

## ANALYSE MULTIDIMENSIONNELLE DES DONNEES

### SUJET DONNE EN 2005/2006

N.B. Calculatrice et documents autorisés

Une étude a été effectuée auprès d'apprentis d'une région de Suisse romande. Il s'agit d'une enquête par questionnaire standardisé soumise à un échantillon de jeunes (16 à 20 ans) qui effectuent une formation dans une entreprise industrielle ou artisanale tout en suivant des cours théoriques dans une école professionnelle. La recherche portait sur la perception de différents aspects du travail et de la formation (choix de la profession, sens du travail, préférences professionnelles, relations avec le patron et les autres apprentis, etc.). Une question leur demandait d'indiquer quelles étaient les trois professions qui leur plaisaient le plus et les trois qui leur plaisaient le moins parmi une liste de quatorze professions dont l'accès pour ces apprentis était difficile voire impossible de par la formation qu'elles requéraient ou l'investissement financier qu'elles nécessitaient ou encore parce qu'elles n'existaient pas dans la région d'où provenaient les sujets.

Pour chacune des quatorze professions, deux modalités ont été sélectionnées : le fait de l'avoir classée parmi celles qui plaisent le plus ou celles qui plaisent le moins (que nous nommerons acceptation et rejet). Un échantillon de 90 sujets tiré au hasard a rempli le questionnaire. Nous obtenons ainsi une matrice rectangulaire composée de 90 lignes (les sujets) et 28 colonnes (deux fois - acceptation et rejet - 14 professions) contenant les codes 0 ou 1 selon que la modalité est absente ou présente. Cette matrice a été soumise à une analyse factorielle des correspondances. Une partie des résultats de cette AFC est donnée dans les tableaux et le graphique ci-dessous :

*Liste des 14 professions :*

Acteur (act), agriculteur (agr), assistant social (ass), avocat (avo), banquier (ban), chimiste (chi), ingénieur (ing), instituteur (ins), journaliste (jou), médecin (méd), mineur (min), photographe (pho), pilote automobile (pil), psychologue (psy).

*Val. Propres & stat. associées.*

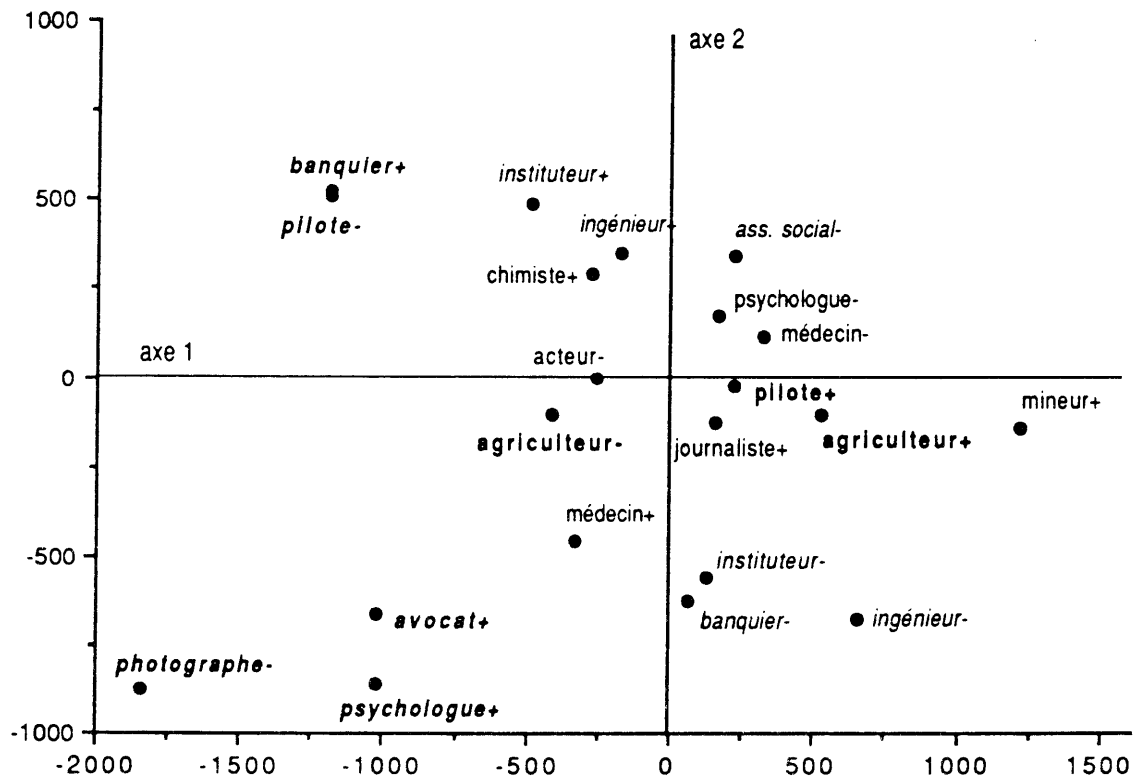
FACTEUR	VALEUR PROPRE	% EXPL.	% CUMUL.	HISTOGRAMME DES VALEURS PROPRES
1	.164	21.98	21.98	*****
2	.088	11.80	33.79	*****
3	.078	10.56	44.35	*****
4	.064	8.64	53.00	*****
5	.051	6.83	59.84	
6	.042	5.71	65.55	
7	.042	5.63	71.19	
8	.032	4.32	75.51	
9	.030	4.07	79.58	

*Scores, contributions et qualités de représentation des individus colonnes*

N.B. Toutes les valeurs du tableau sont indiquées en millièmes. Par exemple la valeur 65 (première case) correspond à 0,065, ou encore à 6,5%. Les colonnes étiquetées "COR" indiquent les qualités de représentation sur les axes ( $\cos^2$ ), et les colonnes étiquetées CTR indiquent les contributions relatives (ou inerties relatives) des modalités.

Modalités	Poids	INR	Facteur 1			Facteur 2			Facteur 3		
			Coord 1	COR	CTR	Coord 2	COR	CTR	Coord 3	COR	CTR
act+	65	22	225	204	20	-24	2	0	48	9	2
agr+	39	35	527	413	66	-106	17	5	-169	42	14
ass+	35	31	-193	56	8	9	0	0	355	191	55
avo+	16	49	-1020	444	98	-664	188	77	-365	57	26
ban+	14	51	-1185	503	116	524	98	42	-285	29	14
chi+	36	37	-272	96	16	293	111	35	-168	36	13
ing+	42	30	-177	59	8	348	228	58	95	17	5
ins+	17	40	-491	140	26	484	136	46	357	74	28
jou+	43	31	159	47	7	-127	30	8	145	39	11
méd+	14	40	-329	49	9	-458	95	32	258	30	11
min+	2	49	1220	78	18	-144	1	0	-3522	654	305
pho+	80	18	248	354	30	-35	7	1	68	27	5
pil+	86	21	267	394	37	33	6	1	-83	38	7
psy+	16	50	-1021	431	98	-863	308	131	-159	10	5
act-	25	37	-255	59	10	-2	0	0	29	1	0
agr-	44	32	-412	307	45	-103	19	5	53	5	2
ass-	28	38	219	47	8	339	114	37	-477	225	81
avo-	58	23	194	126	13	109	39	8	149	74	16
ban-	35	34	61	5	1	-628	539	157	171	40	13
chi-	27	34	408	179	28	-62	4	1	296	94	30
ing-	14	44	653	177	35	-680	191	71	-641	170	71
ins-	31	33	133	22	3	-558	396	109	71	6	2
jou-	17	42	83	4	1	300	50	18	-1048	604	243
méd-	27	33	320	114	17	113	14	4	202	45	14
min-	103	13	-118	150	9	162	281	31	57	35	4
pho-	6	54	-1842	485	120	-878	110	51	-469	31	16
pil-	16	55	-1193	575	143	507	104	48	-122	6	3
psy-	63	23	165	102	11	175	115	22	55	11	2

Représentation des individus-colonnes dans le premier plan factoriel



- 1) Les auteurs utilisent ici une AFC. A quel type de tableau cette méthode s'applique-t-elle généralement ? Pour les données étudiées ici, quelle interprétation peut-on donner des sommes des lignes ? des sommes des colonnes ? Que représente le profil-ligne moyen ?
- 2) Quelles sont les trois professions les plus attractives ? Les moins attractives ? Sur quels éléments des tableaux vous appuyez-vous pour apporter ces réponses ?
- 3) Comment peut-on qualifier la décroissance des valeurs propres ? Sur quel argument peut-on se fonder pour limiter l'étude aux 4 premiers axes factoriels ?
- 4) Etude de la première variable factorielle  
Quels sont les individus-colonnes dont la contribution est supérieure à la moyenne ? Pour chacun d'eux, préciser le signe de la coordonnée correspondante.
- 5) Mener une étude analogue pour la deuxième variable.
- 6) Certaines modalités sont particulièrement mal représentées dans le premier plan factoriel. Indiquer les trois modalités les plus mal représentées dans ce plan.
- 7) Faire une synthèse des deux études précédentes en décrivant les résultats obtenus dans le premier plan factoriel. A partir de l'étude, les auteurs distinguent trois groupes de professions : les professions préférées par les sujets, des professions peu attrayantes et des professions moyennement appréciées. Proposer deux ou trois représentants typiques de chacun des trois groupes.
- 8) Quelle est la modalité qui a la plus grande contribution à la formation du 3ème axe ? Quel est son poids ? Comment peut-on expliquer qu'une modalité ayant un poids aussi faible ait une contribution aussi importante ?
- 9) Toutes les modalités colonnes ont-elles été représentées sur le graphique ? Quels critères ont été utilisés pour éliminer certaines d'entre elles ?
- 10) Un chercheur propose de recoder les données recueillies comme suit : pour chaque profession, on introduit une troisième modalité, "profession non citée". On peut alors considérer que l'on a observé 14 variables nominales, possédant chacune 3 modalités, sur les 90 sujets. Quelle serait alors la méthode d'analyse adaptée à cet ensemble de données ?

### **Quelques commentaires et indications de réponses**

- 1) L'AFC s'applique généralement à un tableau de contingence. Le tableau décrit ici un peu différent, puisque chaque sujet fait 6 choix (3 choix positifs et 3 choix négatifs). Sur chaque ligne, la somme est égale au nombre de choix demandés au sujet, c'est à dire 6 (sauf cas particuliers de réponses incomplètes). La somme sur une colonne représente le nombre de fois où un couple (profession, connotation) a été choisi par les sujets de l'échantillon. Le profil ligne moyen représente les fréquences de choix des couples (profession, connotation) par l'échantillon proposé.
- 2) On se reporte ici à la colonne "poids" du tableau des résultats numériques. Les plus appréciées : pilote, photographe, acteur ; les moins appréciées : mineur, psychologue, avocat.
- 3) Le tableau des valeurs propres n'indique pas la totalité des résultats : seules les premières valeurs propres, qui expliquent environ 80% de l'inertie, sont indiquées. La décroissance des valeurs propres est assez lente. Il semble que les auteurs aient ici choisi de s'arrêter au 4ème axe, car les 4 premiers axes expliquent plus de la moitié de l'inertie.
- 4) On examine ici la colonne "CTR" du facteur 1 et le signe de la valeur indiquée dans la colonne "Coord 1". La contribution moyenne est de  $100/28=3,6\%$ . On relève donc les couples (profession, connotation) pour lesquels la colonne CTR contient une valeur supérieure à 36 :

-	+
ban+	agr+
pil-	pil+
pho-	
psy+	
avo+	
agr-	

5) De même pour le second axe :

-	+
ban-	ing+
psy+	pil-
ins-	ins+
avo+	ban+
ing-	ass-
pho-	

6) Les qualités de représentation dans le premier plan factoriel sont obtenues en additionnant les valeurs indiquées dans les colonnes "COR" relatives aux facteurs 1 et 2. Les trois qualités les plus mauvaises sont jou- (54), ass+ (56), act- (59).

7) Le premier axe oppose les professions de pilote automobile et d'agriculteur aux professions libérales (banquier, avocat, psychologue). Le second axe oppose les professions d'ingénieur, instituteur, banquier (partie positive de l'axe) à celles d'avocat et de psychologue (partie négative). La partie négative de l'axe correspond également au rejet des professions d'ingénieur, instituteur, banquier.

8) La modalité min+ représente 30% de l'inertie de cet axe. Cette modalité a effectivement un poids faible (2 millièmes) : elle a vraisemblablement été choisie une seule fois sur les 540 réponses obtenues dans les 90 questionnaires. La réponse où elle apparaît se trouve donc très éloignée du profil moyen, et l'axe que nous étudions est le 3<sup>e</sup>, les deux premiers axes nous mettent en évidence des effets plus généraux.

9) Le graphique ne comporte que 21 points alors qu'il existe 28 modalités. Certaines modalités (ass+, par exemple) ont été supprimées. Il s'agit de modalités proches de l'origine, dont les étiquettes auraient recouvert des étiquettes déjà présentes. Le but des auteurs était visiblement d'améliorer la lisibilité du graphique.

10) On réalise alors l'étude de 3 variables nominales sur un ensemble de sujets, ce qui peut être réalisé par une analyse des correspondances multiples (ACM).

N.B. Pour une analyse plus fine des résultats de cette AFC cf. Doise W., Clémence, A., Lorenzi-Cioldi F., Représentations sociales et analyses de données, chapitre 3.