

Analyse multivariée

Analyse factorielle
des
correspondances

Présentation de l'AFC

- Etude du lien entre deux variables qualitatives
 - Mesure quantitative du lien par une distance appelée distance du chi-deux ou distance du χ^2
 - Etude de la nature du lien en cherchant les correspondances entre les modalités de la 1^{ère} variable avec la modalité de la 2^{ème} variable.

Etude du lien de 2 caractères qualitatifs

- Les seuls calculs possibles sur des caractères qualitatifs sont des **effectifs et/ou des fréquences**
- Chercher un lien entre deux caractères qualitatifs reviendra à étudier l'ensemble des effectifs des sous populations définies par les couples de modalités (x_i, y_j) prises respectivement par C_1 et C_2 .

C_2	y_1		y_j		y_l
C_1					
x_1	$n_{1,1}$				$n_{1,l}$
x_i	$n_{i,1}$		$n_{i,j}$		$n_{i,l}$
x_k	$n_{k,1}$		$n_{k,j}$		$n_{k,l}$

$n_{i,j}$ est le nombre d'individus ω tels que $C_1(\omega) = x_i$ et $C_2(\omega) = y_j$

Exemple : Etude du lien entre la couleur des yeux et la couleur des cheveux

Cheveux	bruns	chatains	roux	blonds
Yeux				
bleus	11	10	0	8
verts	5	8	1	4
marrons	16	22	3	12

Effectifs marginaux

Pour faire des interprétations sur des correspondances entre des modalités de C_1 et des modalités de C_2 , il faut compléter le tableau avec les effectifs de C_1 sans C_2 et des effectifs de C_2 sans C_1 . Ces effectifs sont appelés effectifs marginaux (en marge de)

	y_1		y_j		y_l	
x_1	$n_{1,1}$				$n_{1,l}$	$n_{1,.}$
x_i	$n_{i,1}$		$n_{i,j}$		$n_{i,l}$	$n_{i,.}$
x_k	$n_{k,1}$		$n_{k,j}$		$n_{k,l}$	$n_{k,.}$
	$n_{.,1}$		$n_{.,j}$		$n_{.,l}$	n

Effectifs marginaux

$$\text{pour } C_1: n_{i,.} = \sum_{j=1}^l n_{i,j} \quad \text{pour } C_2: n_{.,j} = \sum_{i=1}^k n_{i,j}$$

Effectif total

$$n = \sum_{j=1}^l n_{.,j} = \sum_{i=1}^k n_{i,.} = \sum_{i=1}^k \sum_{j=1}^l n_{i,j}$$

Exemple : Etude du lien entre la couleur des yeux et la couleur des cheveux

Yeux	Cheveux				
	bruns	chatains	roux	blonds	
bleus	11	10	0	8	29
verts	5	8	1	4	18
marrons	16	22	3	12	53
	32	40	4	24	100

Comparaison des effectifs non pertinente

Fréquences marginales

Des effectifs ne sont pas directement comparables tandis que des fréquences sont toujours comparables

	y_1		y_j		y_l	
x_1	$f_{1,1}$				$f_{1,l}$	$f_{1,.}$
$f_{i,j}$ est la proportion d'individus ω dans P tels que $C_1(\omega) = x_i$ et $C_2(\omega) = y_j$						
x_i	$f_{i,1}$		$f_{i,j}$		$f_{i,l}$	$f_{i,.}$
x_k	$f_{k,1}$		$f_{k,j}$		$f_{k,l}$	$f_{k,.}$
	$f_{.,1}$		$f_{.,j}$		$f_{.,l}$	1

Fréquences marginales

pour C_1 : $f_{i,.} = \sum_{j=1}^l f_{i,j}$ pour C_2 : $f_{.,j} = \sum_{i=1}^k f_{i,j}$

$$1 = \sum_{j=1}^l f_{.,j} = \sum_{i=1}^k f_{i,.} = \sum_{i=1}^k \sum_{j=1}^l f_{i,j}$$

Exemple : Etude du lien entre la couleur des yeux et la couleur des cheveux

Cheveux	bruns	chatains	roux	blonds	
Yeux					
bleus	0,11	0,1	0	0,08	0,29
verts	0,05	0,08	0,01	0,04	0,18
marrons	0,16	0,22	0,03	0,12	0,53
	0,32	0,4	0,04	0,24	1

Profils et fréquences conditionnelles

Profils lignes	y_1	y_j	y_i	
x_1	$f_{1/1}$		$f_{1/i}$	$f_{1..}$
x_i	$f_{1/i}$	$f_{j/i}$	$f_{i/i}$	$f_{i..}$
x_k	$f_{1/k}$	$f_{j/k}$	$f_{i/k}$	$f_{k..}$
	$f_{.,1}$	$f_{.,j}$	$f_{.,i}$	

Profil ligne : répartition en fréquences du caractère C_2 dans une sous population définie par $P_{i..} = \{\omega / C_1(\omega) = x_i\}$

$$f_{j/i} = \frac{n_{i,j}}{n_{i..}}$$

comparable avec f_j

Profil colonne : répartition en fréquences de C_1 dans une sous population définie par $P_{.,j} = \{\omega / C_2(\omega) = y_j\}$

$$f_{i/j} = \frac{n_{i,j}}{n_{.,j}}$$

comparable avec f_i

		bruns	châtains	roux	blonds
Profils lignes	bleus	0,38	0,34	0	0,28
	verts	0,28	0,44	0,06	0,22
	marrons	0,3	0,42	0,06	0,23
		0,32	0,4	0,04	0,24

Profils colonnes					
	bruns	châtains	roux	blonds	
bleus	0,34	0,25	0	0,33	0,29
verts	0,16	0,2	0,25	0,17	0,18
marrons	0,5	0,55	0,75	0,5	0,53

Distance entre les profils et les profils moyens

- On compare les profils lignes avec le profil moyen

$$d^2(PL_i, PLM) = \sum_{j=1}^s \left(\frac{(f_{j/i} - f_{.,j})^2}{f_{.,j}} \right)$$

- On compare les profils colonnes avec le profil moyen

$$d^2(PC_j, PCM) = \sum_{i=1}^r \left(\frac{(f_{i/j} - f_{i,.})^2}{f_{i,.}} \right)$$

Inertie et distance du chi-deux

- L'inertie du nuage des profils lignes est définie par

$$\sum_{i=1}^r (d^2(PL_i, PLM)) = \sum_{i=1}^r \sum_{j=1}^s \left(\frac{(f_{j/i} - f_{.,j})^2}{f_{.,j}} \right)$$

- L'inertie du nuage des profils colonnes est définie par

$$\sum_{j=1}^s d^2(PC_j, PCM) = \sum_{j=1}^s \sum_{i=1}^r \left(\frac{(f_{i/j} - f_{i,.})^2}{f_{i,.}} \right)$$

Présentation du cas

- On a interrogé 10005 étudiants à qui on a demandé le type d'études qu'ils poursuivent et la catégorie socio-professionnelle (CSP) de leur chef de famille.
- On cherche à savoir
 - s'il y a un lien entre le type d'étude et la CSP.
 - si la réponse est oui alors quelle est la nature de ce lien

Les données du cas

Tableau d'effectifs CSP x Type d'études poursuivies

	Droit	Sci Eco	Lettres	Sciences	Méd. & Dent.	Pharma.	Plur. Disc.	IUT	Totaux
Exp. Agr.	80	36	134	99	65	28	11	58	511
Sal. Agr.	6	2	15	6	4	1	1	4	39
Patron	168	74	312	137	208	53	21	62	1035
Pro. Lib.	470	191	806	400	876	164	45	79	3031
Cad. Moy.	236	99	493	264	281	56	36	87	1552
Employé	145	52	281	133	135	30	20	54	850
Ouvrier	166	64	401	193	127	23	28	129	1131
Pers. Serv.	16	6	27	11	8	2	2	8	80
Autres.	305	115	624	247	301	47	47	90	1776
Totaux	1592	639	3093	1490	2005	404	211	571	10005

Tableau de contingence

	Droit	Sci Eco	Lettres	Sciences	Méd. & Dent.	Pharma.	Plur. Disc.	IUT	Totaux
Exp. Agr.	80	36	134	99	65	28	11	58	511
Sal. Agr.	6	2	15	6	4	1	1	4	39
Patron	168	74	312	137	208	53	21	62	1035
Pro. Lib.	470	191	806	400	876	164	45	79	3031
Cad. Moy.	236	99	493	264	281	56	36	87	1552
Employé	145	52	281	133	135	30	20	54	850
Ouvrier	166	64	401	193	127	23	28	129	1131
Pers. Serv.	16	6	27	11	8	2	2	8	80
Autres.	305	115	624	247	301	47	47	90	1776
Totaux	1592	639	3093	1490	2005	404	211	571	10005

%	Droit	Sci Eco	Lettres	Sciences	Méd# & Dent#	Pharma#	Plur# Disc#	IUT	Totaux
Exp. Agr.	0,80	0,36	1,34	0,99	0,65	0,28	0,11	0,58	5,11
Sal. Agr.	0,06	0,02	0,15	0,06	0,04	0,01	0,01	0,04	0,39
Patron	1,68	0,74	3,12	1,37	2,08	0,53	0,21	0,62	10,35
Pro. Lib.	4,70	1,91	8,06	4,00	8,76	1,64	0,45	0,79	30,30
Cad. Moy.	2,36	0,99	4,93	2,64	2,81	0,56	0,36	0,87	15,51
Employé	1,45	0,52	2,81	1,33	1,35	0,30	0,20	0,54	8,50
Ouvrier	1,66	0,64	4,01	1,93	1,27	0,23	0,28	1,29	11,30
Pers. Serv.	0,16	0,06	0,27	0,11	0,08	0,02	0,02	0,08	0,80
Autres.	3,05	1,15	6,24	2,47	3,01	0,47	0,47	0,90	17,75
Totaux	15,91	6,39	30,92	14,89	20,04	4,04	2,11	5,71	100,00

Profils Lignes

Profils des lignes									
%	Droit	Sci Eco	Lettres	Sciences	Méd# & Dent#	Pharma#	Plur# Disc#	IUT	Total
Exp. Agr.	15,66	7,05	26,22	19,37	12,72	5,48	2,15	11,35	100,00
Sal. Agr.	15,38	5,13	38,46	15,38	10,26	2,56	2,56	10,26	100,00
Patron	16,23	7,15	30,14	13,24	20,10	5,12	2,03	5,99	100,00
Pro. Lib.	15,51	6,30	26,59	13,20	28,90	5,41	1,48	2,61	100,00
Cad. Moy.	15,21	6,38	31,77	17,01	18,11	3,61	2,32	5,61	100,00
Employé	17,06	6,12	33,06	15,65	15,88	3,53	2,35	6,35	100,00
Ouvrier	14,68	5,66	35,46	17,06	11,23	2,03	2,48	11,41	100,00
Pers. Serv.	20,00	7,50	33,75	13,75	10,00	2,50	2,50	10,00	100,00
Autres.	17,17	6,48	35,14	13,91	16,95	2,65	2,65	5,07	100,00
Profil moyen	15,91	6,39	30,92	14,89	20,04	4,04	2,11	5,71	100,00

Contributions à la statistique du khi-2 totale									
%	Droit	Sci Eco	Lettres	Sciences	Méd# & Dent#	Pharma#	Plur# Disc#	IUT	Somme
Exp, Agr,	0,00	0,07	0,76	1,45	2,87	0,55	0,00	5,98	11,69
Sal, Agr,	0,00	0,02	0,15	0,00	0,39	0,04	0,01	0,30	0,91
Patron	0,01	0,20	0,04	0,40	0,00	0,63	0,01	0,03	1,32
Pro, Lib,	0,07	0,01	3,84	1,23	24,92	2,97	1,18	10,71	44,92
Cad, Moy,	0,10	0,00	0,08	0,98	0,61	0,15	0,07	0,01	1,99
Employé	0,15	0,02	0,27	0,07	1,54	0,11	0,05	0,13	2,33
Ouvrier	0,23	0,20	1,58	0,75	9,19	2,36	0,15	13,50	27,97
Pers, Serv,	0,18	0,03	0,04	0,02	0,84	0,10	0,01	0,54	1,76
Autres,	0,37	0,01	2,15	0,24	1,78	1,79	0,51	0,27	7,11
Somme	1,11	0,55	8,91	5,13	42,13	8,70	1,98	31,47	100,00

Profils Colonnes

Profils de colonnes									
%	Droit	Sci Eco	Lettres	Sciences	Méd# & Dent#	Pharma#	Plur# Disc#	IUT	Profil moyen
Exp, Agr,	5,03	5,63	4,33	6,64	3,24	6,93	5,21	10,16	5,11
Sal, Agr,	0,38	0,31	0,49	0,40	0,20	0,25	0,47	0,70	0,39
Patron	10,55	11,58	10,09	9,19	10,37	13,12	9,95	10,86	10,35
Pro, Lib,	29,52	29,89	26,06	26,85	43,69	40,59	21,33	13,84	30,30
Cad, Moy,	14,82	15,49	15,94	17,72	14,02	13,86	17,06	15,24	15,51
Employé	9,11	8,14	9,09	8,93	6,73	7,43	9,48	9,46	8,50
Ouvrier	10,43	10,02	12,96	12,95	6,33	5,69	13,27	22,59	11,30
Pers, Serv,	1,01	0,94	0,87	0,74	0,40	0,50	0,95	1,40	0,80
Autres,	19,16	18,00	20,17	16,58	15,01	11,63	22,27	15,76	17,75
Total	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00

Contributions à la statistique du khi-2 totale									
%	Droit	Sci Eco	Lettres	Sciences	Méd# & Dent#	Pharma#	Plur# Disc#	IUT	Somme
Exp, Agr,	0,00	0,07	0,76	1,45	2,87	0,55	0,00	5,98	11,69
Sal, Agr,	0,00	0,02	0,15	0,00	0,39	0,04	0,01	0,30	0,91
Patron	0,01	0,20	0,04	0,40	0,00	0,63	0,01	0,03	1,32
Pro, Lib,	0,07	0,01	3,84	1,23	24,92	2,97	1,18	10,71	44,92
Cad, Moy,	0,10	0,00	0,08	0,98	0,61	0,15	0,07	0,01	1,99
Employé	0,15	0,02	0,27	0,07	1,54	0,11	0,05	0,13	2,33
Ouvrier	0,23	0,20	1,58	0,75	9,19	2,36	0,15	13,50	27,97
Pers, Serv,	0,18	0,03	0,04	0,02	0,84	0,10	0,01	0,54	1,76
Autres,	0,37	0,01	2,15	0,24	1,78	1,79	0,51	0,27	7,11
Somme	1,11	0,55	8,91	5,13	42,13	8,70	1,98	31,47	100,00

Décomposition de l'inertie sur les axes

Décomposition de l'inertie et du Khi-2

Valeur singulière	Inertie principale	Khi-2	Pourcentage	Pourcentage cumulé	17	34	51	68	85
					-----+				
0,19945	0,03978	398,009	83,5	83,5	*****				
0,07461	0,00557	55,698	11,69	95,19	***				
0,0338	0,00114	11,43	2,4	97,59	*				
0,03092	0,00096	9,566	2,01	99,59	*				
0,01136	0,00013	1,292	0,27	99,86					
0,0076	0,00006	0,577	0,12	99,98					
0,00269	0,00001	0,072	0,02	100					
Total	0,04764	476,644	100						

Il y a 9 CSP et 8 types d'études poursuivies. Degrés de liberté = 56 = (9-1).(8-1).

Le seuil à 5% d'un χ^2 à 56 ddl est de 74,5. La valeur calculée 476,644 est bien supérieure à ce seuil \Rightarrow CSP et Type d'études sont assez liées

$$\sum_{i=1}^9 \sum_{j=1}^8 \frac{\left(\frac{n_{i,j} \cdot n_{.,j}}{n} \right)^2}{\frac{n_{i.,} \cdot n_{.,j}}{n}} = 476,644$$

Contributions et Qualité de représentation

Contributions partielles à l'inertie des profils lignes

	Dim1	Dim2	Dim3
Exp. Agr.	0,0691	0,4709	0,1235
Sal. Agr.	0,0102	0,0007	0,0024
Patron	0,0011	0,0167	0,2467
Pro. Lib.	0,5309	0,0316	0,0694
Cad. Moy.	0,0094	0,0036	0,0248
Employé	0,0227	0,0161	0,0429
Ouvrier	0,3176	0,0213	0,3758
Pers. Serv.	0,017	0,0005	0,065
Autres.	0,022	0,4387	0,0496
Totaux	1,00	1,00	1,00

Carré des cosinus des profils lignes avec les axes

	Dim1	Dim2	Dim3	Totaux
Exp, Agr,	0,4935	0,4708	0,0253	0,9896
Sal, Agr,	0,9309	0,0083	0,0062	0,9454
Patron	0,0671	0,1474	0,4478	0,6623
Pro, Lib,	0,987	0,0082	0,0037	0,9989
Cad, Moy,	0,3952	0,0212	0,0298	0,4462
Employé	0,8121	0,0807	0,044	0,9368
Ouvrier	0,9483	0,0089	0,0322	0,9894
Pers, Serv,	0,8045	0,0036	0,0884	0,8965
Autres,	0,259	0,7212	0,0167	0,9969

Contributions partielles à l'inertie des profils colonnes

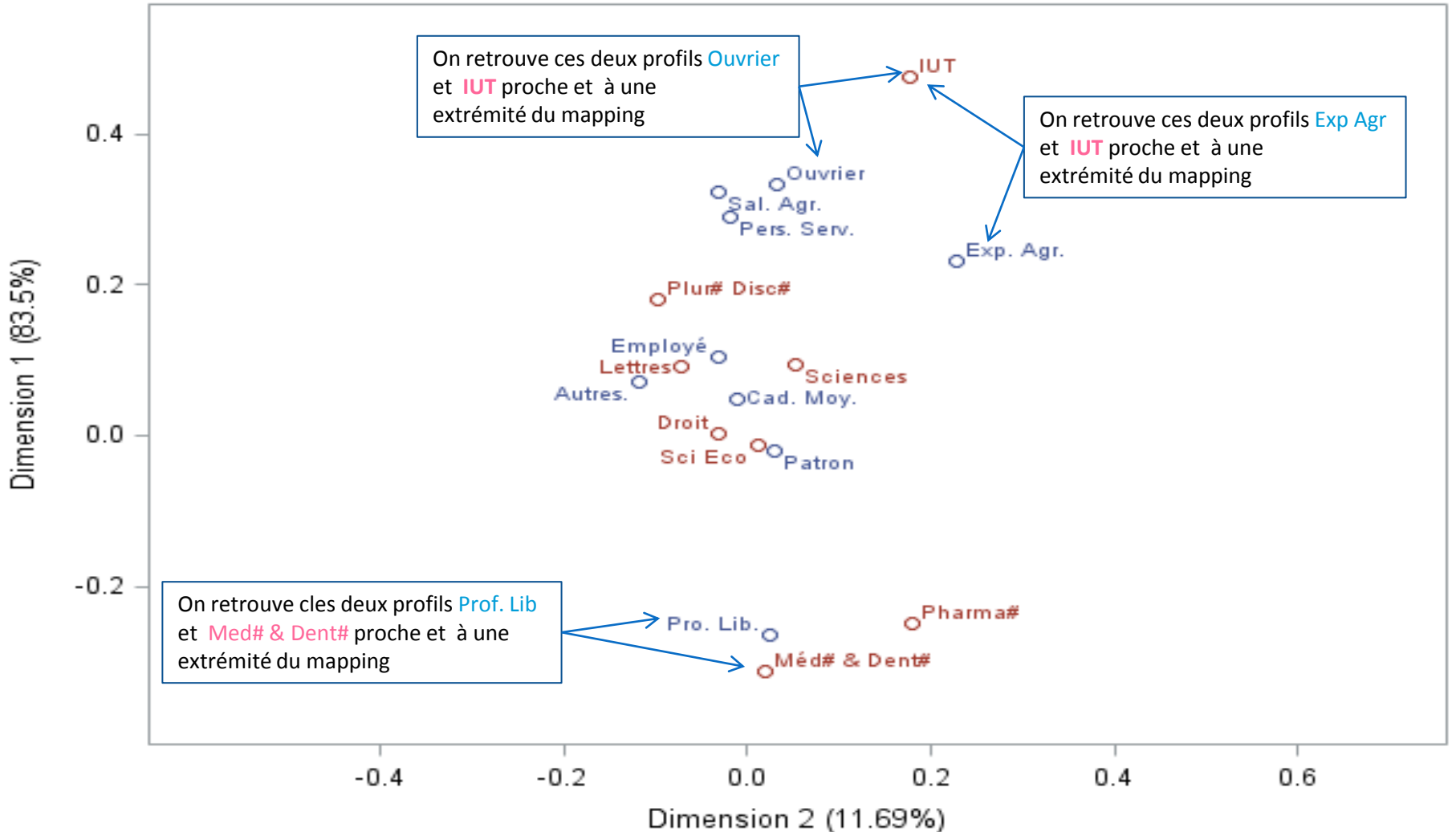
	Dim1	Dim2	Dim3
Droit	0,0001	0,0296	0,2471
Sci Eco	0,0002	0,0016	0,176
Lettres	0,0635	0,2954	0,0238
Sciences	0,0339	0,0706	0,0587
Méd# & Dent#	0,4959	0,0144	0,1912
Pharma#	0,0634	0,2353	0,2451
Plur# Disc#	0,0175	0,0358	0,0177
IUT	0,3256	0,3173	0,0404
Totaux	1,00	1,00	1,00

Carré des cosinus pour des profils colonnes

	Dim1	Dim2	Dim3	Totaux
Droit	0,00	0,31	0,53	0,85
Sci Eco	0,03	0,03	0,76	0,83
Lettres	0,59	0,39	0,01	0,99
Sciences	0,55	0,16	0,03	0,74
Méd# & Dent#	0,98	0,00	0,01	1,00
Pharma#	0,61	0,32	0,07	0,99
Plur# Disc#	0,74	0,21	0,02	0,97
IUT	0,8639	0,1178	0,0031	0,9848

Mappings des profils sur les axes

Correspondence Analysis



Questions ?

