \times 1029}{3784}$

Statistique du KHI-2 de Pearson

$$\chi^2 = \sum_{k=1}^K \sum_{l=1}^L \frac{(n_{kl} - e_{kl})^2}{e_{kl}}$$

Sous H0, la statistique suit une loi du KHI-2 à (K-1)x(L-1) degrés de liberté.

KHI-2 d'écart à l'indépendance - Exemple

Effectifs observés					
CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	80	99	65	58	302
Patron	168	137	208	62	575
Cadre.sup	470	400	876	79	1825
Employé	145	133	135	54	467
Ouvrier	166	193	127	129	615
Total	1029	962	1411	382	3784

Effectifs sous H0					
CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	82.1	76.8	112.6	30.5	302
Patron	156.4	146.2	214.4	58.0	575
Cadre.sup	496.3	464.0	680.5	184.2	1825
Employé	127.0	118.7	174.1	47.1	467
Ouvrier	167.2	156.4	229.3	62.1	615
Total	1029	962	1411	382	3784

KHI-2	320.3
ddl	12
p-value	< 0.0001

Au risque 5%, on rejette l'hypothèse nulle d'indépendance entre Y et X.

CSP\Filière	Droit	Sciences	Médecine	IUT
Exp.agri	0.1	6.4	20.1	24.8
Patron	0.9	0.6	0.2	0.3
Cadre.sup	1.4	8.8	56.2	60.1
Employé	2.6	1.7	8.8	1.0
Ouvrier	0.0	8.6	45.7	72.1

Remarque : Restrictions sur la validité du test.

1. Il faut que 80% des cellules présentent un $e_{kl} \geq 5$ (il faut surtout noter qu'une faible valeur de e_{kl} « gonfle » exagérément la valeur du KHI-2)
2. Quand n est élevé, il est toujours significatif (car les ddl ne tiennent pas compte de n).

KHI-2 n'est pas normalisé, $\chi^2 = 20 \rightarrow$ on ne sait pas vraiment ce que ça veut dire

$$0 \leq \chi^2 \leq n \times \min(K - 1, L - 1)$$

Mesures dérivées du KHI-2

$$\text{PHI-2} \quad \phi^2 = \frac{\chi^2}{n} = \frac{320.3}{3784} = 0.0846$$

0.0846 = Inertie totale. L'information disponible peut être perçue sous l'angle de l'association entre Y et X.

T de Tschuprow : normalisation par les degrés de liberté. Varie entre 0 et 1 sous certaines conditions.

$$t = \sqrt{\frac{\chi^2}{n \times \sqrt{(K-1)(L-1)}}} = \sqrt{\frac{320.3}{3784 \times \sqrt{(5-1)(4-1)}}} = 0.1563$$

V de Cramer : autre normalisation. Varie entre 0 et 1.

$$V = \sqrt{\frac{\chi^2}{n \times \min(K-1, L-1)}} = \sqrt{\frac{320.3}{3784 \times \min(5-1, 4-1)}} = 0.1680$$



Analyse approfondie des associations

Contributions au KHI-2 (à l'inertie totale)

Effectifs observés					
CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	80	99	65	58	302
Patron	168	137	208	62	575
Cadre.sup	470	400	876	79	1825
Employé	145	133	135	54	467
Ouvrier	166	193	127	129	615
Total	1029	962	1411	382	3784

Résidus standardisés				
CSP\Filière	Droit	Sciences	Médecine	IUT
Exp.agri	-0.23	2.54	-4.49	4.98
Patron	0.93	-0.76	-0.44	0.52
Cadre.sup	-1.18	-2.97	7.49	-7.75
Employé	1.60	1.31	-2.97	1.00
Ouvrier	-0.10	2.93	-6.76	8.49

(+) attraction ; (-) répulsion

$$r_{kl} = \frac{n_{kl} - e_{kl}}{\sqrt{e_{kl}}}$$

Suit très approximativement une loi normale (0, 1) c.-à-d. au risque 5%,
 $|r_{kl}| > 2 \rightarrow$ significatif.

Question : quelles sont les associations (de modalités) qui contribuent le plus à l'information ?

Résidus standardisés et contributions au KHI-2.

Effectifs sous H0					
CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	82.1	76.8	112.6	30.5	302
Patron	156.4	146.2	214.4	58.0	575
Cadre.sup	496.3	464.0	680.5	184.2	1825
Employé	127.0	118.7	174.1	47.1	467
Ouvrier	167.2	156.4	229.3	62.1	615
Total	1029	962	1411	382	3784

Contributions au KHI-2 (%)					
CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	0.02%	2.01%	6.29%	7.75%	16.06%
Patron	0.27%	0.18%	0.06%	0.08%	0.59%
Cadre.sup	0.43%	2.75%	17.53%	18.77%	39.49%
Employé	0.80%	0.54%	2.75%	0.31%	4.39%
Ouvrier	0.00%	2.68%	14.26%	22.52%	39.46%
Total	1.52%	8.16%	40.88%	49.44%	

$$c_{kl} = \frac{r_{kl}^2}{\chi^2}$$

Fraction d'information
 (contribution à l'information)
 portée par chaque case.

L'AFC résulte d'une décomposition en valeurs singulières de cette matrice !!!

(<http://www.mathematica-journal.com/2010/09/an-introduction-to-correspondence-analysis/>)

Autre prisme : Test du rapport de vraisemblance

Utiliser l'Indice d'attraction et de répulsion pour une analyse approfondie des associations

Indice d'attraction et répulsion

($i > 1$: attraction ; $i < 1$: répulsion)

$$i_{kl} = \frac{P(Y = k \& X = l)}{P(Y = k) \times P(X = l)} \# \frac{o_{kl}}{e_{kl}}$$

Statistique du rapport de vraisemblance. Sous H_0 : suit une loi du KHI-2 à $(K-1)(L-1)$ ddl.

$$G = 2 \sum_{k=1}^K \sum_{l=1}^L n_{kl} \times \ln(i_{kl})$$

	Indice d'attraction - répulsion			
CSP\Filière	Droit	Sciences	Médecine	IUT
Exp.agri	0.974	1.289	0.577	1.902
Patron	1.074	0.937	0.970	1.068
Cadre.sup	0.947	0.862	1.287	0.429
Employé	1.142	1.120	0.775	1.145
Ouvrier	0.993	1.234	0.554	2.078

G	321.964
ddl	12
p-value	< 0.0001

La statistique de Pearson est une approximation du rapport de vraisemblance... mais elle est plus connue (et utilisée).



Analyse factorielle des correspondances

Positionner les profils entre eux

Identifier les attractions / répulsions

Indispensable pour les grands tableaux !!!



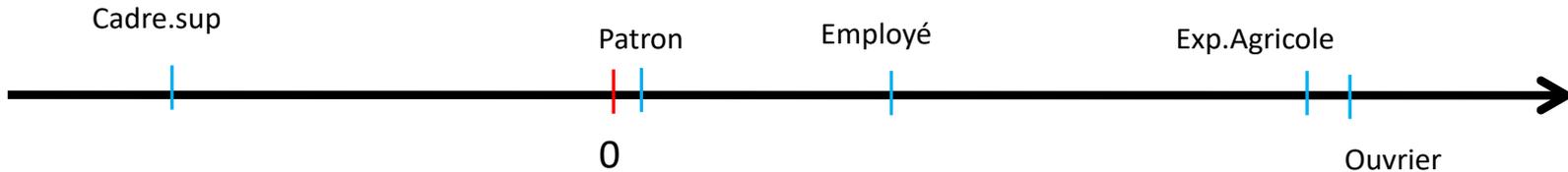
		1	2	3	4	
CSP\Filière		Droit	Sciences	Médecine	IUT	Total
1	Exp.agri	80	99	65	58	302
2	Patron	168	137	208	62	575
3	Cadre.sup	470	400	876	79	1825
4	Employé	145	133	135	54	467
5	Ouvrier	166	193	127	129	615
Total		1029	962	1411	382	3784

1. Quelles sont les CSP qui font les mêmes choix de filières ?
2. Quelles sont les filières qui ont des structures de CSP proches ?
3. Quelle CSP est attirée par quelle filière ? Quelle filière est attractive pour quelle CSP ?

On souhaite obtenir une vision à la fois globale et synthétique.
Idéalement une (des) représentation(s) graphique(s).



Trouver un premier facteur qui permet de positionner les modalités de la manière la plus dispersée possible (pour qu'on les distingue le mieux possible les uns des autres)

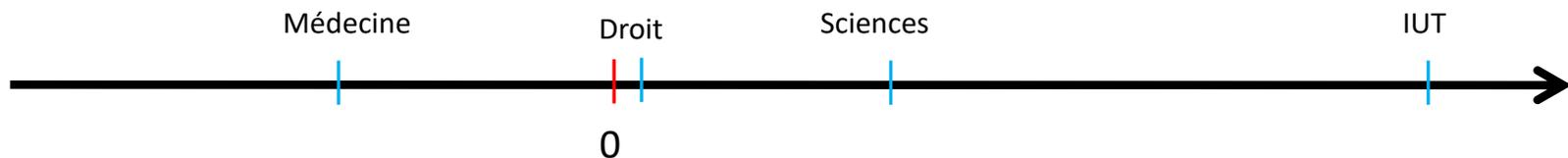


$$\lambda_1 = \sum_{k=1}^K \frac{n_{k.}}{n} \times F_{k1}^2$$

- F_{k1} sont les « coordonnées factorielles » sur le 1^{er} facteur
- La moyenne des points modalités est égale à 0
- λ représente la variance des points modalités
- On cherche à maximiser λ

$$\frac{\lambda_1}{\phi^2}$$

Indique la crédibilité (la fidélité à la réalité) du facteur



$$\lambda_1 = \sum_{l=1}^L \frac{n_{.l}}{n} \times G_{l1}^2$$

- G_{l1} sont les coordonnées factorielles sur le 1^{er} facteur
- Le λ calculé ici coïncide avec celui calculé sur le profil ligne



$$\phi^2 = \frac{\chi^2}{n} = 0.0846 \quad \text{Inertie totale}$$

Values	Poids	Coord.1
Cadre.sup	0.48229	-0.26272
Patron	0.15196	0.02015
Employé	0.12341	0.14209
Exp.agri	0.07981	0.41012
Ouvrier	0.16253	0.45148

$(0.0824/0.0846) = 97.4\%$ de l'information disponible est restituée par ce facteur (axe factoriel).

Lambda 1	0.0824
----------	---------------

$$d^2_1(\text{patron}) = 0.1521 = (0.41012 - 0.02015)^2$$

Remarque : 100% de l'information = tableau initial des distances entres profils.
Effectivement, le positionnement factoriel propose une image assez fidèle ici.

CSP	Exp.agri	Patron	Cadre.sup	Employé	Ouvrier
Exp.agri	0	0.1621	0.4538	0.0874	0.0042
Patron		0	0.0846	0.0245	0.1918
Cadre.sup			0	0.1768	0.5109
Employé				0	0.1154
Ouvrier					0

Distance entre modalités (tableau initial)

CSP	Exp.agri	Patron	Cadre.sup	Employé	Ouvrier
Exp.agri	0	0.1521	0.4527	0.0718	0.0017
Patron		0	0.0800	0.0149	0.1860
Cadre.sup			0	0.1639	0.5101
Employé				0	0.0957
Ouvrier					0

Distance entre modalités (1^{er} facteur)

L'AFC devient décisive lorsque le nombre de modalité est élevé !!!

Values	Poids	Coord.1
Cadre.sup	0.48229	-0.26272
Patron	0.15196	0.02015
Employé	0.12341	0.14209
Exp.agri	0.07981	0.41012
Ouvrier	0.16253	0.45148

Lambda 1	0.0824
----------	--------

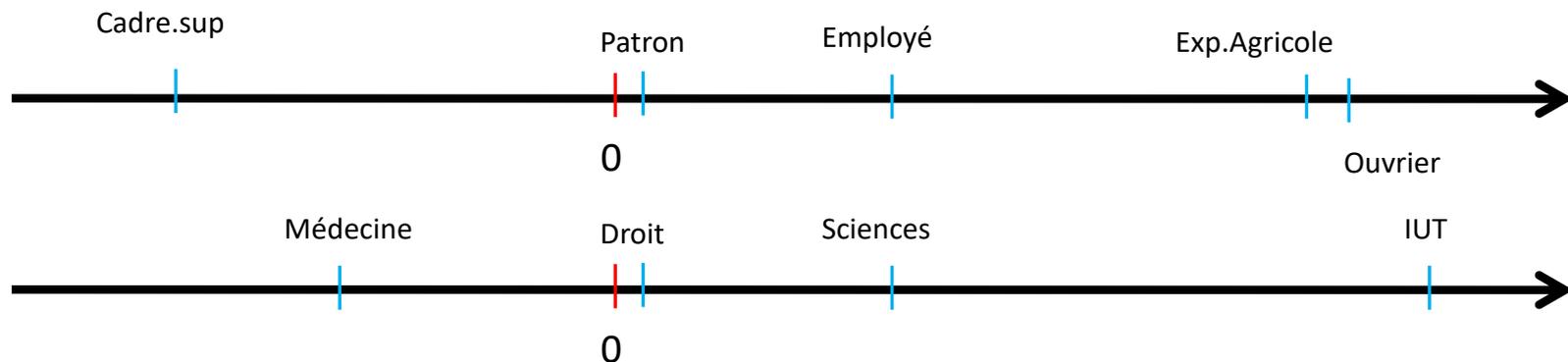
Values	Poids	Coord.1
Médecine	0.37289	-0.30313
Droit	0.27193	0.02799
Sciences	0.25423	0.16046
IUT	0.10095	0.64017

Lambda 1	0.0824
----------	--------

1. λ_1 est bien le même (calculs sur les modalités lignes ou colonnes)
2. OK pour le positionnement relatif des modalités des lignes (resp. des colonnes).
3. Mais que peut-on dire en termes d'associations lignes/colonnes ? Est-ce que leurs proximités sur l'axe est une indication ?

Ex.A – Les enfants d'ouvriers sont attirés par les IUT ?

Ex.B – Les enfants de patrons sont attirés par le droit ?



On peut obtenir les coordonnées des modalités lignes à partir de l'ensemble des points colonnes (et inversement) : moyenne pondérée déflatée par la racine de l'inertie.

$$F_{k1} = \frac{1}{\sqrt{\lambda_1}} \sum_{l=1}^L \frac{n_{kl}}{n_k} \times G_{l1}$$

Coordonnées des modalités colonnes sur le 1^{er} axe

Values	Droit	Sciences	Médecine	IUT
coord 1	0.02799	0.16046	-0.30313	0.64017

Profil ligne des « ouvriers »

	Droit	Sciences	Médecine	IUT
Ouvrier	0.2699	0.3138	0.2065	0.2098

Coordonnée de « ouvriers » à partir de la relation

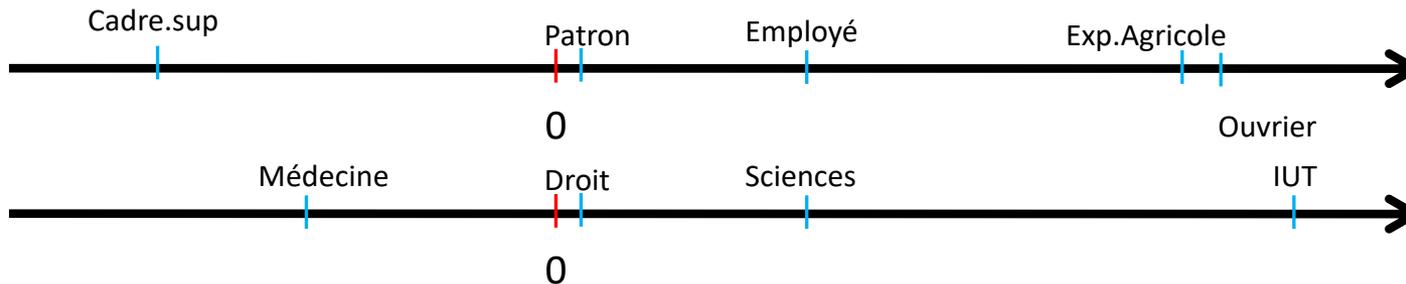
$$F_{51} = \frac{1}{\sqrt{0.0824}} (0.2699 \times 0.02799 + \dots + 0.2098 \times 0.64017) = 0.45148$$

Conclusion :

1. Le rapprochement des coordonnées des modalités lignes / colonnes est licite grâce à cette relation
2. Mais il ne peut se faire que globalement c.-à-d. une modalité ligne doit être située par rapport à l'ensemble des modalités colonnes (et inversement)



AFC, analyse des associations – Relations quasi-barycentriques (suite)



Ex.A – Les enfants d’ouvriers sont attirés par les IUT ?

Réponse : OUI. Les enfants d’ouvriers sont plus attirés par les IUT que l’ensemble des étudiants.

Ex.B - Les enfants de patrons sont attirés par le droit ?

Réponse : NON. Par rapport à l’ensemble des étudiants, les enfants de patrons ne sont attirés par aucune filière en particulier (idem pour Employé d’ailleurs).

C'est bien l'information fournie par les profils lignes

	Droit	Sciences	Médecine	IUT
Ouvrier	0.270	0.314	0.207	0.210
Patron	0.292	0.238	0.362	0.108
Marge (Total)	0.272	0.254	0.373	0.101

C'est aussi l'information fournie par le tableau des résidus standardisés

CSP\Filière	Résidus standardisés			
	Droit	Sciences	Médecine	IUT
Exp.agri	-0.23	2.54	-4.49	4.98
Patron	0.93	-0.76	-0.44	0.52
Cadre.sup	-1.18	-2.97	7.49	-7.75
Employé	1.60	1.31	-2.97	1.00
Ouvrier	-0.10	2.93	-6.76	8.49

Indices reconstitués à partir du 1er axe

CSP\Filière	Droit	Sciences	Médecine	IUT	Coord.1
Exp.agri	1.040	1.229	0.567	1.915	0.41012
Patron	1.002	1.011	0.979	1.045	0.02015
Cadre.sup	0.974	0.853	1.277	0.414	-0.26272
Employé	1.014	1.079	0.850	1.317	0.14209
Ouvrier	1.044	1.252	0.523	2.007	0.45148
Coord.1	0.02799	0.16046	-0.30313	0.64017	

Lambda.1 0.082394

Indices reconstitués à partir des H premiers axes.

$$\hat{i}_{kl} = 1 + \sum_{h=1}^H \frac{F_{kh} \times G_{lh}}{\sqrt{\lambda_h}}$$

Estimation des indices à partir du 1^{er} axe

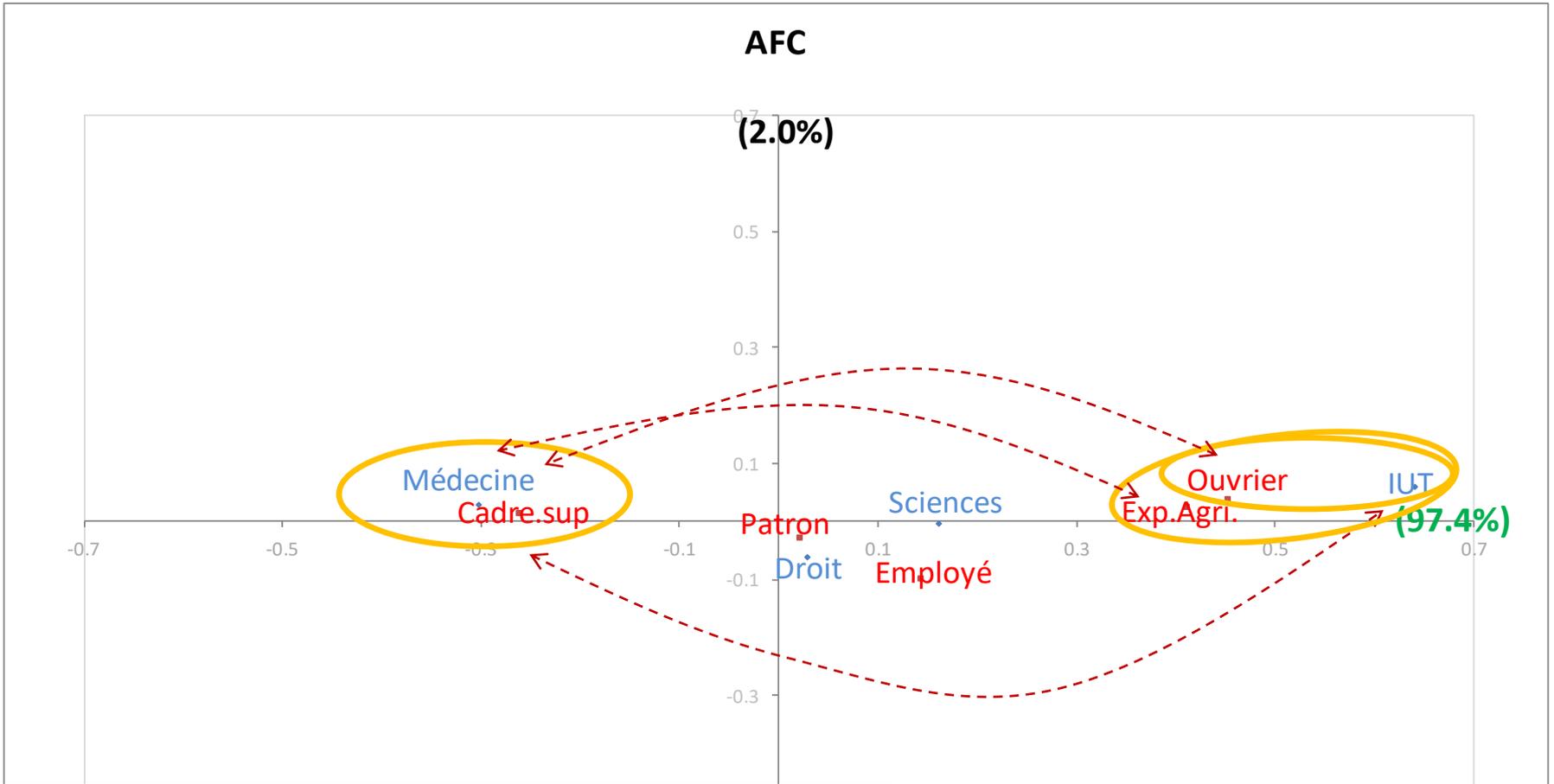
- A. 2 modalités s'attirent (se repoussent) si leurs coordonnées sont de même signe (de signe contraire) sur les axes.
- B. Le trait est d'autant plus marqué que les valeurs sont élevées (en valeur absolue).

Indices calculés sur le tableau initial des données

	Indice d'attraction - répulsion			
CSP\Filière	Droit	Sciences	Médecine	IUT
Exp.agri	0.974	1.289	0.577	1.902
Patron	1.074	0.937	0.970	1.068
Cadre.sup	0.947	0.862	1.287	0.429
Employé	1.142	1.120	0.775	1.145
Ouvrier	0.993	1.234	0.554	2.078



AFC, analyse des associations – Représentation simultanée dans le plan



Row	Column	Value	Expected	Std.Resid.	Contrib.	Cumul (%)
Ouvrier	IUT	129	62.1	8.49	22.52	22.52
Cadre.sup	IUT	79	184.2	-7.75	18.77	41.29
Cadre.sup	Médecine	876	680.5	7.49	17.53	58.82
Ouvrier	Médecine	127	229.3	-6.76	14.26	73.08
Exp.agri	IUT	58	30.5	4.98	7.75	80.83
Exp.agri	Médecine	65	112.6	-4.49	6.29	87.12

AFC, analyse des associations – Décomposition du χ^2

Effectifs observés

CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	80	99	65	58	302
Patron	168	137	208	62	575
Cadre.sup	470	400	876	79	1825
Employé	145	133	135	54	467
Ouvrier	166	193	127	129	615
Total	1029	962	1411	382	3784

Effectifs sous H0 (Indépendance)

CSP\Filière	Droit	Sciences	Médecine	IUT
Exp.agri	82.1	76.8	112.6	30.5
Patron	156.4	146.2	214.4	58.0
Cadre.sup	496.3	464.0	680.5	184.2
Employé	127.0	118.7	174.1	47.1
Ouvrier	167.2	156.4	229.3	62.1

$$\chi_{total}^2 = 320.27$$

Vérifier la qualité de restitution de l'information sur le 1^{er} axe

Effectifs reconstitués à partir du 1er axe

CSP\Filière	Droit	Sciences	Médecine	IUT
Exp.agri	85.4	94.4	63.8	58.4
Patron	156.7	147.8	209.8	60.7
Cadre.sup	483.6	395.8	869.3	76.3
Employé	128.8	128.2	148.0	62.1
Ouvrier	174.6	195.8	120.0	124.6

$$\chi_1^2 = \sum_k \sum_l \frac{(\hat{n}_{kl}^{(1)} - e_{kl})^2}{e_{kl}} = 311.7810$$

$$\phi_1^2 = \frac{\chi_1^2}{n} = \lambda_1 = 0.0824$$

Formule de reconstitution
(h premiers axes)

$$\hat{n}_{kl}^{(h)} = \frac{n_{k.} \cdot n_{.l}}{n} \left(1 + \sum_h \frac{F_{hk} \times G_{hl}}{\sqrt{\lambda_h}} \right)$$

$$(\hat{n}_{kl}^1 - n_{kl})$$

Quantifier l'information non restituée sur le 1^{er} axe

Reste - Ecart au tableau observé

CSP\Filière	Droit	Sciences	Médecine	IUT
Exp.agri	-5.4	4.6	1.2	-0.4
Patron	11.3	-10.8	-1.8	1.3
Cadre.sup	-13.6	4.2	6.7	2.7
Employé	16.2	4.8	-13.0	-8.1
Ouvrier	-8.6	-2.8	7.0	4.4

$$\chi_{résiduel}^2 = \sum_k \sum_l \frac{(\hat{n}_{kl}^{(1)} - n_{kl})^2}{e_{kl}} = 8.4885$$



$$\chi_{total}^2 = \chi_1^2 + \chi_{résiduel}^2$$



Analyse factorielle des correspondances

Quelques éléments sur les calculs



Soit
$$M = \frac{1}{\sqrt{n}} S$$

où S est la matrice des résidus standardisés

L'AFC consiste à calculer la décomposition en valeurs singulières de M

$$M = U \Delta V^T$$

U ($K \times K$) contient les K vecteurs singuliers à gauche (modalités lignes). U est orthonormée.

Δ ($K \times L$) est une matrice dont les éléments situés sur la diagonale correspondent aux valeurs singulières.

V ($L \times L$) contient les L vecteurs singuliers à droite (modalités colonnes). V est orthonormée.

A quoi correspond cette opération ?

(On perçoit bien l'analyse croisée ligne/colonne ici)

Une valeur singulière δ_h est telle que

$$M \vec{v}_h = \delta_h \vec{u}_h$$
$$M^T \vec{u}_h = \delta_h \vec{v}_h$$

Concrètement ?

1. On cherche à produire des vecteurs de projections de manière à ce que la dispersion des modalités lignes (colonnes) soit la plus grande possible sur l'axe.
2. La dispersion doit être la même pour les modalités lignes et les modalités colonnes.
3. Les facteurs sont orthogonaux deux à deux.



$$\lambda_h = \delta_h^2$$

L'inertie expliquée par l'axe (la valeur propre associée à l'axe) = carré de la valeur singulière

$\lambda_h \leq 1$ (cf. Lebart et al., pages 85 et 86)

$$F_{kh} = \frac{u_{kh} \times \delta_h}{\sqrt{\frac{n_{k.}}{n}}}$$

Obtention des coordonnées des modalités lignes à partir des vecteurs singuliers à gauche.

$$G_{lh} = \frac{v_{lh} \times \delta_h}{\sqrt{\frac{n_{.l}}{n}}}$$

Obtention des coordonnées des modalités colonnes à partir des vecteurs singuliers à droite.



```

#charger le fichier de données
etudiants <- read.table(file="afc_etudiants.txt",row.names=1,header=T,sep="\t")
print(etudiants)

#transformer le data.frame en tableau croisé
etudiants <- as.table(etudiants)

#effectif total
n <- sum(etudiants)
print(n)

#marge ligne
sum.row <- apply(etudiants,1,sum)
sum.row <- matrix(sum.row,nrow=length(sum.row),ncol=1)

#marge colonne
sum.col <- apply(etudiants,2,sum)
sum.col <- matrix(sum.col,nrow=1,ncol=length(sum.col))

#tableau sous indépendance
e <- (sum.row %*% sum.col)/n
print(e)

#résidus standardisés (au facteur 1/racine(n) près)
res.std <- (1.0/sqrt(n))*(etudiants-e)/sqrt(e)
print(res.std)

#nombre d'axes max.
nb.axes <- min(nrow(res.std)-1,ncol(res.std)-1)

#décomposition en valeurs singulières
etu.svd <- svd(res.std,nu=nb.axes,nv=nb.axes)
print(etu.svd)

```

	Droit	Sciences	Medecine	IUT
Exp.agri	-0.003810533	0.04122993	-0.072936528	0.081002388
Patron	0.015129514	-0.01234475	-0.007115688	0.008434451
Cadre.sup	-0.019177531	-0.04827641	0.121818945	-0.126038226
Employe	0.025975610	0.02129813	-0.048214003	0.016231633
Ouvrier	-0.001558365	0.04764791	-0.109844873	0.138055472

```

> print(etu.svd)
$d
[1] 2.870429e-01 4.127286e-02 2.323375e-02 1.121893e-17

$u
      [,1]      [,2]      [,3]
[1,] 0.40363357 -0.1796988 0.4655040
[2,] 0.02736552 0.2510939 -0.7865605
[3,] -0.63561972 -0.2624212 0.1852881
[4,] 0.17390042 0.8284107 0.3211899
[5,] 0.63409564 -0.3866923 -0.1647239

$v
      [,1]      [,2]      [,3]
[1,] 0.05084468 0.76654132 -0.3713418
[2,] 0.28186205 0.03340234 0.8156039
[3,] -0.64485656 -0.43885508 -0.1366763
[4,] 0.70861035 -0.46765896 -0.4221553

```



```
#vérification orthonormalité des vecteurs de projection
```

```
U <- matrix(etu.svd$u,nrow=5,ncol=3)
```

```
print(t(U)%*%U)
```

```
V <- matrix(etu.svd$v,nrow=4,ncol=3)
```

```
print(t(V)%*%V)
```

```
#les valeurs propres (inerties des axes)
```

```
print(round(etu.svd$d^2,4))
```

```
#profils marginaux
```

```
profil.row <- sum.row/n
```

```
profil.col <- sum.col/n
```

```
#coordonnées des modalités lignes - 1er facteur
```

```
print(round(etu.svd$d[1]*U[,1]/sqrt(profil.row),5))
```

```
#coordonnées des modalités colonnes - 1er facteur
```

```
print(round(etu.svd$d[1]*V[,1]/sqrt(profil.col),5))
```

```
> print(t(U)%*%U)
      [,1]      [,2]      [,3]
[1,] 1.000000e+00 -2.498002e-16 -3.608225e-16
[2,] -2.498002e-16 1.000000e+00 -2.359224e-16
[3,] -3.608225e-16 -2.359224e-16 1.000000e+00
> V <- matrix(etu.svd$v,nrow=4,ncol=3)
> print(t(V)%*%V)
      [,1]      [,2]      [,3]
[1,] 1.000000e+00 5.551115e-17 -1.665335e-16
[2,] 5.551115e-17 1.000000e+00 -1.665335e-16
[3,] -1.665335e-16 -1.665335e-16 1.000000e+00
```

```
> print(round(etu.svd$d^2,4))
[1] 0.0824 0.0017 0.0005 0.0000
```

```
> print(round(etu.svd$d[1]*U[,1]/sqrt(profil.row),5))
      [,1]
[1,] 0.41012
[2,] 0.02015
[3,] -0.26272
[4,] 0.14209
[5,] 0.45148
```

```
> print(round(etu.svd$d[1]*V[,1]/sqrt(profil.col),5))
      [,1]      [,2]      [,3]      [,4]
[1,] 0.02799 0.16046 -0.30313 0.64017
```



Analyse factorielle des correspondances

Pratique de l'AFC



AFC – Choisir le nombre d’axes

1. Nombre d’axes maximum que l’on peut produire

$$H_{max} = \min(K-1 ; L-1)$$

CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	80	99	65	58	302
Patron	168	137	208	62	575
Cadre.sup	470	400	876	79	1825
Employé	145	133	135	54	467
Ouvrier	166	193	127	129	615
Total	1029	962	1411	382	3784

Peut être déduite des (L-1) colonnes et la marge

Peut être déduite des (K-1) lignes et la marge

2. ~ Règle de Kaiser :

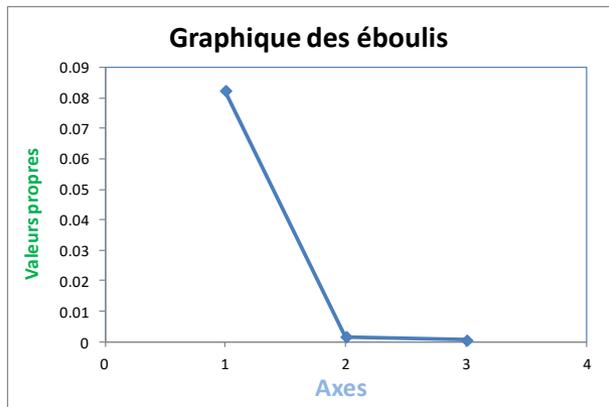
(peu utilisée)

Retenir un axe si le pourcentage d’inertie est supérieur à $(1/H_{max})$

$(1/3) = 33\%$ pour notre exemple

Axis	Eigen value	% explained
1	0.0824	97.35%
2	0.0017	2.01%
3	0.0005	0.64%
Tot.	0.0846	-

3. Scree plot : Graphique des éboulis et règle du coude



4. Pragmatique : au moins 2 axes pour obtenir une représentation graphique

$$CTR_h(k) = \frac{\frac{n_{k.} d_h^2(k)}{n}}{\sum_{k=1}^K \frac{n_{k.} d_h^2(k)}{n}} = \frac{n_{k.} d_h^2(k)}{n \lambda_h}$$

Rows analysis				Inertie %		97.35%	2.01%
Characterization				Coord.		Contributions (%)	
Values	Weight	Sq. Dist.	Inertia	coord 1	coord 2	ctr 1	ctr 2
Exp.agri	0.0798	0.1704	0.0136	0.4101	0.0263	16.29	3.23
Patron	0.1520	0.0033	0.0005	0.0202	-0.0266	0.07	6.30
Cadre.sup	0.4823	0.0693	0.0334	-0.2627	0.0156	40.40	6.89
Employé	0.1234	0.0301	0.0037	0.1421	-0.0973	3.02	68.63
Ouvrier	0.1625	0.2055	0.0334	0.4515	0.0396	40.21	14.95

Contribution : influence relative de la modalité dans la définition de l'axe factoriel.
 Part d'inertie de la modalité dans l'inertie totale de l'axe.
 Somme_k[CTR_h(k)] = 100%.

Informations sur les modalités : importants pour relativiser les résultats de l'AFC

Positionnement dans le plan factoriel

Columns analysis				Inertie %		97.35%	2.01%
Characterization				Coord.		Contributions (%)	
Values	Weight	Sq. Dist.	Inertia	coord 1	coord 2	ctr 1	ctr 2
Droit	0.2719	0.0047	0.0013	0.0280	-0.0607	0.26	58.76
Sciences	0.2542	0.0272	0.0069	0.1605	-0.0027	7.94	0.11
Médecine	0.3729	0.0928	0.0346	-0.3031	0.0297	41.58	19.26
IUT	0.1010	0.4145	0.0418	0.6402	0.0608	50.21	21.87

Modalités lignes

Characterization				Coord.		COS ²	
Values	Weight	Sq. Dist.	Inertia	coord 1	coord 2	cos 1	cos 2
Exp.agri	0.0798	0.1704	0.0136	0.4101	0.0263	0.99 (0.99)	0.00 (0.99)
Patron	0.1520	0.0033	0.0005	0.0202	-0.0266	0.12 (0.12)	0.21 (0.34)
Cadre.sup	0.4823	0.0693	0.0334	-0.2627	0.0156	1.00 (1.00)	0.00 (1.00)
Employé	0.1234	0.0301	0.0037	0.1421	-0.0973	0.67 (0.67)	0.31 (0.99)
Ouvrier	0.1625	0.2055	0.0334	0.4515	0.0396	0.99 (0.99)	0.01 (1.00)

$$COS_h^2(k) = \frac{\frac{n_k}{n} d_h^2(k)}{\frac{n_k}{n} d^2(k)} = \frac{d_h^2(k)}{d^2(k)}$$

Qualité : information de la modalité reproduite par l'axe.
 Inertie de la modalité sur l'axe divisée par l'inertie totale de la modalité.
 Somme_h[COS²_h(k)] = 100%.

Modalités colonnes

Characterization				Coord.		COS ²	
Values	Weight	Sq. Dist.	Inertia	coord 1	coord 2	cos 1	cos 2
Droit	0.2719	0.0047	0.0013	0.0280	-0.0607	0.17 (0.17)	0.78 (0.94)
Sciences	0.2542	0.0272	0.0069	0.1605	-0.0027	0.95 (0.95)	0.00 (0.95)
Médecine	0.3729	0.0928	0.0346	-0.3031	0.0297	0.99 (0.99)	0.01 (1.00)
IUT	0.1010	0.4145	0.0418	0.6402	0.0608	0.99 (0.99)	0.01 (1.00)

Attention : compte tenu des CTR et COS², on serait tenté de penser qu'il y a quelque chose entre « Employé » et « Droit » sur le 2nd axe.
 → NON, car l'inertie associée à l'axe factoriel est très faible (2.01%)
 → Cela sera confirmé par le tableau des contributions au KHI-2



A lire toujours pour confirmer les informations fournies par l'AFC !!!

Contributions au KHI-2 (%)					
CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	0.02%	2.01%	6.29%	7.75%	16.06%
Patron	0.27%	0.18%	0.06%	0.08%	0.59%
Cadre.sup	0.43%	2.75%	17.53%	18.77%	39.49%
Employé	0.80%	0.54%	2.75%	0.31%	4.39%
Ouvrier	0.00%	2.68%	14.26%	22.52%	39.46%
Total	1.52%	8.16%	40.88%	49.44%	

Présentée sous une forme linéaire (et triée) dans Tanagra pour une meilleure lecture

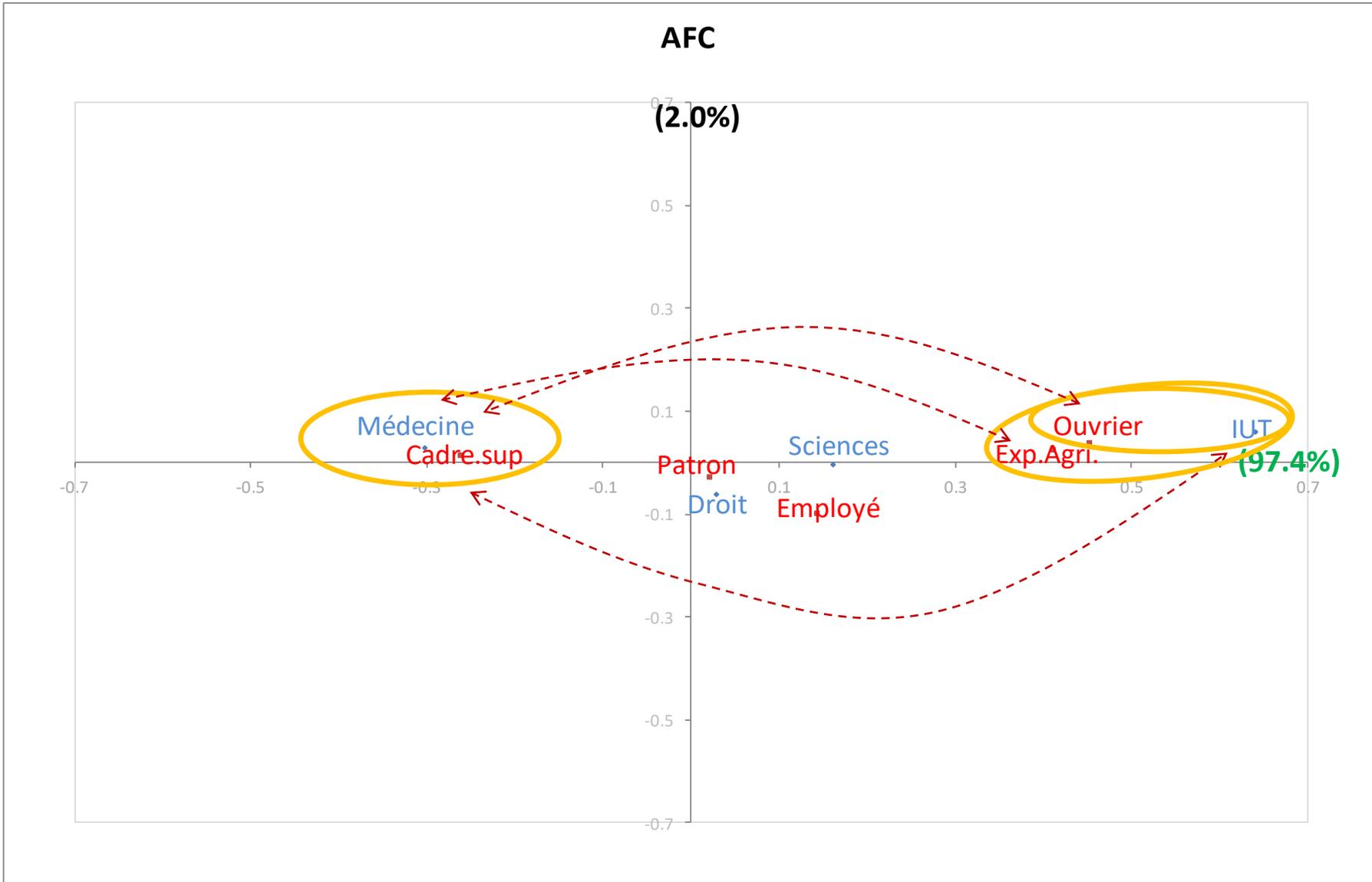
[Seules les contributions supérieures à la moyenne sont affichées c.-à-d. Contrib % > 100 / (K x L)]

CHI-2 contributions

Id	Row	Column	Value	Expected	Std.Resid.	Contrib. (%)	Cumul (%)
1	Ouvrier	IUT	129	62.1	8.49	22.52	22.52
2	Cadre.sup	IUT	79	184.2	-7.75	18.77	41.29
3	Cadre.sup	Médecine	876	680.5	7.49	17.53	58.82
4	Ouvrier	Médecine	127	229.3	-6.76	14.26	73.08
5	Exp.agri	IUT	58	30.5	4.98	7.75	80.83
6	Exp.agri	Médecine	65	112.6	-4.49	6.29	87.12

Il n'y a rien entre « Employé » et « Droit » !!!





CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	80	99	65	58	302
Patron	168	137	208	62	575
Cadre.sup	470	400	876	79	1825
Employé	145	133	135	54	467
Ouvrier	166	193	127	129	615
Bourgeois	638	537	1084	141	2400

Ligne supplémentaire : les enfants de « Bourgeois »

Passage aux profils lignes

Profils lignes

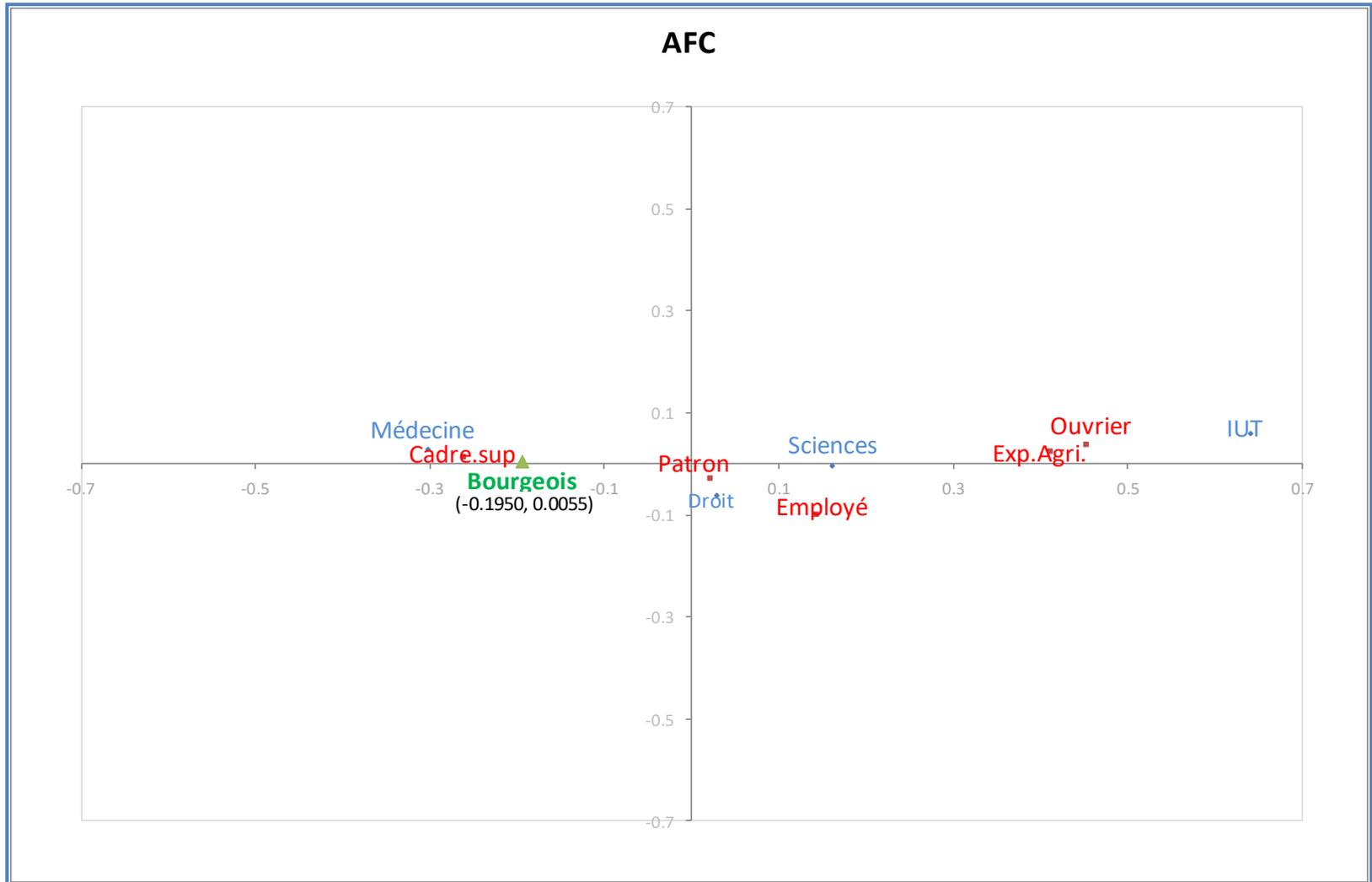
CSP\Filière	Droit	Sciences	Médecine	IUT	Total
Exp.agri	0.265	0.328	0.215	0.192	1
Patron	0.292	0.238	0.362	0.108	1
Cadre.sup	0.258	0.219	0.480	0.043	1
Employé	0.310	0.285	0.289	0.116	1
Ouvrier	0.270	0.314	0.207	0.210	1
Bourgeois	0.266	0.224	0.452	0.059	1

Coordonnées des « Filières » sur le 1^{er} axe , avec $\lambda_1=0.0824$

Values	Coord.1 (G)
Droit	0.0280
Sciences	0.1605
Médecine	-0.3031
IUT	0.6402

Positionnement de « bourgeois » sur le premier axe à partir de la relation

$$F_{k1} = \frac{1}{\sqrt{\lambda_1}} \sum_{l=1}^L \frac{n_{kl}}{n_k} \times G_{l1} = \frac{1}{\sqrt{0.0824}} (0.266 \times 0.0280 + \dots + 0.059 \times 0.6402) = -0.1950$$



Le profil des enfants de « Bourgeois » se rapproche de celui des enfants de « Cadre sup. » et, comme eux, ils sont attirés par la « Médecine ».

Certains logiciels (ex. Tanagra) fournissent les coefficients des fonctions scores qui s'appliquent directement aux profils

$$F_{kh} = \sum_{l=1}^L \frac{n_{kl}}{n_k} \times a_{lh} \quad \text{Où} \quad a_{lh} = \frac{G_{lh}}{\sqrt{\lambda_h}}$$

Fonctions scores fournis par les logiciels

Factor score coefficients for supplementary row

From column values (relative frequency)

Column	Factor 1	Factor 2	Factor 3
Droit	0.097502	-1.469969	0.712131
Sciences	0.559017	-0.066254	-1.617643
Médecine	-1.056027	0.718656	0.223806
IUT	2.230238	1.471908	1.328652

Profil à positionner (« Bourgeois » = 6^{ème} modalité)

	Droit	Sciences	Médecine	IUT
Bourgeois	0.266	0.224	0.452	0.059

$$F_{61} = 0.266 \times 0.097502 + \dots + 0.059 \times 2.230238 = -0.1950$$

$$F_{62} = 0.266 \times (-1.469969) + \dots + 0.059 \times 1.471908 = 0.0055$$

$$F_{63} = 0.266 \times 0.712131 + \dots + 0.059 \times 1.328652 = 0.0065$$

Remarque : On peut calculer la qualité de représentation (QLT = COS²) sur les H premiers axes en faisant le rapport entre le rapport entre la somme du carré des distances à l'origine sur chaque axe et le carré de la distance par rapport au profil moyen (*le profil moyen du tableau originel, sans les lignes supplémentaires*)



1. L'AFC est destinée à l'analyse des grands tableaux croisés
2. C'est une méthode descriptive
3. L'AFC s'applique au-delà des tableaux de contingence
4. Il faut que ce soit un tableau croisé de valeurs positives
5. Et que les notions de marges et de profils soient applicables
6. L'AFC s'intéresse aux structures des relations lignes/colonnes
7. Il peut fournir des résultats même si le χ^2 n'est pas significatif



Analyse factorielle des correspondances

Un exemple réaliste



Prof/Media	Radio	Tel.	Quot.Nat.	Quot.Reg.	Press.Mag.	Press.TV	Total
Agriculteur	96	118	2	71	50	17	354
Petit.Patr.	122	136	11	76	49	41	435
Prof.Cad.Sup	193	184	74	63	103	79	696
Prof.Int.	360	365	63	145	141	184	1258
Employe	511	593	57	217	172	306	1856
Ouvr.Qualif.	385	457	42	174	104	220	1382
Ouvr.Non-Qual.	156	185	8	69	42	85	545
Inactif	1474	1931	181	852	642	782	5862
Total	3297	3969	438	1667	1303	1714	12388

Vont jouer un rôle (trop ?) important

Profils lignes

Prof	Radio	Tel.	Quot.Nat.	Quot.Reg.	Press.Mag.	Press.TV	Total	Weight	Sq. Dist.	Inertia	%inertia
Agriculteur	0.271	0.333	0.006	0.201	0.141	0.048	1	0.0286	0.1293	0.0037	17%
Petit.Patr.	0.280	0.313	0.025	0.175	0.113	0.094	1	0.0351	0.0304	0.0011	5%
Prof.Cad.Sup	0.277	0.264	0.106	0.091	0.148	0.114	1	0.0562	0.1890	0.0106	48%
Prof.Int.	0.286	0.290	0.050	0.115	0.112	0.146	1	0.1016	0.0142	0.0014	6%
Employe	0.275	0.320	0.031	0.117	0.093	0.165	1	0.1498	0.0098	0.0015	7%
Ouvr.Qualif.	0.279	0.331	0.030	0.126	0.075	0.159	1	0.1116	0.0138	0.0015	7%
Ouvr.Non-Qual.	0.286	0.339	0.015	0.127	0.077	0.156	1	0.0440	0.0250	0.0011	5%
Inactif	0.251	0.329	0.031	0.145	0.110	0.133	1	0.4732	0.0029	0.0014	6%
Total	0.266	0.320	0.035	0.135	0.105	0.138	1				

Profils colonnes

Prof	Radio	Tel.	Quot.Nat.	Quot.Reg.	Press.Mag.	Press.TV	Total
Agriculteur	0.029	0.030	0.005	0.043	0.038	0.010	0.029
Petit.Patr.	0.037	0.034	0.025	0.046	0.038	0.024	0.035
Prof.Cad.Sup	0.059	0.046	0.169	0.038	0.079	0.046	0.056
Prof.Int.	0.109	0.092	0.144	0.087	0.108	0.107	0.102
Employe	0.155	0.149	0.130	0.130	0.132	0.179	0.150
Ouvr.Qualif.	0.117	0.115	0.096	0.104	0.080	0.128	0.112
Ouvr.Non-Qual.	0.047	0.047	0.018	0.041	0.032	0.050	0.044
Inactif	0.447	0.487	0.413	0.511	0.493	0.456	0.473
Total	1	1	1	1	1	1	1
Weight	0.2661	0.3204	0.0354	0.1346	0.1052	0.1384	
Sq. Dist.	0.0029	0.0033	0.2944	0.0243	0.0284	0.0272	
Inertia	0.0008	0.0011	0.0104	0.0033	0.0030	0.0038	
%inertia	3%	5%	47%	15%	13%	17%	

Va jouer un rôle (trop ?) important

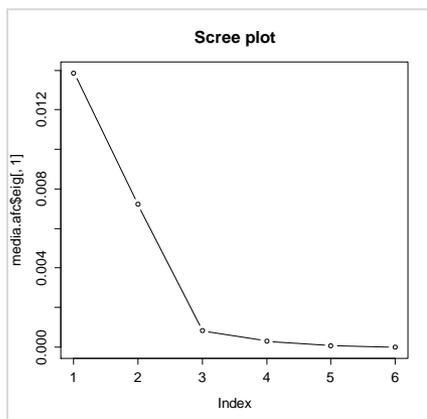


Eigen values

Matrix trace = 0.0223

SQRT(Matrix trace) = 0.1493

Axis	Eigen value	% explained	% cumulated
1	0.013857	62.20%	62.20%
2	0.007211	32.37%	94.56%
3	0.000825	3.70%	98.27%
4	0.000304	1.36%	99.63%
5	0.000083	0.37%	100.00%
Tot.	0.022279	-	-



2 facteurs suffisent amplement

Rows analysis

Characterization				Coord.		Contributions (%)		COS	
Values	Weight	Sq. Dist.	Inertia	coord 1	coord 2	ctr 1	ctr 2	cos 1	cos 2
Agriculteur	0.02858	0.12929	0.00369	0.16615	-0.30961	5.7	38.0	0.21 (0.21)	0.74 (0.95)
Petit.Patr.	0.03511	0.03039	0.00107	0.06838	-0.14315	1.2	10.0	0.15 (0.15)	0.67 (0.83)
Prof.Cad.Sup	0.05618	0.189	0.01062	-0.42998	-0.06087	75.0	2.9	0.98 (0.98)	0.02 (1.00)
Prof.Int.	0.10155	0.01417	0.00144	-0.1066	0.0326	8.3	1.5	0.80 (0.80)	0.08 (0.88)
Employe	0.14982	0.00981	0.00147	0.01573	0.09547	0.3	18.9	0.03 (0.03)	0.93 (0.95)
Ouvr.Qualif.	0.11156	0.01382	0.00154	0.04371	0.10138	1.5	15.9	0.14 (0.14)	0.74 (0.88)
Ouvr.Non-Qual.	0.04399	0.02497	0.0011	0.1178	0.09486	4.4	5.5	0.56 (0.56)	0.36 (0.92)
Inactif	0.4732	0.00285	0.00135	0.03258	-0.0334	3.6	7.3	0.37 (0.37)	0.39 (0.76)

Columns analysis

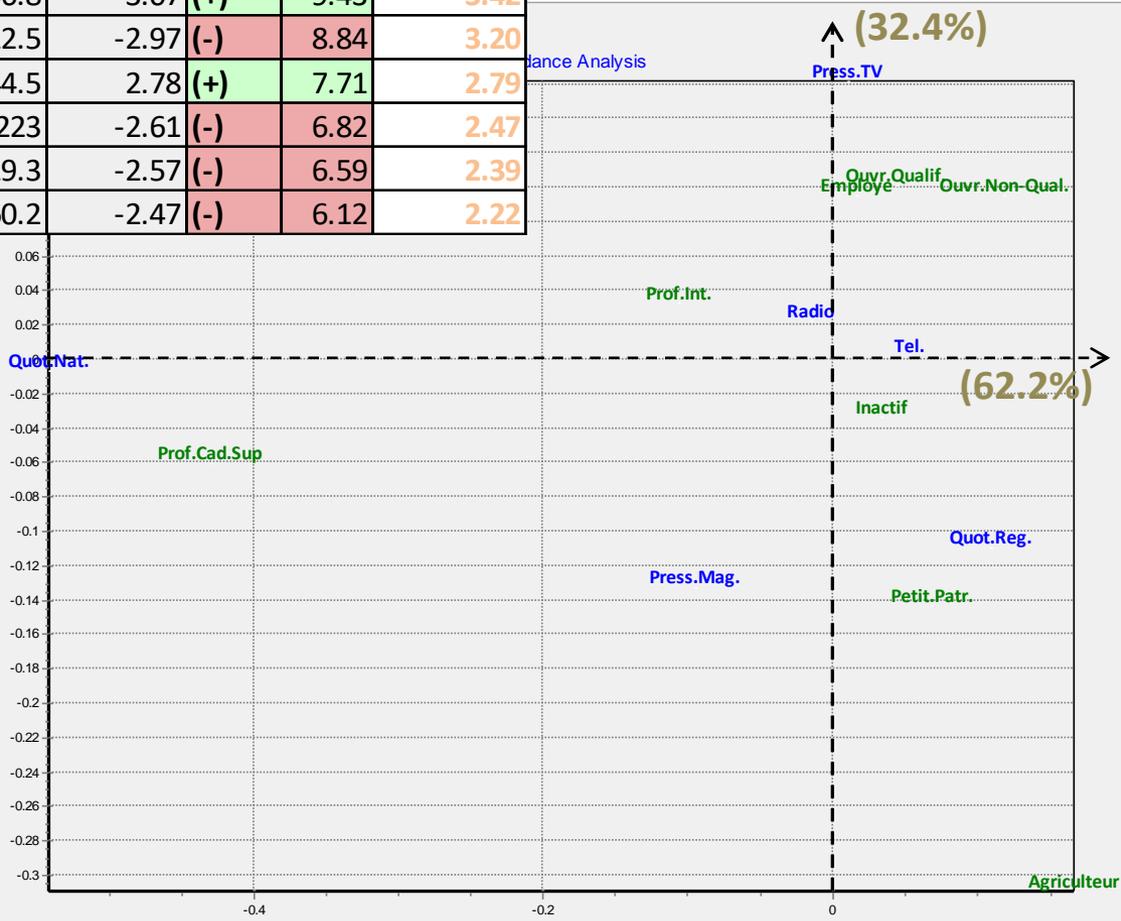
Characterization				Coord.		Contributions (%)		COS	
Values	Weight	Sq. Dist.	Inertia	coord 1	coord 2	ctr 1	ctr 2	cos 1	cos 2
Radio	0.26614	0.0029	0.00077	-0.01494	0.02211	0.43	1.8	0.08 (0.08)	0.17 (0.25)
Tel.	0.32039	0.00334	0.00107	0.05328	0.00208	6.56	0.02	0.85 (0.85)	0.00 (0.85)
Quot.Nat.	0.03536	0.2944	0.01041	-0.54067	-0.00621	74.59	0.02	0.99 (0.99)	0.00 (0.99)
Quot.Reg.	0.13457	0.02434	0.00327	0.10883	-0.10965	11.5	22.44	0.49 (0.49)	0.49 (0.98)
Press.Mag.	0.10518	0.02838	0.00298	-0.09481	-0.1325	6.82	25.61	0.32 (0.32)	0.62 (0.94)
Press.TV	0.13836	0.02724	0.00377	0.00975	0.16161	0.09	50.11	0.00 (0.00)	0.96 (0.96)



CHI-2 contributions

Id	Row	Column	Value	Expected	Std.Resid	Contrib.	%
1	Prof.Cad.Sup	Quot.Nat.	74	24.6	9.96 (+)	99.13	35.92
2	Agriculteur	Press.TV	17	49	-4.57 (-)	20.88	7.57
3	Prof.Cad.Sup	Press.Mag.	103	73.2	3.48 (+)	12.12	4.39
4	Ouvr.Qualif.	Press.Mag.	104	145.4	-3.43 (-)	11.77	4.26
5	Agriculteur	Quot.Reg.	71	47.6	3.39 (+)	11.46	4.15
6	Prof.Cad.Sup	Quot.Reg.	63	93.7	-3.17 (-)	10.04	3.64
7	Employe	Press.TV	306	256.8	3.07 (+)	9.43	3.42
8	Agriculteur	Quot.Nat.	2	12.5	-2.97 (-)	8.84	3.20
9	Prof.Int.	Quot.Nat.	63	44.5	2.78 (+)	7.71	2.79
10	Prof.Cad.Sup	Tel.	184	223	-2.61 (-)	6.82	2.47
11	Ouvr.Non-Qual.	Quot.Nat.	8	19.3	-2.57 (-)	6.59	2.39
12	Petit.Patr.	Press.TV	41	60.2	-2.47 (-)	6.12	2.22

Remarque : Il faudrait mettre « Prof.Cad.Sup » et /ou « Quot.Nat » en éléments supplémentaires. Ces modalités « écrasent » complètement le tableau ... et masquent les autres informations intéressantes.



Analyse factorielle des correspondances

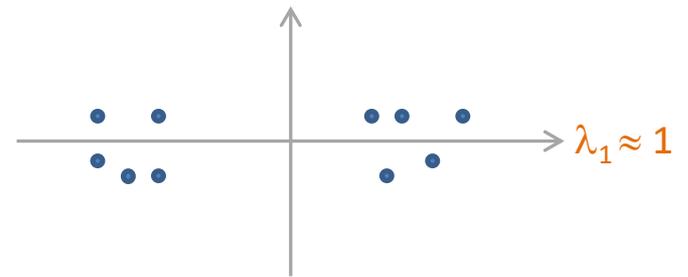
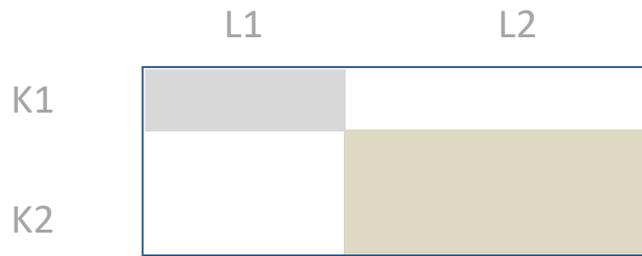
Quelques formes caractéristiques

(Lebart et al., page 93 ; Tenenhaus, pages 241 à 248)



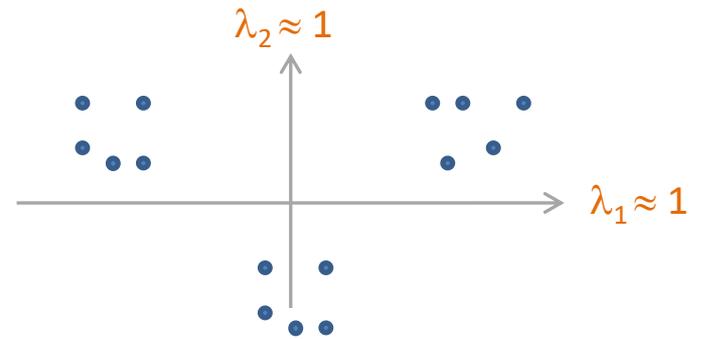
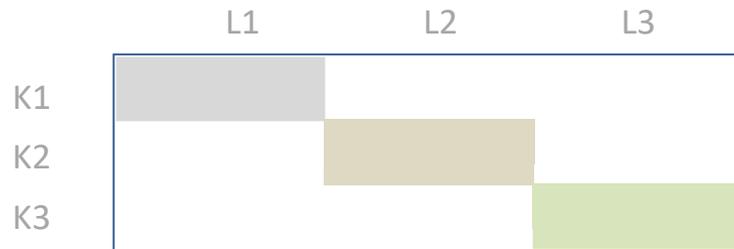
Deux blocs distincts dans le tableau (après réorganisation)

→ peut être intéressant de réaliser deux analyses distinctes



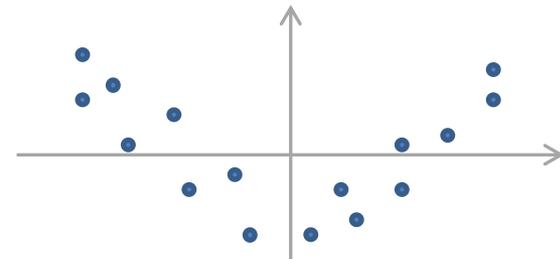
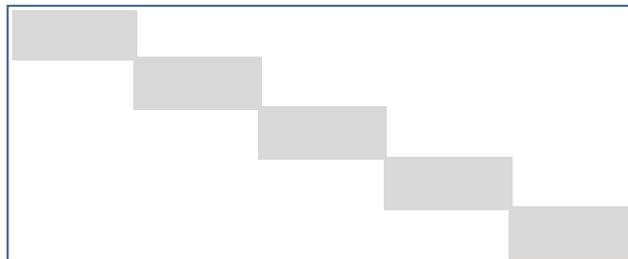
Des blocs distincts dans le tableau (après réorganisation)

→ peut être intéressant de réaliser des analyses distinctes



Diagonale chargée (après réorganisation)

→ Souvent le cas lorsqu'il y a un ordonnancement sous-jacent aux modalités



« Effet Guttman » : redondance de l'information
Relation non linéaire entre les facteurs

Logiciels

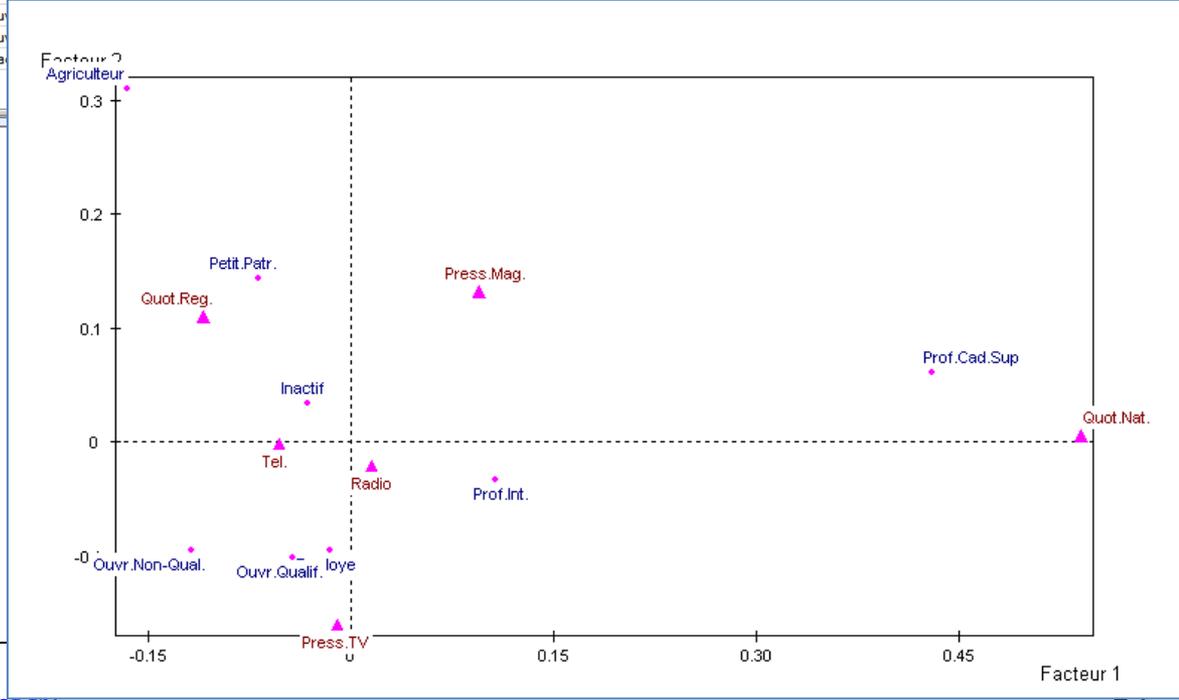


SPAD – La « Rolls-Royce » de l'analyse de données « à la française »

The screenshot shows the SPAD 7.4 software interface. The top menu bar includes 'Projet', 'Edition', 'Affichage', 'Diagramme', 'Dessin', and 'Outils'. The 'Projet' window shows 'Analyse factorielle' with a 'Diagramme' icon. The 'Diagramme' window shows a flow from 'media_prof_afc. Lebart & al.-Tab 1.3-10-p.104' to 'AFC'. The 'Méthodes' window lists various statistical methods under 'Analyses factorielles', with 'AFC - Analyse Factorielle des Correspondances Simples' selected. The 'Données' window displays a table with 8 rows and 7 columns.

Index	Prof	Radio	Tel.	Quot.Nat.	Quot.Reg.	Press.Mag.	Press.TV
1	Agriculteur	96	118	2	71	50	17
2	Petit.Patr.	122	136	11	76	49	41
3	Prof.Cad.Sup	193	184	74	63	103	79
4	Prof.Int.	360	365	63	145	141	184
5	Employe	511	593	57	217	172	306
6	Ouvr.						
7	Ouvr.						
8	Ina						

Des calculs et une présentation qui font référence !



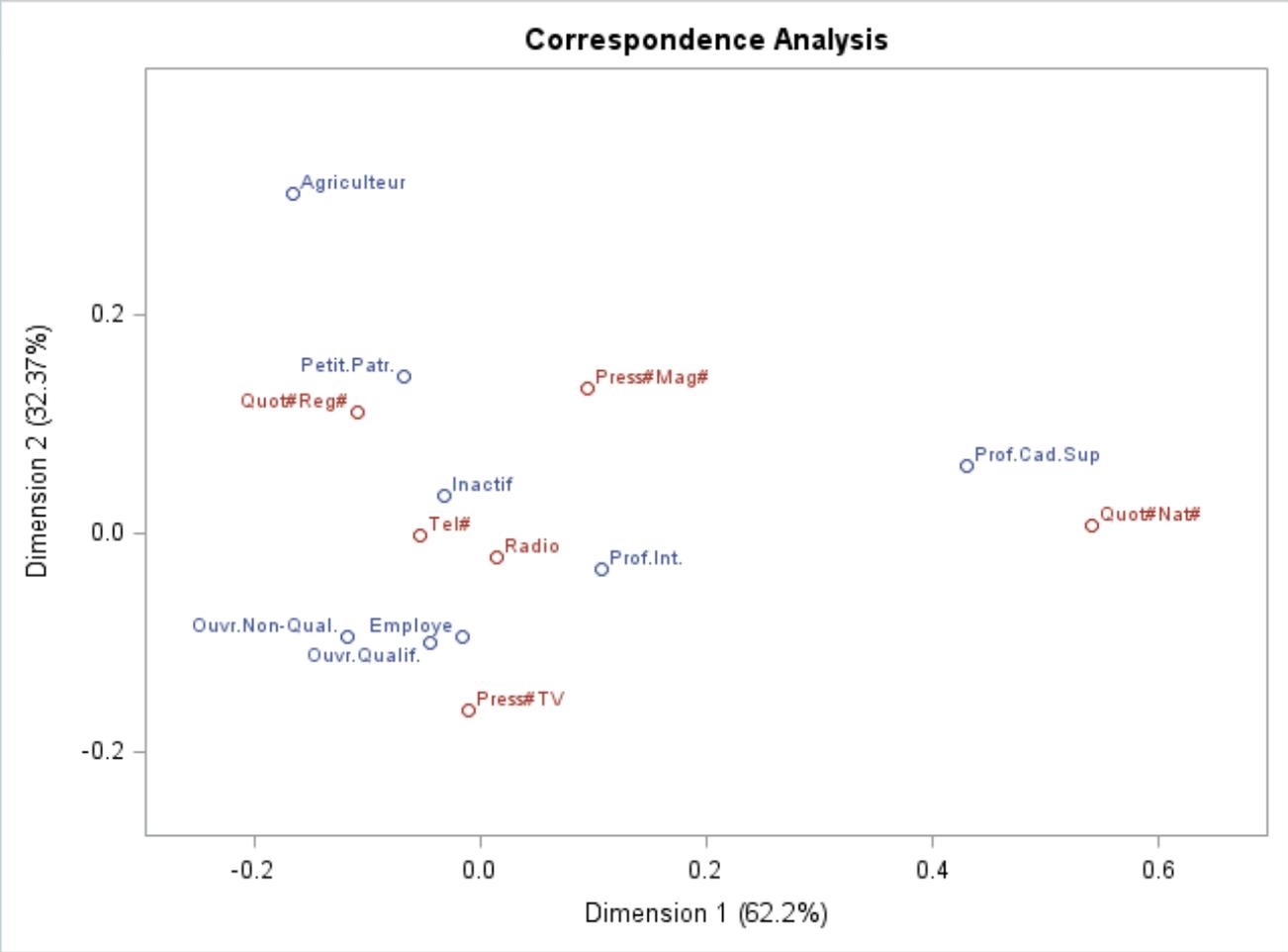
```
proc corresp data = mesdata.media dimens = 2;
var Radio -- Press_TV;
id Prof;
run;
```

Le Système SAS
The CORRESP Procedure

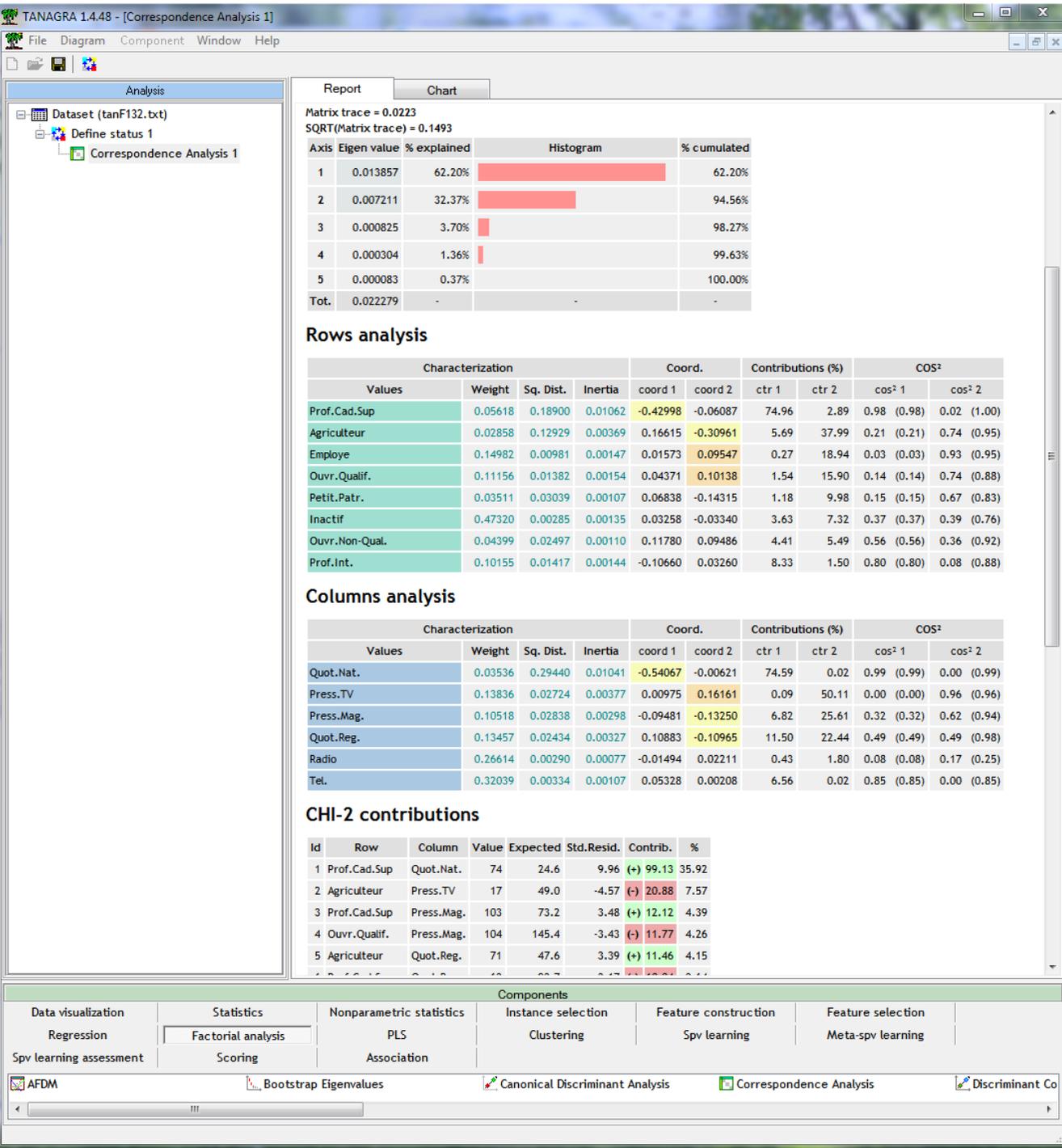
Décomposition de l'inertie et du Khi-2

Valeur singulière	Inertie principale	Khi-2	Pourcentage	Pourcent. cumulé	12	24	36	48	60
					-----+-----+-----+-----+-----				
0.11772	0.01386	171.664	62.20	62.20	*****				
0.08492	0.00721	89.326	32.37	94.56	*****				
0.02872	0.00082	10.217	3.70	98.27	**				
0.01743	0.00030	3.764	1.36	99.63	*				
0.00909	0.00008	1.024	0.37	100.00					
Total	0.02228	275.995	100.00						

Degrés de liberté = 35



Fournit les indicateurs usuels. Un des rares à afficher explicitement la décomposition du χ^2 .

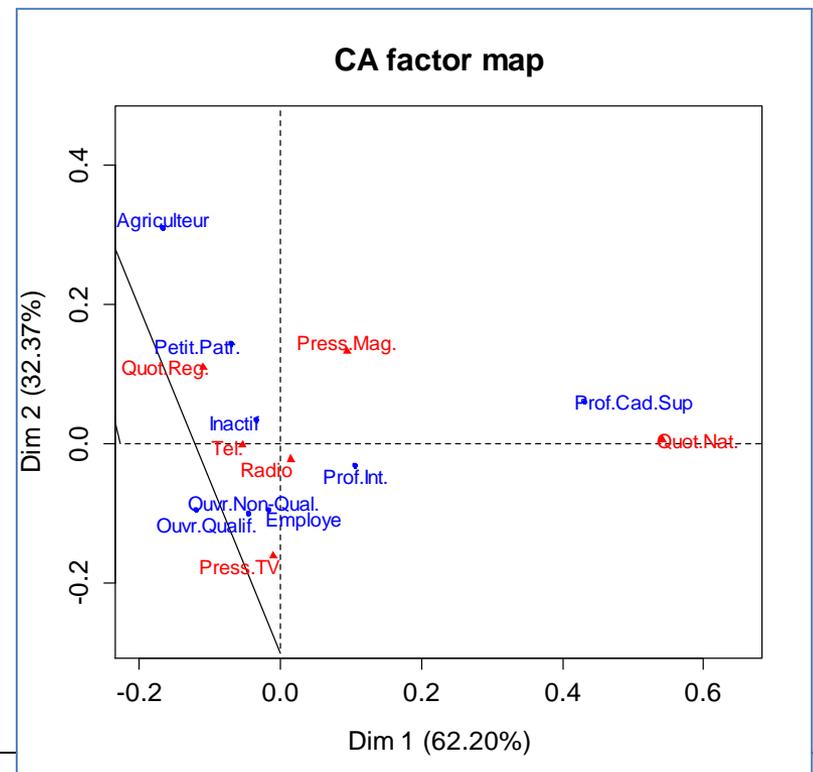
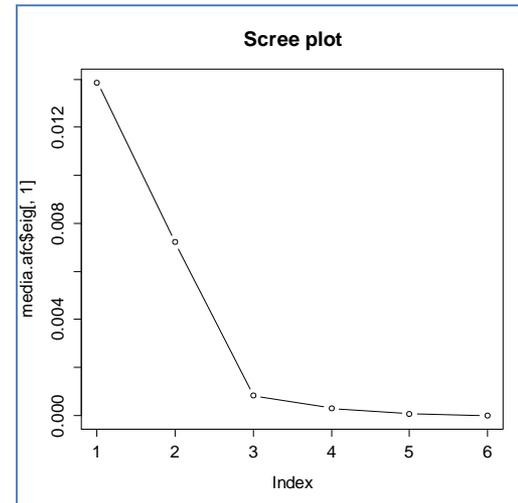


Tanagra – Axé sur la facilité d'utilisation

Tri des modalités les plus contributives pour chaque axe. Utilisation de codes couleurs pour une meilleure identification des similitudes et différences.

Exclusivité : tableau des contributions au χ^2 avec les résidus standardisés.

```
#charger le package
library(FactoMineR)
#lancer l'AFC
media.afc <- CA(media,ncp=2,graph=FALSE)
#liste des objets disponibles
print(attributes(media.afc))
#quelques statistiques sur le tableau de données
print(media.afc$call)
#tableau des valeurs propres et Scree plot
print(media.afc$eig)
plot(media.afc$eig[,1],type="b",main="Scree plot")
#coordonnées, contributions et cos2 - lignes
print(media.afc$row)
#coordonnées, contributions et cos2 - colonnes
print(media.afc$col)
#graphique
plot(media.afc,cex=0.75)
```



Bibliographie

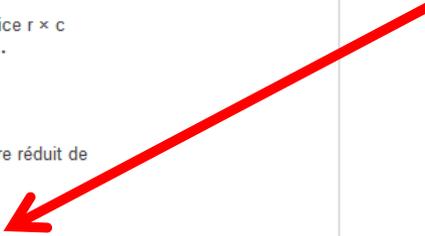


The screenshot shows a Google search interface with the query "analyse factorielle des correspondances". The search results are displayed under the "Web" tab. A blue oval highlights the text "Environ 48 100 résultats (0,30 secondes)". The results list several links:

- [Analyse factorielle des correspondances - Wikipédia](http://fr.wikipedia.org/wiki/Analyse_factorielle_des_correspondances)
fr.wikipedia.org/wiki/Analyse_factorielle_des_correspondances
- [\[PDF\] Analyse Factorielle des Correspondances \(AFC\) - WikiStat](http://wikistat.fr/pdf/st-m-explo-afc.pdf)
wikistat.fr/pdf/st-m-explo-afc.pdf
- [\[PDF\] introduction à l'analyse factorielle des correspondances](http://web.univ-pau.fr/RECHERCHE/SET/.../INTRODUCTION_AFC.pdf)
web.univ-pau.fr/RECHERCHE/SET/.../INTRODUCTION_AFC.pdf
- [Cours analyse factorielle des correspondances - YouTube](http://www.youtube.com/playlist?list=PL02D3C245C7AAD789)
www.youtube.com/playlist?list=PL02D3C245C7AAD789
- [\[PDF\] 3 Analyse Factorielle des Correspondances](http://geai.univ-brest.fr/~carpentier/2004-2005/PSRS83B-2.pdf)
geai.univ-brest.fr/~carpentier/2004-2005/PSRS83B-2.pdf

De très nombreux supports de cours d'excellentes factures sont disponibles.
Profusion d'exemples traités.

Y compris sur YouTube !



Escofier B., Pagès J., « Analyses factorielles simples et multiples », Dunod, 2008 ; chapitre 3, pages 63 à 83.

Lebart L., Morineau A., Piron M., « Statistique exploratoire multidimensionnelle », Dunod, 3^{ème} édition, 2000 ; Section 1.3, pp. 67 à 107.

Saporta G., « Probabilités, Analyse des Données et Statistique », Tehcnip, 2006 ; chapitre 9, pp. 201 à 217.

Tenenhaus M., « Statistique : Méthodes pour décrire, expliquer et prévoir », Dunod, 2006 ; chapitre 7, pages 197 à 249.



- « [AFC – Association médias et professions](#) » (Mars 2008)
- « [Analyse factorielle des correspondances avec R](#) » (Mai 2009)
- « [Analyse des correspondances – Comparaisons](#) » (Déc. 2012)

