ENQUETE SUR L'UTILISATION DES LOGICIELS DE STATISTIQUE - ASU 1992 -

M. BECUE[1], M.K. DIALLO[2], D. GRANGE[3], L. HAEUSLER[4], Y. LECHEVALLIER[5], Y. MAJJAD[3], M. RINGENBACH[3], V. PEREZ[4], F. SERMIER[6]

[1] Departament d'Estadistica i investigacio Operativa
Facultat d'Informatica de Barcelona
Universitat Politècnica de Catalunya
c/Pau Gargallo, 5 08028 Barcelona

[2] Université de Conakry
BP 1147 Conakry
Guinée

[3] Unité Statistique
Centre de Calcul du CNRS
23, rue du Loess
67200 Strasbourg

[4] CISIA
1, avenue Herbillon
94160 Saint Mandé

[5] INRIA
Rocquencourt
78153 Le Chesnay

[6] 9, rue Sophie Germain
75014 Paris

Résumé :
En Mai 1992, le groupe "Logiciels et Statistique" de l'ASU faisait une enquête pour connaître l'opinion des utilisateurs de logiciels de statistique. 416 questionnaires ont été obtenus et traités par un groupe de travail. Cet article présente les résultats.

Mots-clés :
enquête, logiciels de statistique, question ouverte, thémascope.

1. Contexte de l'enquête

Dès sa création, en Janvier 1992, le groupe "Logiciels et Statistique" de l'ASU s'est intéressé à la comparaison de logiciels de statistique. Les travaux sont assez rares dans ce domaine ou commencent à dater [1], [2], [3], [4]. Plus nombreuses sont les publications commerciales mais celles-ci s'intéressent plus à l'aspect convivial et informatique du produit qu'à son contenu statistique [5], [6], [7], [8], [9], [10], [11], [12], [13].
Au printemps 1992, le groupe "Logiciels et Statistique" décidait d'effectuer une enquête auprès des statisticiens et utilisateurs de logiciels statistiques, afin de connaître, d'une part, les logiciels utilisés, et d'autre part, l'opinion des utilisateurs sur ces logiciels et l'usage qui en est fait.

Le groupe "Constitution du Questionnaire" (J.C. Dauphin, D. Grangé, L. Lebart, A. Morin, A. Morineau et F. Sermier) se mettait alors en place pour constituer le questionnaire. N'ayant à sa disposition aucune base de sondage et ne disposant que de très faibles moyens financiers, le groupe a utilisé toutes les bonnes volontés pour diffuser le plus largement possible ce questionnaire. Une distribution a d'abord été effectuée aux XXIV èmes journées de l'ASU à Bruxelles en Mai 1992, puis au Congrès Distancia à Rennes en Juin 1992. Des sociétés ou clubs d'utilisateurs de logiciels (Addad, CISIA, Statgraphics, S-Plus et de SAS/STAT) ont accepté de faire parvenir ce questionnaire à leurs clients et nous les en remercions. Ce mode de diffusion est, bien évidemment, le point faible de cette étude et entraînera un biais d'enquête que nous retrouverons dans les résultats à tel point que certaines sociétés ont cru bon de nous en avertir, en particulier la société Eole, dès septembre 1992.

416 questionnaires étaient parvenus au 15 septembre 1992. Tout membre de l'ASU pouvait disposer des données et participer au traitement de l'enquête dans le but exclusif de collaborer à un travail de groupe. Les membres du groupe ont décidé, dans un premier temps, de travailler sur les données selon leur intérêt. Ils étaient libres du choix des logiciels et du matériel. Ce mode de travail s'est révélé très motivant et donc fructueux. En effet, les membres du groupe, provenant de divers horizons, n'avaient pas le même intérêt pour ces données et ont eu des regards complémentaires. Par ailleurs ils ont utilisé des logiciels et des environnements différents, ce qui a permis d'explorer les données sous des angles variés [14], [15] [16], [17].

F. Sermier s'est essentiellement intéressé au panorama des logiciels cités et aux ordinateurs utilisés. Y. Lechevallier, et M.K. Diallo souhaitaient voir par qui étaient utilisées les

techniques statistiques citées. L. Haeusler et V. Perez décrivaient les types d'utilisation des logiciels. M. Becue essayait de cerner ce qu'est un "bon logiciel de statistique", en exploitant la question ouverte. D. Grangé et M. Ringenbach exploraient la fiche signalétique de l'enquêté. Y. Majjad s'intéressait aux opinions des utilisateurs sur leurs logiciels.

2. Les répondants

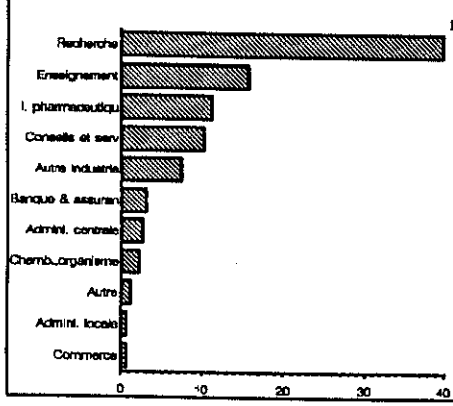
Les 416 personnes qui ont répondu au questionnaire, proviennent essentiellement : d'organismes de recherche (44,5 %), d'enseignement (16 %), de l'industrie pharmaceutique (11 %) ou des conseils et services (10 %). En conséquence 62 % travaillent dans des organismes de plus de 500 personnes.

Par contre, parmi ces 416 répondants, on trouve peu de chercheurs en statistique (89). Mais ils sont plus nombreux (201) à utiliser l'outil statistique pour faire de la recherche dans un autre domaine. 201 également ont une activité de Conseils et Services alors que les enseignants sont assez rares dans cet échantillon puisque 42 % des répondants ont une activité d'enseignement occasionnelle et que pour 29 % elle est inexistante ou non renseignée. 28 personnes ne se reconnaissent pas dans les activités proposées. Malheureusement les informations renseignées en clair ne permettent pas de cerner ces activités mais indiquent plutôt des domaines d'activités (électronique, toxicologie,.....). On peut penser que ces personnes ont peut-être des activités de production plutôt que de recherche, ou conseils.

Parmi les principaux domaines d'activité, la rubrique "recherche ou enseignement scientifique" vient nettement en tête 34 %, suivie de médecine, pharmacie, biologie 14 % puis de l'agro-alimentaire 9 %. 35 ne sont pas reconnus dans les 10 domaines proposés. Parmi ceux-ci on trouve l'économie, les sciences politiques, la poste, le tourisme, l'agriculture, l'administration, les services SSII, l'agronomie et l'informatique.
216 répondants ont un diplôme en statistique qui est essentiellement Bac + 5 suivi par l'ISUP. Les autres formations statistiques, Bac + 4, ENSAE SEA, ENSAE CGSA, IUT se répartissent entre 5 et 7 %.

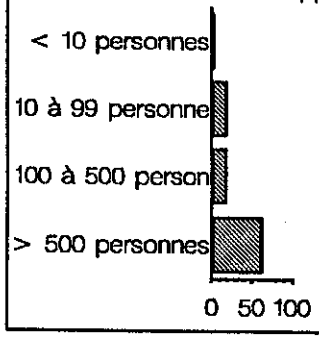
Parmi les personnes qui n'ont pas de diplôme en statistique, 64 % ont une formation en statistique inférieure à 2 ans. Mais l'ensemble des répondants utilise les méthodes statistiques depuis au moins 4 ans et en moyenne depuis 9 ans. 59 % des répondants ont une autre formation ou diplôme

Q1 – Quel est le secteur principal d'activité de l'organisme ou vous travaillez ?



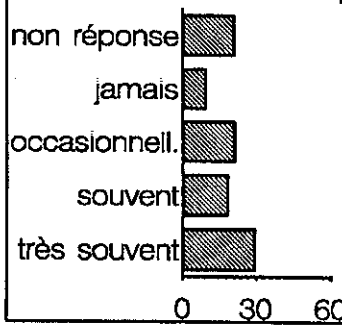
	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
Recherche	185	185	44.47	44.47
Enseignement	66	251	15.87	60.34
I. pharmaceutiqu	47	298	11.30	71.63
Conseils et serv	43	341	10.34	81.97
Autre industrie	31	372	7.45	89.42
Banque & assuran	13	385	3.13	92.55
Admini. centrale	11	396	2.64	95.19
Chamb_organisme	9	405	2.16	97.36
Autre	5	410	1.20	98.56
Admini. locale	3	413	0.72	99.28
Commerce	3	416	0.72	100.00

Q2 – Quelle est la taille de l'organisme qui vous emploie ?



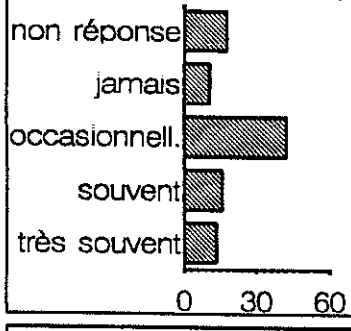
	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
< 10 personnes	12	12	2.93	2.93
10 à 99 personne	74	86	18.09	21.03
100 à 500 person	69	155	16.87	37.90
> 500 personnes	254	409	62.10	100.00

Q3 – Profil d'activité : Conseil et études



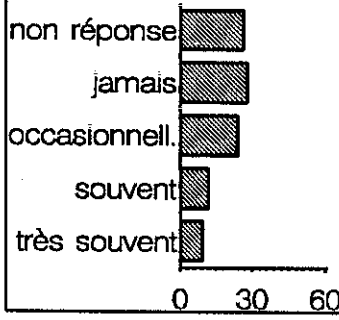
	FREQ.	CUM. FREQ.	PCT.
non réponse	88	88	21.10
jamais	39	127	9.35
occasionnell.	89	216	21.34
souvent	77	293	18.47
très souvent	124	417	29.74

Q3 – Profil d'activité : Enseignement et formation



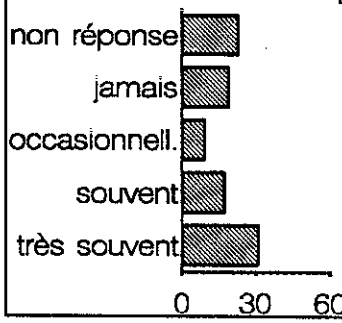
	FREQ.	CUM. FREQ.	PCT.
non réponse	74	74	17.75
jamais	45	119	10.79
occasionnell.	176	295	42.21
souvent	66	361	15.83
très souvent	56	417	13.43

Q3 – Profil d'activité : Recherche en statistiques



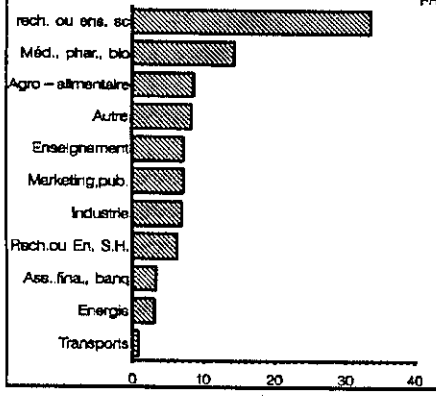
	FREQ.	CUM. FREQ.	PCT.
non réponse	110	110	26.38
jamais	117	227	28.06
occasionnell.	101	328	24.22
souvent	50	378	11.99
très souvent	39	417	9.35

Q3 – Profil d'activité : Recherche dans un autre domaine



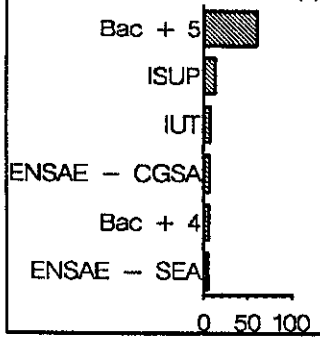
	FREQ.	CUM. FREQ.	PCT.
non réponse	96	96	23.02
jamais	80	176	19.18
occasionnell.	39	215	9.35
souvent	73	288	17.51
très souvent	129	417	30.94

Q4 – Quel est votre principal domaine d'activité ?



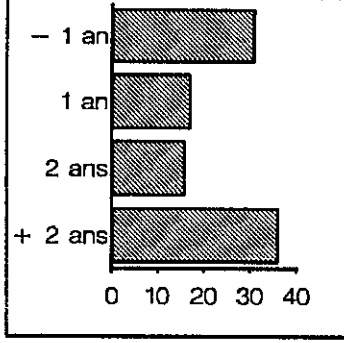
	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
rech. ou ens. sc	140	140	33.65	33.65
Méd., phar., bio	60	200	14.42	48.09
Agro – alimentaire	36	236	8.65	56.73
Autre	35	271	8.41	65.14
Enseignement	30	301	7.21	72.36
Marketing,pub.	30	331	7.21	79.57
Industrie	29	360	6.97	86.54
Rech.ou En. S.H.	26	386	6.25	92.79
Ass.fina., banq	14	400	3.37	96.15
Energie	13	413	3.13	99.28
Transports	3	416	0.72	100.00

Q5_1 – Avez–vous un diplôme spécifique en statistique ?



	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
Bac + 5	133	133	61.29	61.29
ISUP	29	162	13.36	74.65
IUT	17	179	7.83	82.49
ENSAE – CGSA	15	194	6.91	89.40
Bac + 4	12	206	5.53	94.93
ENSAE – SEA	11	217	5.07	100.00

Q5_2 – Durée de la formation title2 si pas de diplôme axis2 label = none

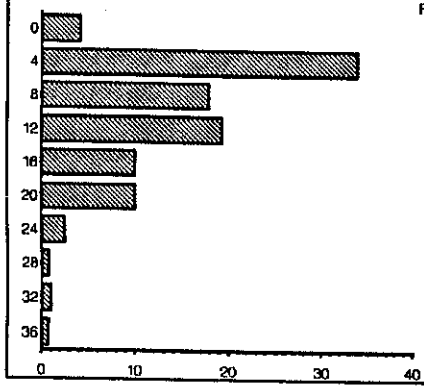


	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
– 1 an	62	62	31	31
1 an	34	96	17	48
2 ans	32	128	16	64
+ 2 ans	72	200	36	100

Bien que cette enquête ait été distribuée au Congrès de l'ASU à Bruxelles, qui avait environ 500 participants, seulement 119 répondants (29 %) sont membres de l'ASU et 111 ont participé à un congrès de l'ASU au cours des 2 dernières années. 62 % des répondants ne sont membres d'aucune société savante. Les participants aux congrès de l'ASU, statisticiens théoriciens, ne sont peut-être pas des utilisateurs de logiciels. Ce questionnaire a sans doute mieux touché les utilisateurs de logiciels grâce à la diffusion faite par les sociétés de logiciels, ce qui expliquerait le fort pourcentage de non statisticiens dans cette enquête.

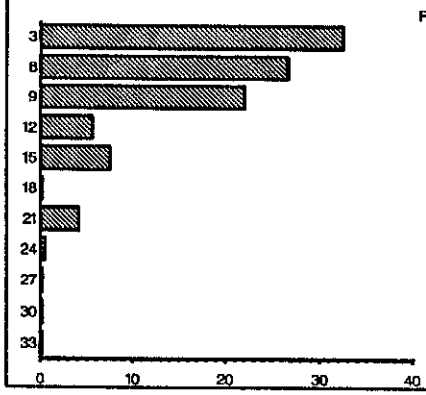
Par contre, 149 enquêtés participent à un club d'utilisateurs de logiciels. La procédure DEMOD de SPAD.N (15) nous permet d'en donner le profil suivant : ils utilisent SAS prioritairement, sont membres de l'ASU, de l'IIS, participent aux Congrès de l'ASU. Ils appartiennent à l'industrie pharmaceutique, médecine, pharmacie, biologie, banque, assurance, finance. Ils sont statisticiens de formation, essentiellement Bac + 5. Ils préfèrent travailler par langage de commande, ne travaillent jamais sur Mac et très souvent sur site central. Ils utilisent souvent le modèle log-linéaire, très souvent les statistiques descriptives, les analyses factorielles et les classifications et occasionnellement font du contrôle de qualité. Ce groupe compte 45 % de femmes alors que celles-ci ne représentent que 30 % des enquêtés.

Q7 — Nombre d'années d'utilisation des méthodes statistiques dans votre activité professionnelle ?



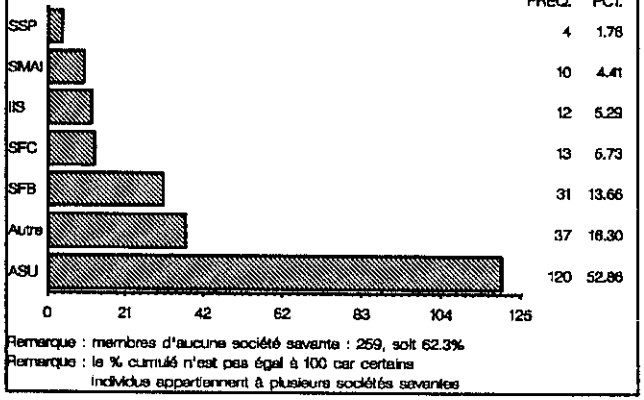
	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
0	17	17	4.15	4.15
4	139	156	33.90	38.05
8	73	229	17.80	55.85
12	79	308	19.27	75.12
16	41	349	10.00	85.12
20	41	390	10.00	95.12
24	10	400	2.44	97.56
28	3	403	0.73	98.29
32	4	407	0.98	99.27
36	3	410	0.73	100.00

Q8 — Nombre d'années d'utilisation d'au moins un logiciel de statistique ?



	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
3	133	133	32.52	32.52
6	109	242	26.65	59.17
9	90	332	22.00	81.17
12	23	355	5.62	86.80
15	31	386	7.58	94.38
18	1	387	0.24	94.62
21	17	404	4.16	98.78
24	2	406	0.49	99.27
27	1	407	0.24	99.51
30	1	408	0.24	99.76
33	1	409	0.24	100.00

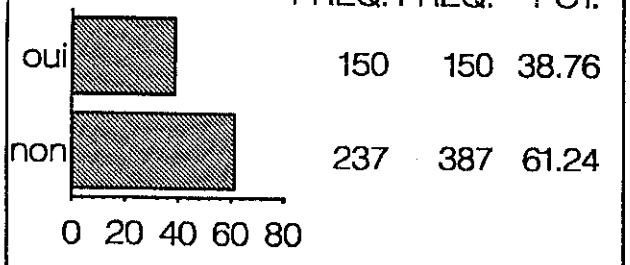
Q9 — Etes-vous membre d'une société savante de statistique ?



	FREQ.	PCT.
SSP	4	1.78
SIMAI	10	4.41
IIS	12	5.29
SFC	13	6.73
SFB	31	13.66
Autre	37	16.30
ASU	120	52.86

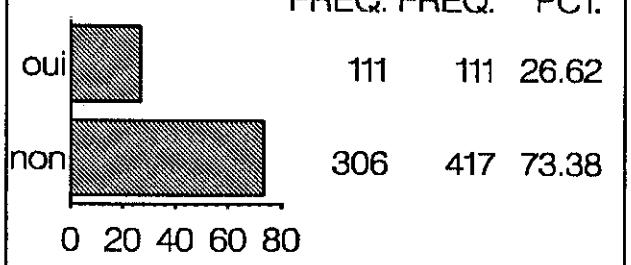
Remarque : membres d'aucune société savante : 259, soit 62.3%
Remarque : le % cumulé n'est pas égal à 100 car certains individus appartiennent à plusieurs sociétés savantes

Q10 — Membre d'un club d'utilisateurs de logiciel statistique



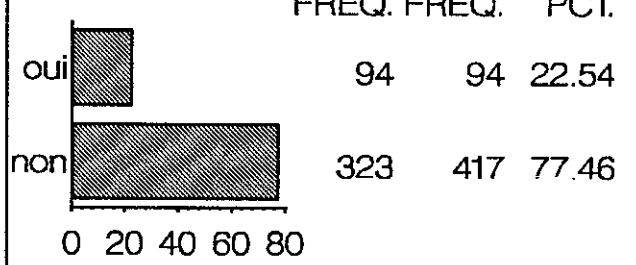
	FREQ.	CUM. FREQ.	PCT.
oui	150	150	38.76
non	237	387	61.24

Q11_1 — Participation à un congrès de l'ASU les 2 dernières années

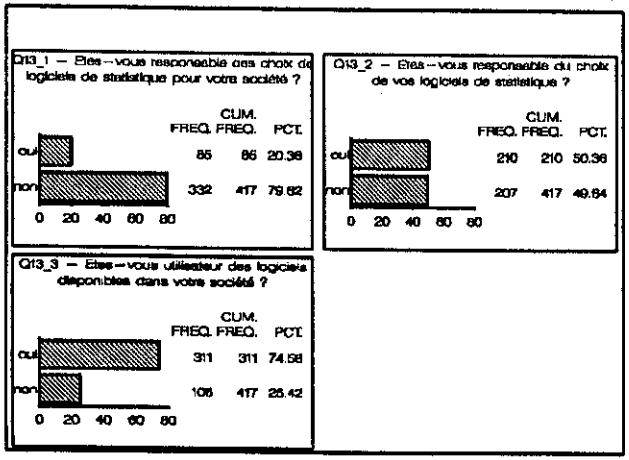


	FREQ.	CUM. FREQ.	PCT.
oui	111	111	26.62
non	306	417	73.38

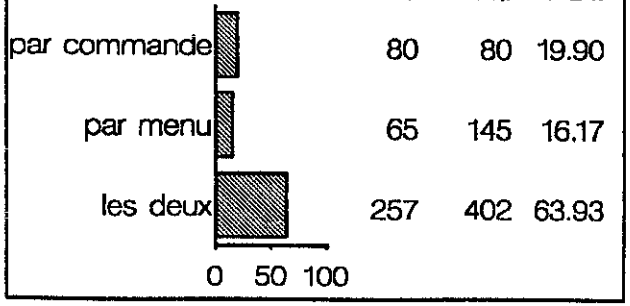
Q11_2 — Participation à un autre congrès statistique que l'ASU



	FREQ.	CUM. FREQ.	PCT.
oui	94	94	22.54
non	323	417	77.46

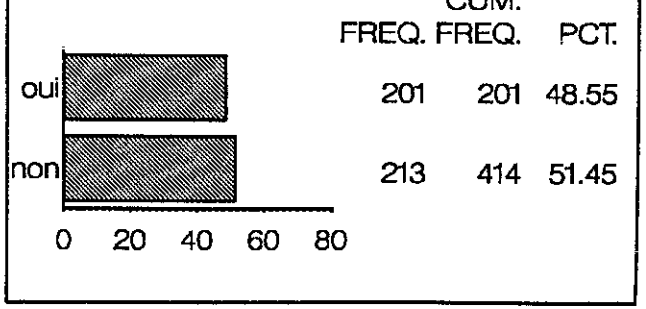


Q14 — Préférez-vous utiliser vos logiciels ?

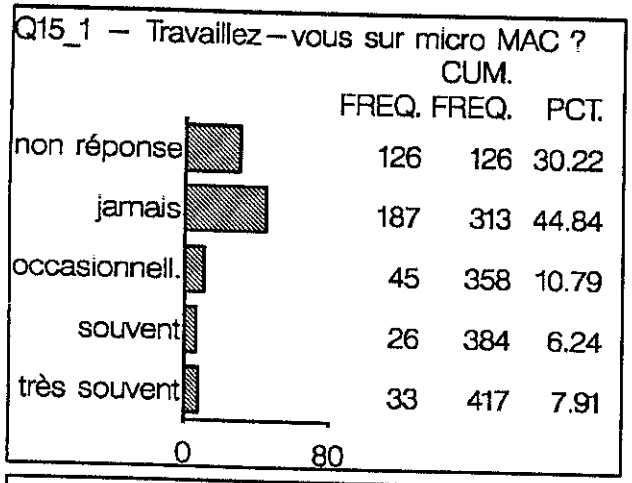


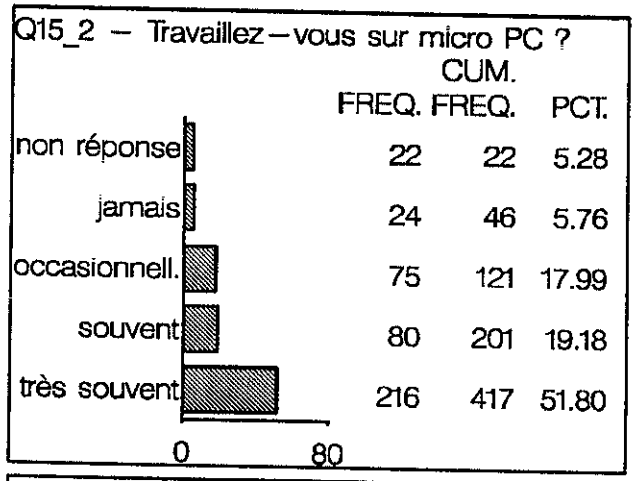
	FREQ.	CUM. FREQ.	PCT.
par commande	80	80	19.90
par menu	65	145	16.17
les deux	257	402	63.93

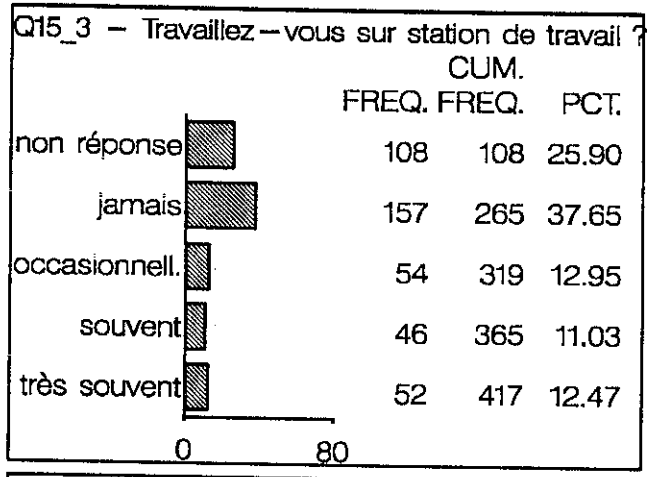
Q16 — Avez-vous un ordinateur personnel à domicile ou portable ?

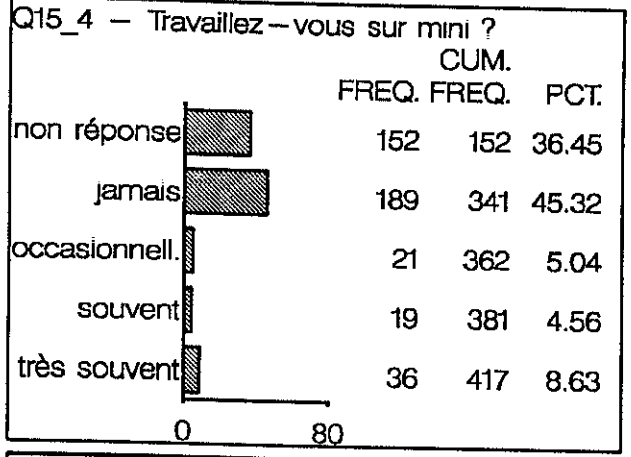


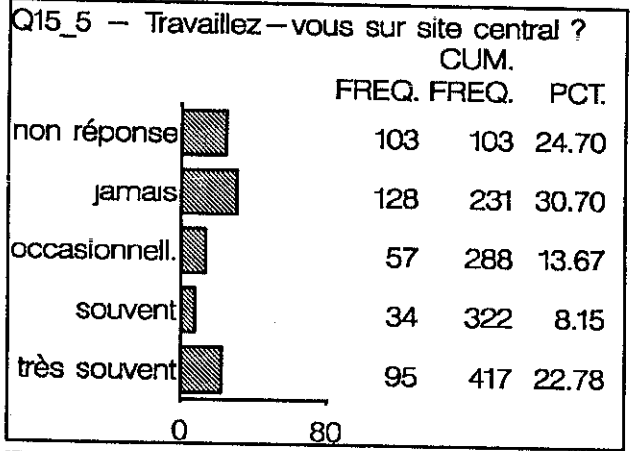
	FREQ.	CUM. FREQ.	PCT.
oui	201	201	48.55
non	213	414	51.45

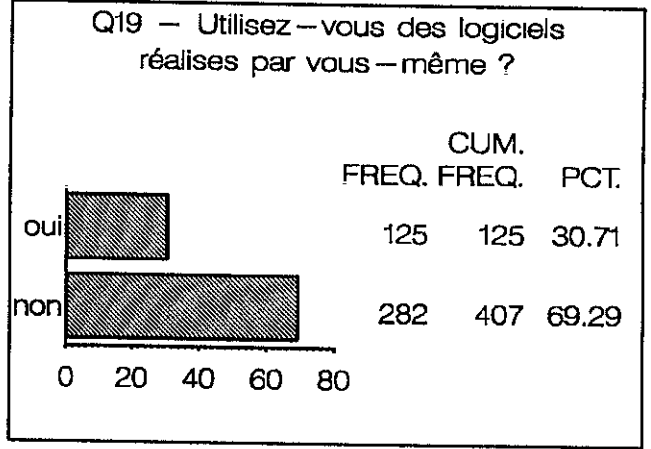

Q15_1 — Travaillez—vous sur micro MAC ?
FREQ. CUM. FREQ. PCT.
non réponse 126 126 30.22
jamais 187 313 44.84
occasionnell. 45 358 10.79
souvent 26 384 6.24
très souvent 33 417 7.91
0 80

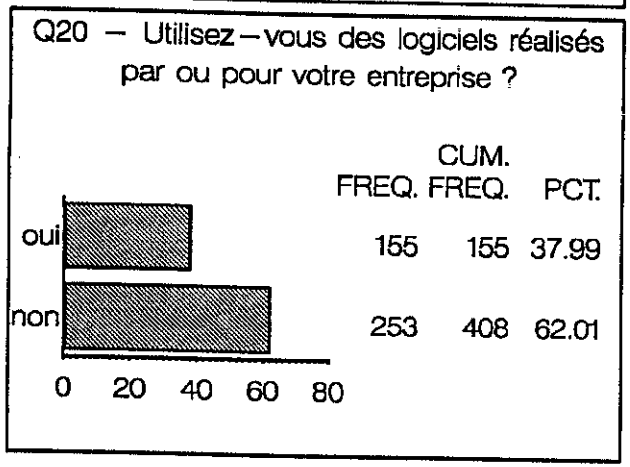

Q15_2 — Travaillez—vous sur micro PC ?
FREQ. CUM. FREQ. PCT.
non réponse 22 22 5.28
jamais 24 46 5.76
occasionnell. 75 121 17.99
souvent 80 201 19.18
très souvent 216 417 51.80
0 80

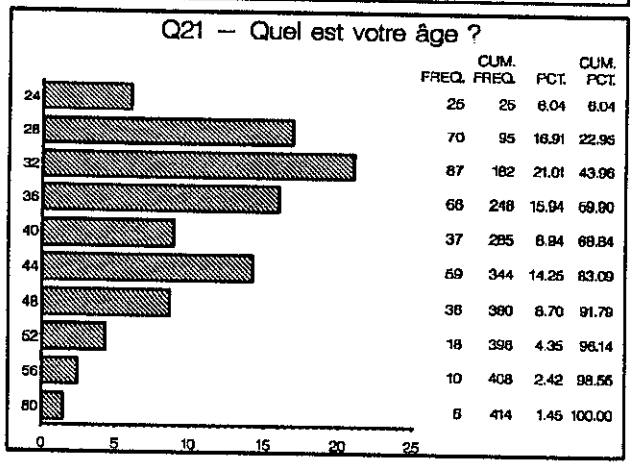

Q15_3 — Travaillez—vous sur station de travail ?
FREQ. CUM. FREQ. PCT.
non réponse 108 108 25.90
jamais 157 265 37.65
occasionnell. 54 319 12.95
souvent 46 365 11.03
très souvent 52 417 12.47
0 80

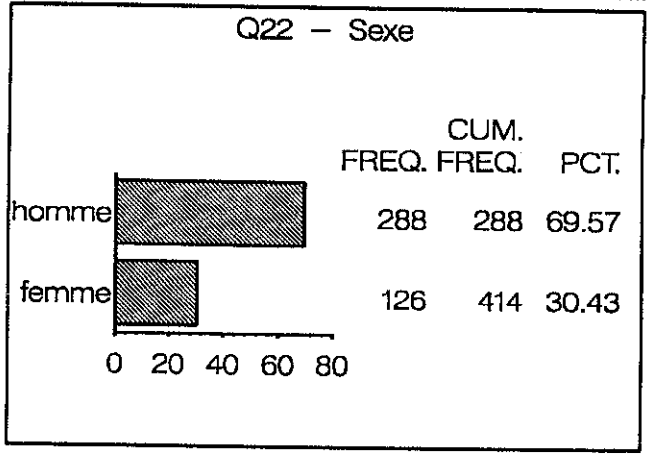

Q15_4 — Travaillez—vous sur mini ?
FREQ. CUM. FREQ. PCT.
non réponse 152 152 36.45
jamais 189 341 45.32
occasionnell. 21 362 5.04
souvent 19 381 4.56
très souvent 36 417 8.63
0 80


Q15_5 — Travaillez—vous sur site central ?
FREQ. CUM. FREQ. PCT.
non réponse 103 103 24.70
jamais 128 231 30.70
occasionnell. 57 288 13.67
souvent 34 322 8.15
très souvent 95 417 22.78
0 80


Q19 — Utilisez—vous des logiciels réalises par vous—même ?
FREQ. CUM. FREQ. PCT.
oui 125 125 30.71
non 282 407 69.29
0 20 40 60 80


Q20 — Utilisez—vous des logiciels réalisés par ou pour votre entreprise ?
FREQ. CUM. FREQ. PCT.
oui 155 155 37.99
non 253 408 62.01
0 20 40 60 80


Q21 — Quel est votre âge ?
FREQ. CUM. FREQ. PCT. CUM. PCT.
25 25 6.04 6.04
70 95 16.91 22.95
87 182 21.01 43.96
66 248 15.94 59.90
37 285 8.94 68.84
59 344 14.25 83.09
36 380 8.70 91.79
18 396 4.35 96.14
10 408 2.42 98.55
6 414 1.45 100.00
0 5 10 15 20 25


Q22 — Sexe
FREQ. CUM. FREQ. PCT.
homme 288 288 69.57
femme 126 414 30.43
0 20 40 60 80

Une très large majorité des répondants travaille dans un environnement micro PC. Moins nombreux sont ceux qui travaillent sur site central. Puis viennent les stations de travail et enfin peu nombreux sont ceux qui travaillent sur Mini ou sur Mac. 31 % utilisent des logiciels qu'ils ont réalisés et 38 % des logiciels réalisés par ou pour leur entreprise.

La moitié des répondants dispose d'un ordinateur personnel.

Enfin 70 % sont des hommes. Ils peuvent être décrits de la façon suivante : de façon significative, on trouve plus d'hommes que dans l'échantillon global dans les rubriques : préfèrent travailler par menu, ne sont pas membres de l'ASU, possèdent un ordinateur personnel, ne font pas partie d'un club d'utilisateurs, utilisent souvent un Mac, font occasionnellement des statistiques non paramétriques et sont responsables du choix de leurs logiciels.

Alors que les femmes se trouvent en proportion importante dans d'autres rubriques et peuvent être ainsi décrites : elles sont membres de l'ASU, ne possèdent pas d'ordinateur personnel à domicile, proviennent de l'industrie pharmaceutique, utilisent très souvent les tests non paramétriques, ont une formation Bac + 4 et participent à un club d'utilisateurs.

3. Les logiciels

Dans les deux parties du questionnaire des questions avaient trait aux logiciels utilisés par le répondant :

- dans la première partie, la question Q17 demandait de donner les
"noms des logiciels commercialisés que vous utilisez, et éventuellement le numéro de version, en indiquant l'ordre de fréquence d'utilisation (1 = le plus fréquent)"
La question était ouverte et permettait de citer tout logiciel commercial utilisé par le répondant, accompagné d'un rang d'utilisation.

- de la même manière, la question Q18 demandait les
"noms des logiciels dont vous disposez et que vous n'utilisez jamais et pourquoi ?"
Six possibilités de réponses étaient prévues. Nous n'avons pas exploité l'information éventuelle fournie sur les raisons de la non-utilisation.

- la deuxième partie, quant à elle, demandait à l'utilisateur de décrire le mode d'utilisation et de donner son opinion sur les 3 logiciels commerciaux qu'il utilise le plus souvent. Dans cette partie, le répondant indique sur quelle machine il utilise le logiciel.

Nous obtenons donc plusieurs bases de dénombrement des réponses :

dans la première partie, sur les **416 questionnaires** reçus, il a été cité

152 **logiciels** différents à la question Q17 ou Q18, soit au total :

1123 citations en Q17 d'un logiciel utilisé (soit 2,7 logiciels utilisés par répondant)
et **229** citations en Q18 d'un logiciel inutilisé (environ 0,6 logiciels inutilisés en moyenne):

dans la deuxième partie, les répondants ont rempli au total :

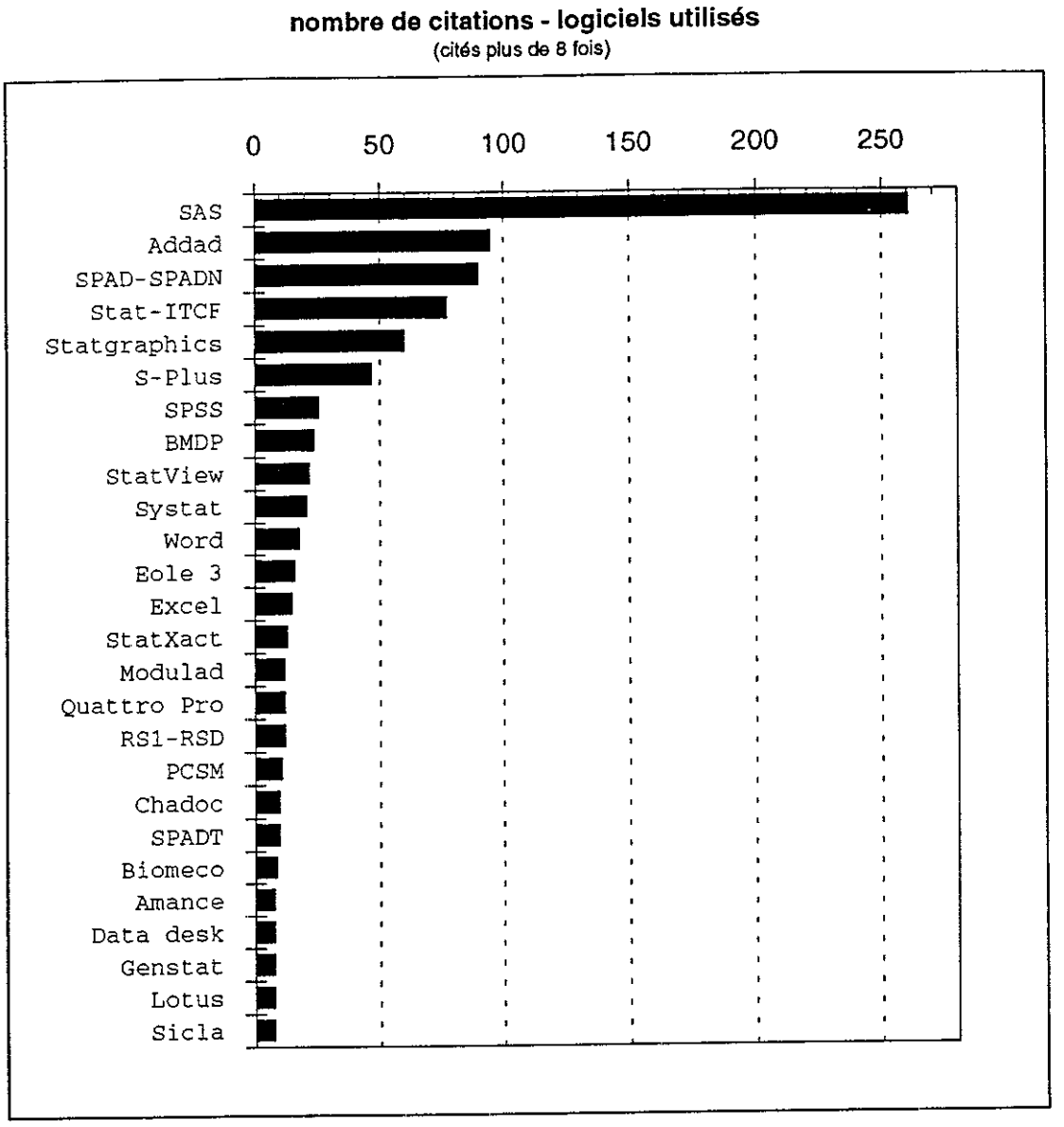
888 descriptions de logiciels (2,1 logiciels décrits par répondant), ce qui, en tenant compte des différents matériels donnait

1059 associations de logiciels à un type de machine (Mac, PC, station, Mini ou mainframe).

3.1 logiciels utilisés

Sur 1123 citations de logiciels utilisés, SAS (toutes versions et tous systèmes confondus) domine largement avec 261 citations, soit 23,2% du total des citations et 62,7% des questionnaires. Il est suivi, loin derrière,avec un peu moins de 100 citations chacun, par Addad et SPAD.N. Après les trois logiciels suivants : Stat-ITCF, Statgraphics et S-Plus, on tombe très rapidement à moins de 30 citations, soit moins de 2,5% du total des citations et moins de 6,5% des répondants.

On voit que la distribution est plutôt concentrée, avec une courbe de concentration du type 20-80 (20% des logiciels cumulant 80% du total des citations). *(voir détail des comptages par logiciel en annexe 1)*.



Il est intéressant de noter que le caractère ouvert de la question Q17 a permis l'apparition de logiciels que nous n'aurions sans doute pas proposés a priori dans une liste de logiciels statistiques : s'il n'est pas trop surprenant de voir apparaître les tableurs, grapheurs ou gestionnaires de base de données, il est plus étonnant de voir des utilitaires comme PC-Shell ou PC-Tools (cités 4 fois au total) ou même Word cité 18 fois ce qui le met en 11 ème position des logiciels statistiques (il est vrai qu'il dispose d'un module grapheur).

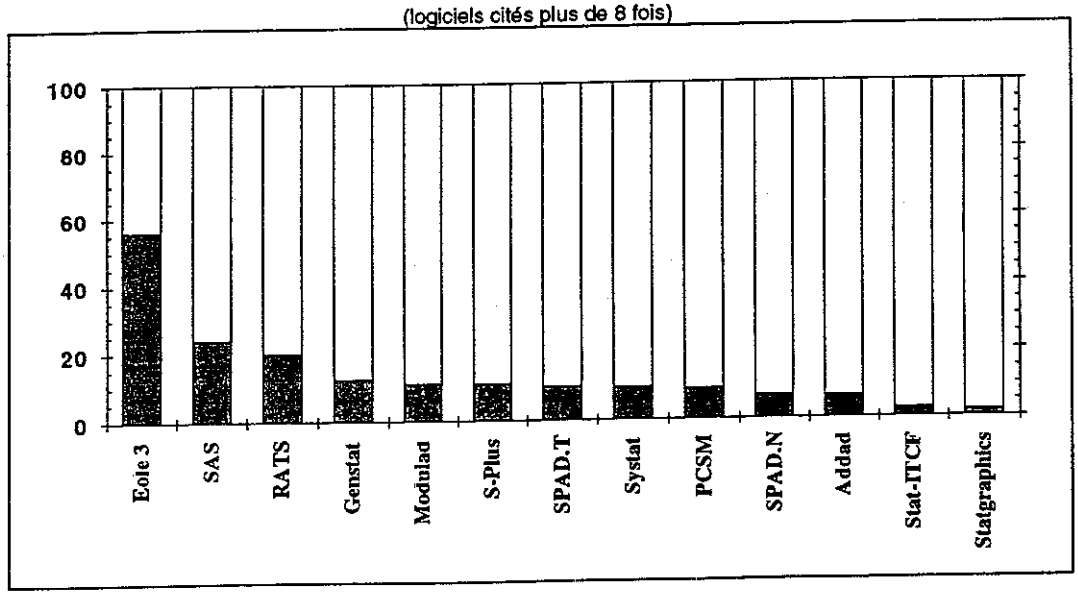
autres logiciels utilisés pour les statistiques

	logiciel	cités en partie 1	étudiés en partie 2
tableurs (total 41)	Excel	15	10
	Quattro Pro	12	12
	Lotus	8	3
	Multiplan	4	3
	Symphony	1	0
	WingZ	1	0
intégrés	Open Access + Works	2	1
grapheurs (total 11)	Harvard Graphics	4	0
	Chart	3	3
	Cricket Graph	3	1
	Décisionnel Graphique	1	1
SGBD (total 10)	dBase	5	1
	Nomad	2	2
	Fox Pro	1	1
	Oracle	1	1
	Paradox	1	0
divers (total 24)	Norton, PC-Tools	4	1
	Windows	2	1
	Word	18	2

Sur les 416 répondants, 99 (soit 23,8%) ne citent qu'un seul logiciel commercial utilisé. Il s'agit pour l'essentiel (62 réponses) d'utilisateurs de SAS. En rapportant ce chiffre au nombre total de citations de SAS, on obtient 23,8% des utilisateurs de SAS qui n'utilisent pas d'autre produit commercial. Ce pourcentage n'est dépassé que par les utilisateurs d'Eole : 9 sur 16 n'utilisent que ce produit (notons que nous avons reçu -tardivement- du distributeur du produit une enveloppe contenant 14 questionnaires ; il nous semble que ceci peut expliquer la position parfois singulière d'Eole dans plusieurs de nos résultats).

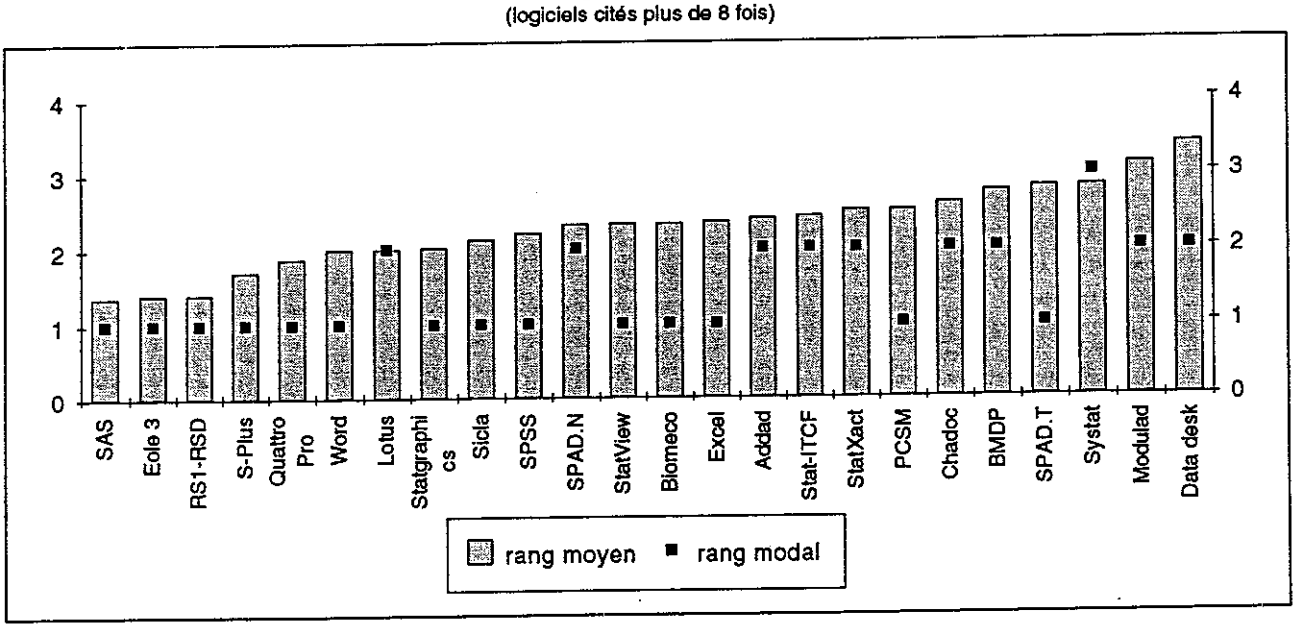
taux d'utilisation monoproduit

(logiciels cités plus de 8 fois)



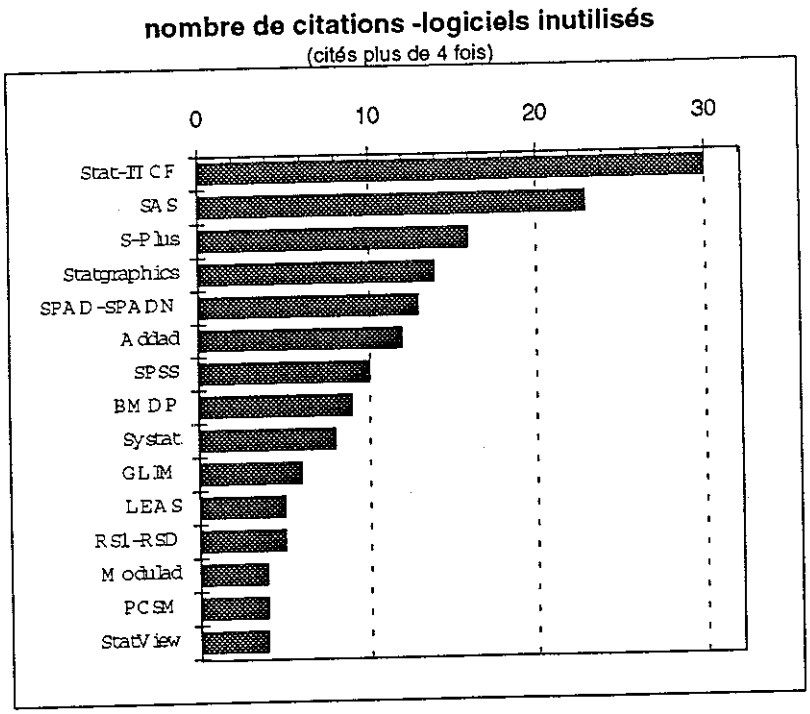
Enfin, lorsqu'on étudie les citations des logiciels utilisés en relation avec leur rang d'utilisation, on retrouve de nouveau SAS en tête avec un rang moyen d'utilisation de 1,36, suivi d'Eole qui là aussi se singularise avec un rang moyen d'utilisation de 1,40. Parmi les autres logiciels les plus fréquemment cités, on retrouve SPAD.N et Addad avec des rangs moyens très voisins de 2,32 et 2,40. Le rang d'utilisation le plus fréquemment cité pour ces deux logiciels est 2. Ils apparaissent donc nettement comme des logiciels "de complément", par rapport à une utilisation intensive de SAS.

classement des logiciels par rang d'utilisation
(logiciels cités plus de 8 fois)



3.2 logiciels inutilisés

30 répondants (soit 7,2%) déclarent disposer de Stat-ITCF et ne pas l'utiliser. Ce nombre est élevé mais peu étonnant compte tenu du mode de diffusion et du coût du logiciel. De même 23 répondants (5,5%) disposent de SAS et ne l'utilisent pas. Dans ce cas les mêmes considérations de coût nous amènent à penser que le logiciel est disponible sur le site central mais que l'utilisateur n'en fait pas usage.

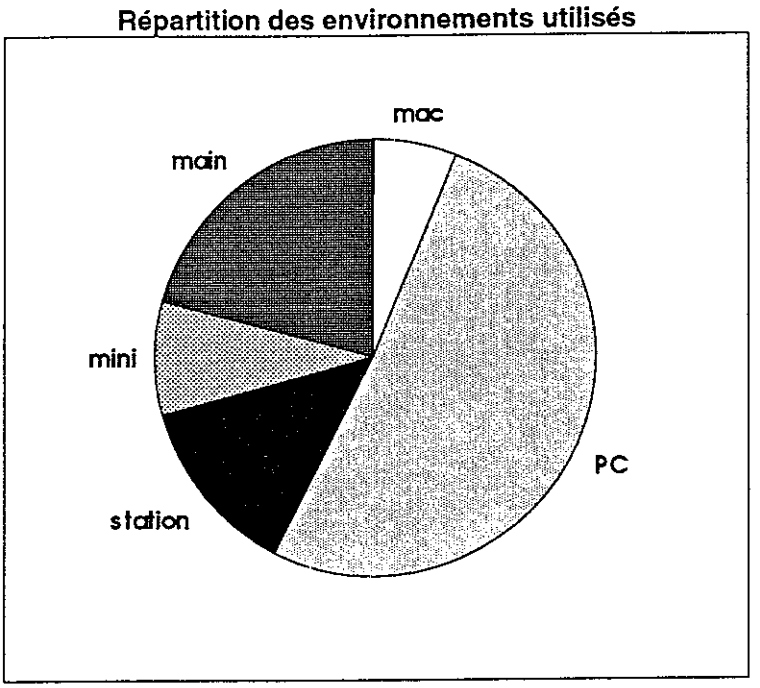


3.3 environnement du logiciel

Maintenant nous abordons la deuxième partie du questionnaire, dans laquelle le répondant décrit son logiciel et déclare le type de machine sur laquelle il l'utilise (question L4 : "dans quel environnement utilisez-vous ce logiciel?").

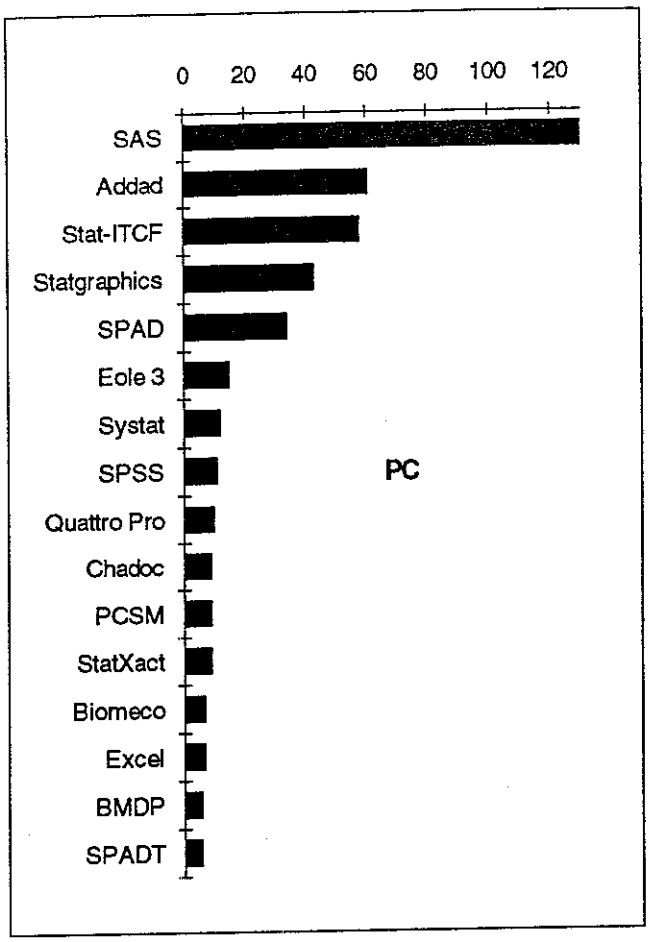
La question prévoyait la possibilité de réponses multiples. Au total, on dénombre 1059 associations entre un environnement et un logiciel, ce qui, rapporté aux 888 logiciels décrits, donne une moyenne de 1,2 environnements différents par logiciel.

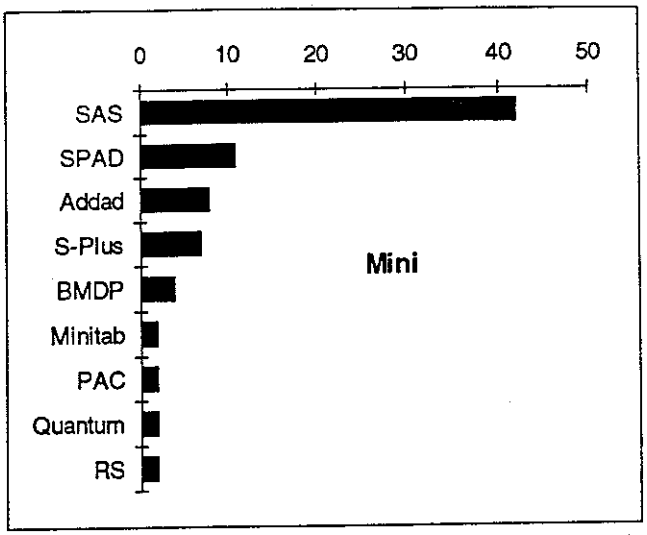
par type de machine	Macintosh	PC	station	mini	mainframe	total
nombre de logiciels cités	23	86	21	17	32	179
nombre de citations de logiciels	66	545	142	88	218	1059

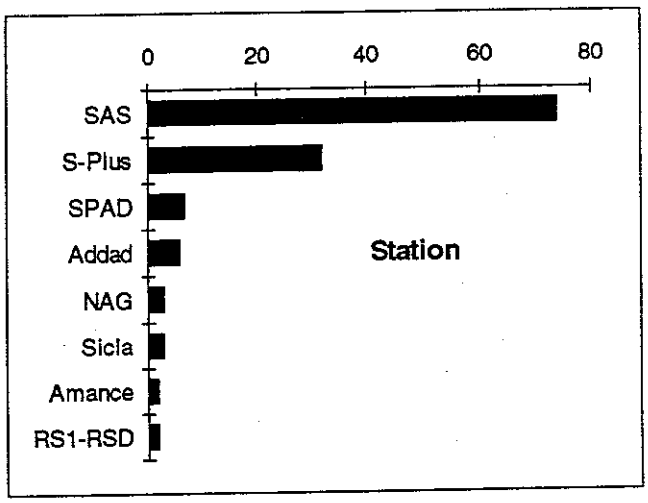


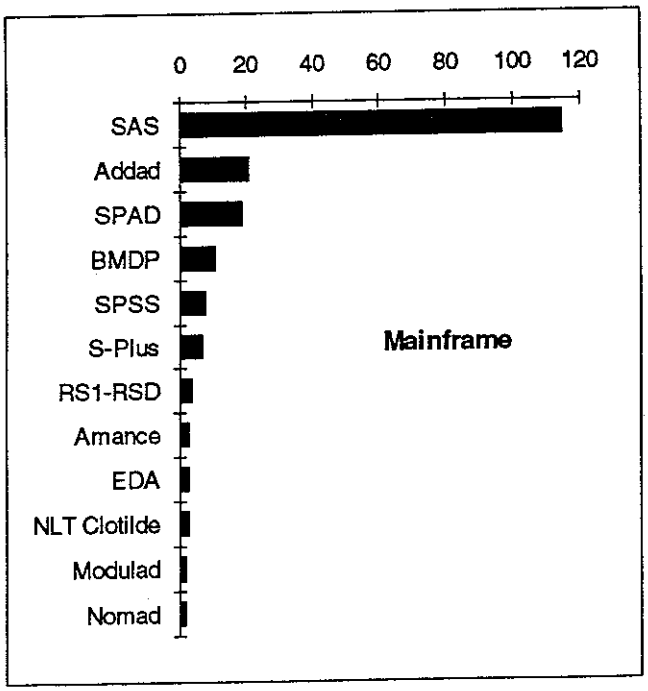
Les graphiques suivants donnent les classements des logiciels selon le type de machine utilisée. On retrouve la domination de SAS sur les 4 environnements où ce logiciel existe. SPAD.N et Addad se retrouvent en position 2 à 5 sur chacun des 5 environnements, leur position est voisine sur mainframe, mini et station de travail (après S-Plus dans ce dernier cas).

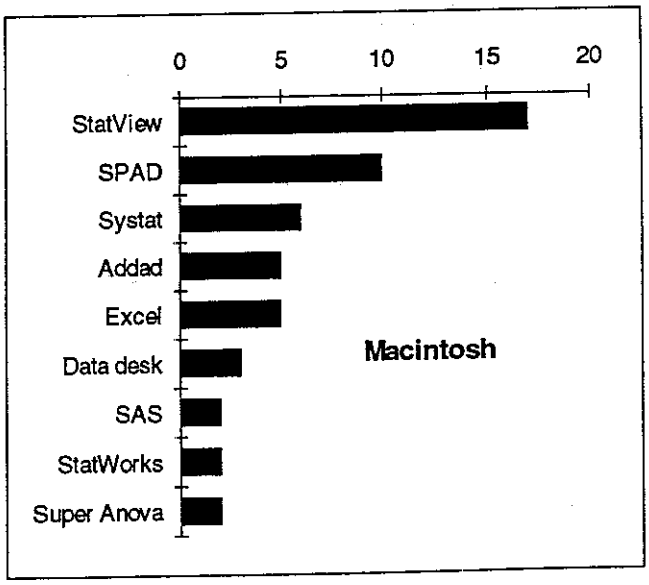
classement des logiciels selon le type de machine











4. Quels logiciels pour quelles techniques ?

Au cours de l'enquête ASU, il était demandé de décrire les trois logiciels les plus utilisés. Cette description comprenait :

- le nom du logiciel
- les techniques utilisées par la personne dans ce logiciel
- des éléments d'évaluation de la qualité de ce logiciel dans divers domaines, notamment l'assistance, la documentation et son interfaçage avec les autres logiciels.

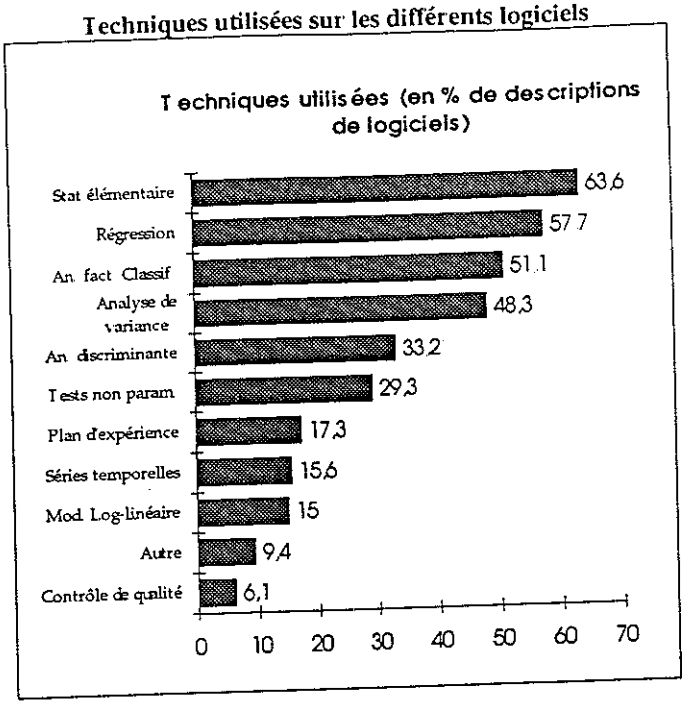
Pour chaque utilisateur, on a donc la description de sa propre utilisation d'au plus trois logiciels. Pour analyser cette information, nous avons considéré comme unité statistique de base une *description de logiciel*. Les 416 utilisateurs ayant répondu à l'enquête ont fourni au total 888 descriptions de logiciels, soit 2,1 logiciels décrits en moyenne par utilisateur.

Les logiciels qui ont été les plus fréquement décrits dans l'enquête sont : SAS (153 utilisateurs), Addad (83), SPAD.N (72), StatITCF (57), Statgraphics (46) et S-Plus (40). Plus marginalement (entre 10 et 17 descriptions), on trouve également SPSS, BMDP, Statview, Systat et Eole. Les autres logiciels ont été décrits moins de 10 fois. Rappelons que l'enquête réalisée à partir des fichiers de l'ASU, du club-utilisateur de SAS, et des clients de SPAD.N et ADDAD a certainement biaisé les résultats. Les interprétations devront surtout prendre en compte les différences entre les logiciels, entre les types d'utilisations, et accorder relativement peu d'importance au niveau de base des résultats.

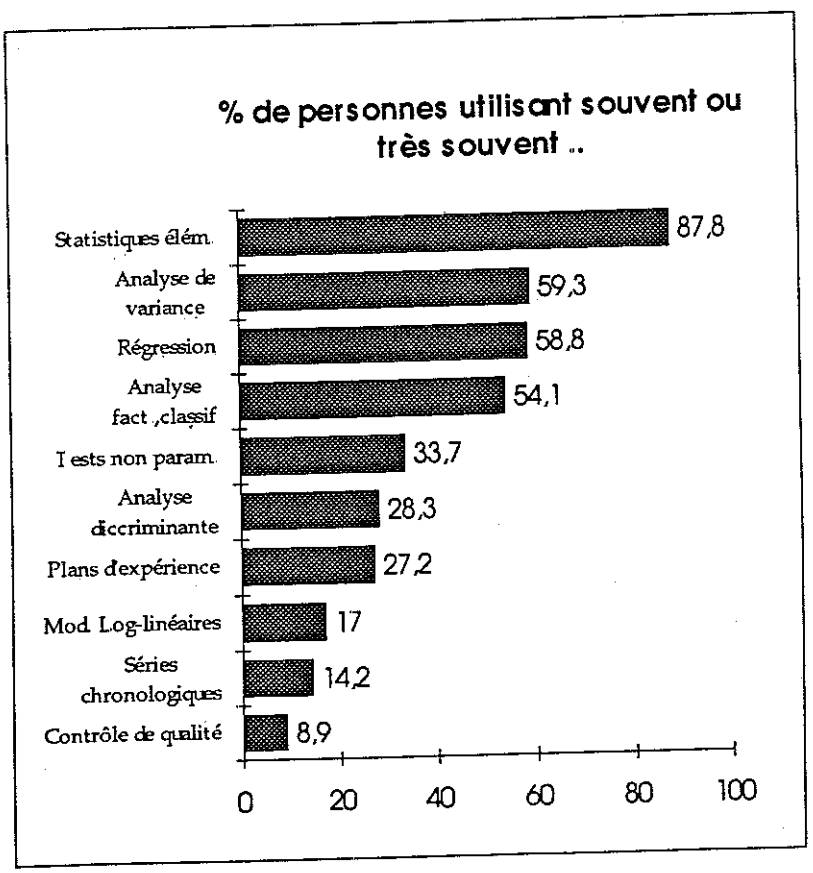
Ce chapitre est consacré aux techniques utilisées dans les différents logiciels. On essaie de voir si certains logiciels ont des utilisations très spécifiques, et lesquelles. Constructeurs et utilisateurs pourront ensuite évaluer l'écart éventuel entre l'ensemble des techniques proposées par les logiciels, et celles qui sont effectivement mises en oeuvre par les utilisateurs.

Le tableau des "descriptions de logiciels" a été soumis dans un premier temps à une analyse des correspondances multiples, les variables actives étant les techniques utilisées dans les logiciels. Dans un deuxième temps, des classes ont été construites regroupant des utilisations de logiciels proches quant aux techniques employées.

Techniques utilisées sur les différents logiciels



La répartition des techniques recensées par le biais des descriptions de logiciels est assez proche de leur répartition en pourcentage d'utilisateurs. Les statistiques descriptives ne sont cependant utilisées que dans 64% des logiciels décrits, alors que 88% des utilisateurs les pratiquent (on se demande d'ailleurs comment font les autres !). Les plans d'expérience sont également moins présents dans les descriptions par logiciel que parmi les utilisateurs.

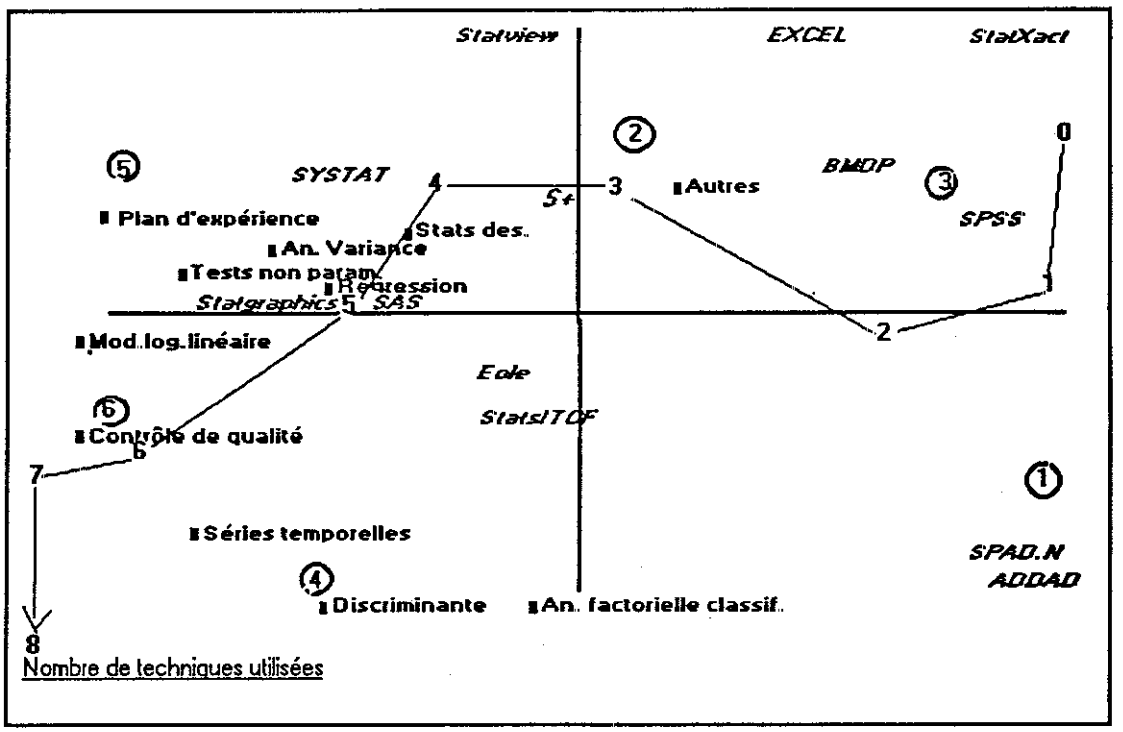


4.1 Analyse des correspondances multiples

Une analyse des correspondances multiples a été réalisée. Les variables actives sont les suivantes : utilisation de ...

Statistiques descriptives élémentaires
Régression
Plan d'expérience
Analyse de la variance
Séries chronologiques, prévision
Contrôle de qualité
Analyses factorielles et classifications
Tests non paramétriques
Analyse discriminante
Modèle log-linéaire, survie
Autre technique utilisée

Plan factoriel (1,2)



Sur le plan factoriel (1,2) les logiciels utilisés ont été positionnés en éléments illustratifs ainsi que le nombre de techniques utilisées dans chaque logiciel et les classes de descriptions de ces logiciels. Ces classes ont étés construites à partir des coordonnées factorielles.

Le premier axe du plan factoriel oppose les logiciels qui ne sont utilisés que pour une seule technique, ou une "autre technique", à ceux qui sont utilisés pour une armada de méthodes différentes, et essentiellement des techniques de modélisation, modèle log-linéaire, contrôle de qualité, séries temporelles, plan d'expérience. SAS, Statgraphics et Systat sont les logiciels les mieux représentés du côté de la modélisation et de la multi-utilisation. Le deuxième axe oppose analyses factorielles et classifications, analyse discriminante et séries chronologiques aux autres techniques.

La classification va permettre d'affiner l'interprétation de ce premier plan factoriel.

4.2 Description des 6 classes obtenues

Classe 1 (27%) : la spécialisation, surtout analyses factorielles et classifications

Cette classe est celle des utilisations spécialisées : peu de techniques sont mises en oeuvre dans le même logiciel (1,5 contre 3,5 en moyenne). Ce sont les analyses factorielles et les classifications qui sont le plus souvent retenues dans ces logiciels (77%). Les autres techniques sont peu utilisées, surtout les statistiques descriptives (12%), la régression (10%) et l'analyse de variance (8%).

Logiciels : ADDAD ou SPAD.N mais pas SAS

Les logiciels spécialisés, ou employés de manière spécialisée, sont en premier lieu ADDAD (29%) et SPAD.N (24%). SAS correspond rarement à ce type d'utilisation, (6% contre 28% en moyenne), ainsi que Statgraphics (1% ; 5%) et S-Plus (1% ; 5%).

Insatisfaction vis-à-vis des logiciels

Ce mode d'utilisation va souvent de pair avec une certaine insatisfaction des utilisateurs sur la gestion des données et l'interfaçage de ces logiciels. L'aide à l'écran apparaît souvent inexistante (43% contre 24% en moyenne), la gestion des données peu satisfaisante (39% contre 24% en moyenne), l'édition des données ainsi que les échanges avec d'autres systèmes de gestion de base de données inexistantes (20 et 18% ; 10 et 10%). L'échange des données avec d'autres logiciels et la réaction du logiciel à une erreur sont déclarés peu

satisfaisants (35% et 38% ; 24% et 28%). On conçoit aisément que ces logiciels ne soient employés que pour une ou deux techniques : leur mise en oeuvre ne se justifie que pour des emplois très spécifiques.

Classe 2 (33%) : statistiques descriptives et régressions

Les logiciels sont essentiellement utilisés pour les statistiques descriptives (88% ; 64%), ainsi que la régression (72% ; 58%) mais pas du tout pour les séries chronologiques ou la prévision (0% ; 16%), les plans d'expérience (2% ;17%) ou le contrôle de qualité (0% ; 6%). Les autres méthodes sont également très peu utilisées.

Logiciels : ni SPAD.N, ni ADDAD

Les logiciels employés pour utiliser ces techniques statistiques (3 en moyenne) sont très rarement SPAD.N (4% ; 8%) et ADDAD (3% ; 9%).

Secteur d'activité : recherche

53% des utilisations de logiciels de cette classe ont été faites par des personnes travaillant dans la recherche contre 45% sur la population. Par contre, l'utilisation est faible dans le domaine de l'industrie non pharmaceutique (4% ; 7%).

Classe 3 (8%) : autres techniques

Les utilisations de cette classe se portent toutes sur d'autres techniques statistiques que celles énumérées dans le questionnaire (100% ; 9%). Il s'agit la plupart du temps de techniques graphiques, de segmentation, de modèles non linéaires, d'analyses textuelles, de tests ou d'analyse de données sur variables instrumentales.

Logiciels : autres

Les logiciels utilisés sur ces techniques sont pour la plupart différents des 14 logiciels les plus utilisés (47% ; 25%).

Secteur d'activité : énergie

Le pourcentage d'utilisation des logiciels dans l'énergie est plus important dans la classe que dans la population : il atteint 8% des utilisations contre 2% dans la population.

Classe 4 (10%) : séries chronologiques et régressions

Ces utilisations comportent toutes l'emploi des séries chronologiques (100% ; 16%). La plupart comptent aussi la régression (95% ; 58%) et les statistiques descriptives (85% ; 64%). D'autres techniques sont également plus souvent mises en oeuvre qu'en moyenne : l'analyse discriminante (51% ; 33%), les analyses factorielles et classifications (65% ; 51%) et l'analyse de variance (62% ; 48%). Par contre le contrôle de qualité n'est jamais employé (0% ; 6%). Il s'agit donc d'un type d'utilisation assez vaste de l'ensemble des techniques statistiques courantes.

Logiciels : ni SPAD.N, ni ADDAD

Les logiciels utilisés pour mettre en oeuvre ces nombreuses techniques (5,3 en moyenne contre 3,5 dans la population), ne sont pratiquement jamais SPAD.N (1% ; 8%) ni ADDAD (1% ; 9%). Cela n'est guère étonnant, ces deux logiciels ne comportant aucun module d'analyse des séries chronologiques.

Secteur d'activité : enseignement et formation

30% des utilisations de logiciels sont faites par des enseignants ou des formateurs contre 19% sur la population totale.

Classe 5 (16%). Plans d'expérience, analyse de variance, régressions

La caractéristique principale de cette classe est l'utilisation massive des plans d'expérience (70% contre 17% en moyenne). Dans les logiciels décrits, on utilise pratiquement toujours l'analyse de variance (94% ; 48%), la régression (87% ; 58%) ainsi que les statistiques descriptives (87% ; 64%). De plus, dans une majorité de cas, les utilisateurs se servent également des modèles log-linéaires ou de survie (56% ; 15%) et des tests non paramétriques (56% ; 29%). Par contre, les séries chronologiques (4% ; 16%) et le contrôle de qualité (0% ; 6%) ne sont pas utilisés dans les logiciels cités. Comme on le voit, le champ des techniques employées est vaste : 5,4 techniques par logiciel décrit.

Logiciels : SAS

SAS est le logiciel employé pour plus de la moitié des utilisations de cette classe (58% ; 28%). Par contre, ADDAD et SPAD.N ne sont pas employés (0% et 1%).

Participation aux groupes utilisateurs des logiciels

La plupart des utilisations de logiciels sont faites par des personnes qui connaissent l'existence des groupes d'utilisateurs (65% ; 46%) et y participent (42% ; 22%). Trouvant

l'assistance technique téléphonique de faible qualité (21% ; 10%), ils considèrent la documentation en brochure indispensable (76% ; 64%). Cette documentation sur l'aspect statistique est considérée comme bonne (56% ; 44%).

Secteur d'activité : industrie pharmaceutique

25% des utilisations de logiciels sont faites par des enquêtés qui travaillent dans l'industrie pharmaceutique contre 10% dans la population. Par contre, ces utilisations sont absentes en marketing ou dans la publicité (0% ; 7%).

Classe 6 (6%) Contrôle de qualité

La caractéristique principale de la classe est l'utilisation par tous du contrôle de qualité (100% ; 6%). Les autres techniques sont également couramment utilisées : séries chronologiques ou prévision (54% ; 16%), plans d'expérience (48% ; 17%), analyse de variance (79% ; 48%), tests non paramétriques (58% ; 29%), régression (79% ; 58%), analyse discriminante (50% ; 33%) et statistiques descriptives (83% ; 64%). C'est la classe dans laquelle le nombre de techniques employées est le plus important (6,4 en moyenne).

Logiciels : Statgraphics

Le logiciel le plus caractéristique de cette classe est Statgraphics (23% ; 5%).

Utilisateurs satisfaits

37% des utilisations sont faites par des personnes qui considèrent que l'échange des données avec d'autres logiciels est très satisfaisant (19% dans la population) et 40% par des enquêtés très satisfaits par l'édition des données (20% dans la population).

Secteur d'activité : industrie

25% des utilisations de logiciels de cette classe sont réalisées par des personnes qui travaillent dans l'industrie contre 6% sur la population totale. Par contre, seulement 17% des utilisations contre 45% dans la population sont faites dans le domaine de la recherche.

5. Qui utilise quoi ?

5.1 Introduction sur les techniques statistiques

La première partie du questionnaire comprend 22 questions portant sur la connaissance de l'utilisateur : l'âge, le sexe, les formations reçues, les activités professionnelles, les activités associatives savantes et non lucratives, les techniques statistiques utilisées, l'environnement informatique et les logiciels utilisés.

Cette première partie a donné naissance à un tableau de données comportant 416 individus et 60 variables (45 qualitatives et 15 quantitatives).

La distribution de ces variables sur l'échantillon révèle quelques modalités rares et de nombreuses non-réponses partielles : aussi les modalités ayant des effectifs faibles ont été regroupées et une modalité supplémentaire caractérisant cette non-réponse a été introduite afin de tenir compte de cette absence de réponse. Les variables quantitatives ont été transformées en qualitatives et l'ensemble des variables regroupées en cinq thèmes :

1- le thème signalétique
2- le traitement statistique
3- l'expérience en statistique
4- l'environnement informatique
5- les logiciels utilisés

Le thème analysé ici est le thème du traitement statistique qui contient les informations relatives aux méthodes statistiques utilisées ; il est bâti sur la question suivante :

"Quelles sont les techniques statistiques que vous utilisez et avec quelle fréquence?".

Q12-1 statistiques descriptives
Q12-2 régressions
Q12-3 plan d'expérience
Q12-4 analyse de la variance
Q12-5 séries chronologiques, prévision
Q12-6 contrôle de qualité
Q12-7 analyses factorielles et classifications
Q12-8 tests non paramétriques
Q12-9 analyse discriminante
Q12-10 modèle log-linéaire, survie
Q12-11 autre. Précisez :_____

La réponse à ces questions permet d'obtenir une description de la fréquence et de la diversité des techniques statistiques mises en oeuvre. Les réponses concernant les "autres" méthodes étant peu nombreuses, cette variable a été éliminée ainsi il a été retenu 10 variables actives, toutes les autres intervenant en supplémentaires. Sur ce thème il a été effectué deux analyses, chacune utilisant la même méthodologie : une analyse factorielle suivie de classifications.

La première a été réalisée sur l'ensemble des dix questions de ce thème. Le codage des modalités est le suivant :

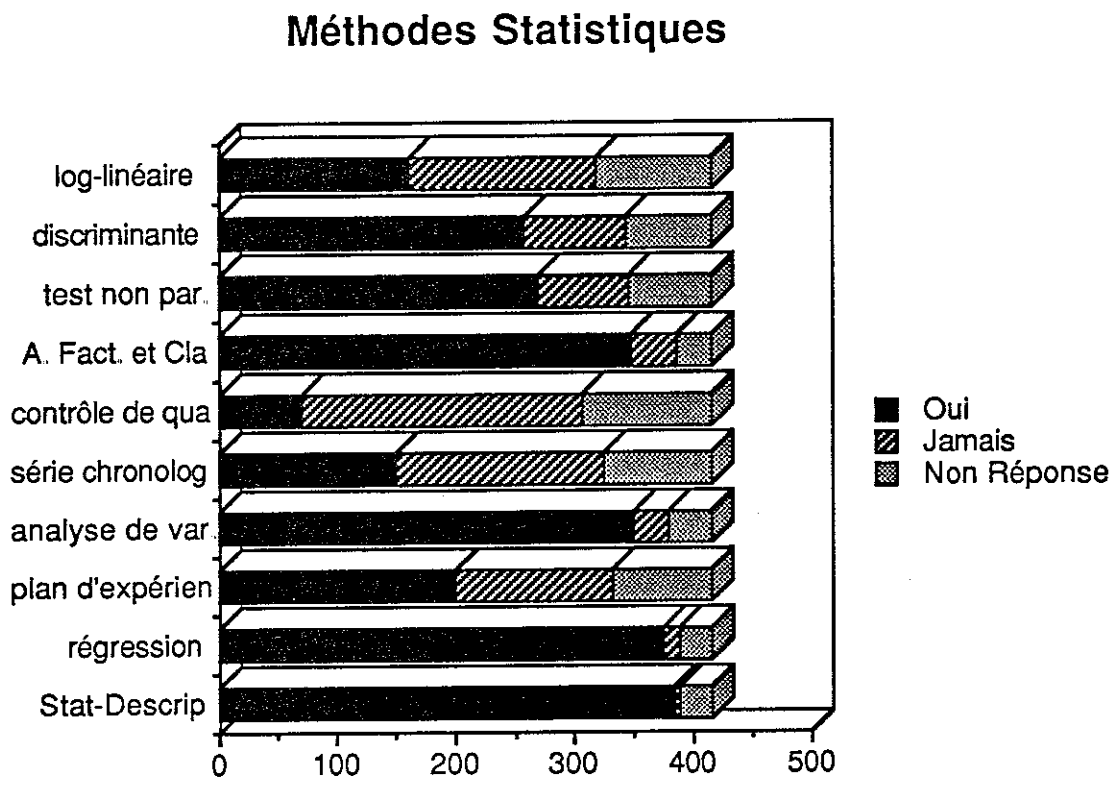
1 : modalité associée à la réponse "Jamais"
2 : modalité associée à la réponse "Occasionnellement"
3 : modalité associée à la réponse "Souvent"
4 : modalité associée à la réponse "Très Souvent"
5 : modalité associée à la réponse "non-réponse"

Ainsi qx_y représente la modalité numéro y de la variable numéro x. Par exemple la modalité q3_1 correspond à la réponse "Jamais" à la question Q12-3.

La seconde a été réalisée sur les mêmes questions, mais nous avons regroupé certaines modalités de ces variables. Ce regroupement est réalisé de la manière suivante :

1 : modalité associée à la réponse "Jamais"
2 : modalité associée à la réponse "Occasionnellement", "Souvent" et "Très Souvent".
L'abrégé de cette modalité 2 est "Oui".
5 : modalité associée à la "non-réponse"

Voici la ventilation de ces questions après ce recodage :



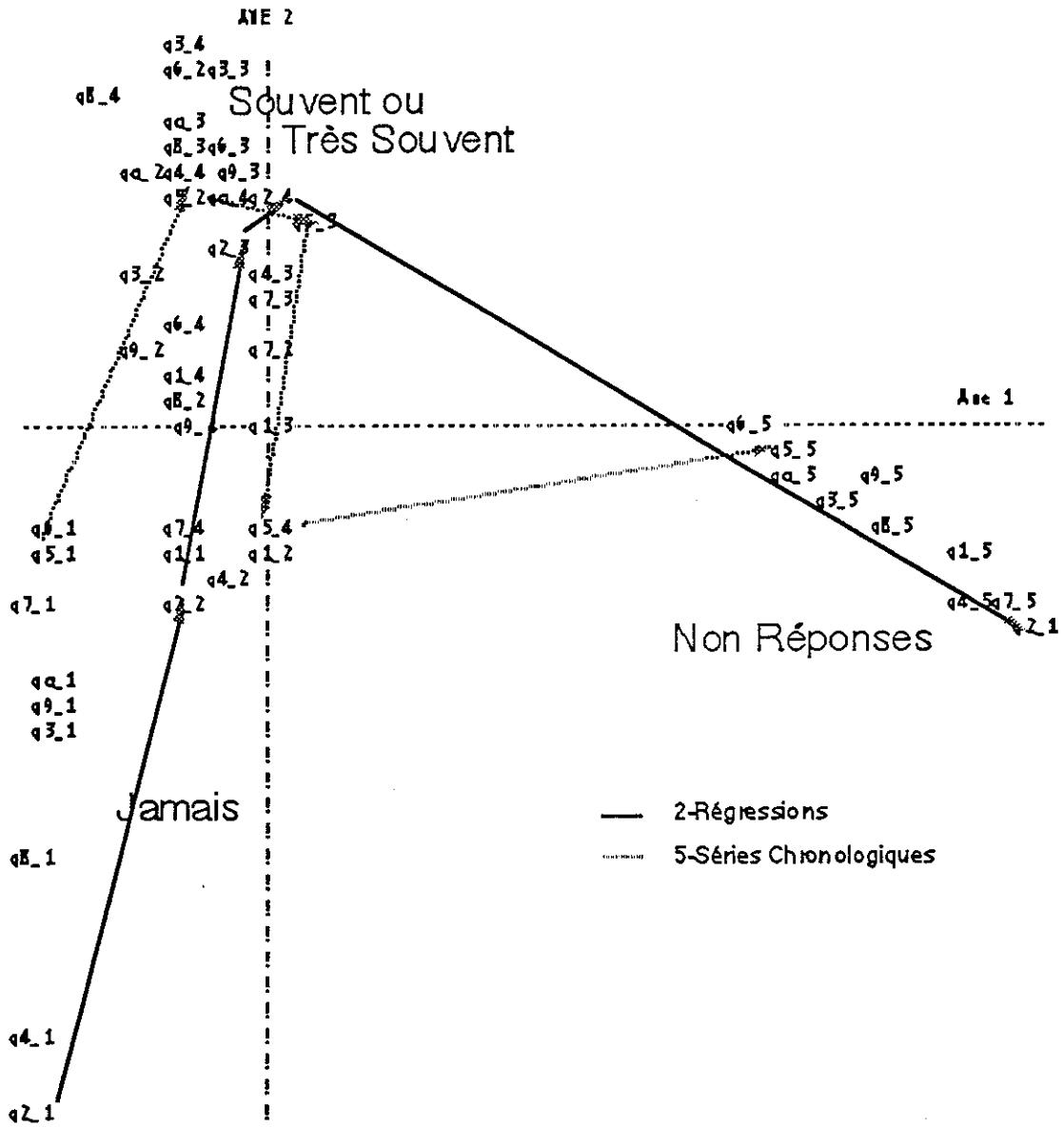
Afin de permettre une meilleure interprétation de ces questions nous avons créé trois nouvelles variables. La première variable comptabilise le nombre de réponses positives à ces 10 questions. Une réponse positive est d'avoir sélectionné l'une des modalités suivantes "Occasionnellement", "Souvent" ou "Très Souvent". La deuxième variable comptabilise les réponses associées à la modalité "Jamais". La troisième variable comptabilise les réponses associées à la modalité "non-réponse".

5.2 Analyse factorielle de l'utilisation des techniques statistiques

La première valeur propre est très grande et elle représente presque 15 % de l'inertie. Les trois valeurs propres suivantes représentent à peu près la même part d'inertie (environ 5%). Sur le premier plan factoriel nous avons un bel effet Guttman. Le premier facteur est associé aux modalités "non-réponse" du questionnaire, ces modalités se trouvent à droite de ce plan factoriel.

Le deuxième axe représente le degré d'utilisation des méthodes statistiques ; en bas de cet axe nous avons les modalités associées à la réponse "Jamais" ; en haut de cet axe nous avons les modalités associées aux réponses "Souvent" et "Très Souvent".

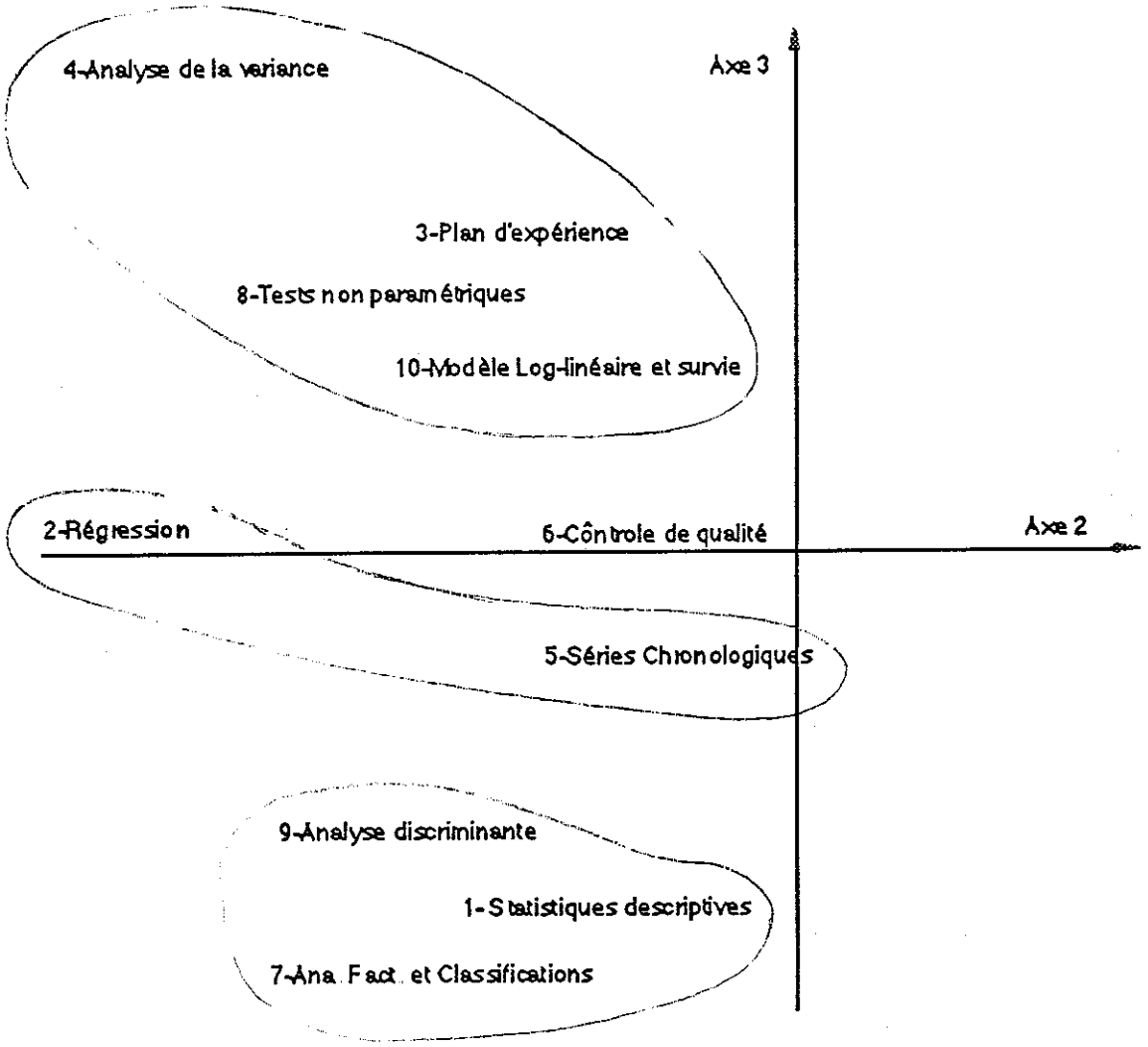
Plan factoriel (1-2)



Les variables Q12_2 (régressions), Q12_3 (plan d'expérience) et Q12_4 (analyse de la variance) sont les parfaits exemples de cet effet Guttman. Par contre les variables Q12_5 (séries chronologiques, prévision) et Q12_6 (contrôle de qualité) ont la modalité "Très Souvent" proche de la modalité "Jamais" et donc nous avons un retour vers le centre du

plan factoriel de la modalité "Très Souvent" ce qui perturbe un peu cet effet Guttman. Ceci peut s'expliquer par la mono-utilisation de ces techniques et la spécialisation des utilisateurs de ces techniques. Plus précisément ces utilisateurs font appel très souvent à une panoplie restreinte de techniques en excluant les autres.

Plan factoriel (2-3)



Sur ce plan factoriel (2-3) les modalités de la variable Q12_1 sont reliées par un trait continu, celles de la variable Q12_5 pour un trait en pointillé.

Ce plan factoriel est obtenu après recodage des variables en trois modalités (2ème analyse). Sur ce plan factoriel nous avons visualisé la modalité "Oui" de chacune des techniques car la modalité "Non Réponse" est toujours au centre et elle est située entre les deux autres modalités ("Jamais" et "Oui") de cette variable. Ce plan factoriel est plus intéressant car il élimine l'effet de la modalité "Non Réponse" et permet de mieux analyser

comportement des utilisateurs des méthodes statistiques. C'est sur ce plan factoriel que nous pouvons interpréter facilement les classes obtenues par la classification, il y apparaît trois grandes familles de variables :

Le premier groupe se situe sur le deuxième axe factoriel et il comprend les variables Q12_2 et Q12_5 qui sont associées aux méthodes de régression et de séries chronologiques.

Le deuxième groupe, se composant des variables Q12_1, Q12_7 et Q12_9, se situe sur l'axe factoriel numéro 3. Les variables Q12_7 et Q12_9 caractérisent à elles seules 61% de la contribution à l'axe 3 qui lui explique 7,6% de l'inertie totale. Ce groupe représente les méthodes d'analyse factorielles, de classification, de discrimination et les techniques descriptives.

Le troisième groupe est composé, en fonction des contributions, essentiellement des variables Q12_4 et Q12_8, puis des variables Q12_3 et Q12_10. Il se situe sur la deuxième diagonale de ce plan factoriel. Ce groupe représente les méthodes d'analyse de variance, les plans d'expérience, les tests non paramétriques et les méthodes log-linéaires.

L'axe factoriel numéro 3 oppose les utilisateurs travaillant avec les méthodes du deuxième groupe aux utilisateurs utilisant les méthodes du troisième groupe.

Du côté des analyses factorielles, des classifications et de l'analyse discriminante les activités d'assurance, finance, banque et de gestion sont bien représentées ; le secteur de l'industrie pharmaceutique se trouve quant à lui du côté des plans d'expérience et des méthodes log-linéaires ceci va être confirmé par la classification.

5.3 Étude de la partition en 5 classes

Cette partition a été obtenue en appliquant la Méthode des Nuées Dynamiques sur l'ensemble des variables actives de l'analyse factorielle précédente. L'interprétation de cette partition est réalisée à partir des variables ayant servi à son obtention ; puis chacune de ces classes seront expliquées en utilisant toutes les variables de la première partie du questionnaire. La classe, associée à la modalité "non-réponse", est bien représentée sur le premier plan factoriel car elle contribue à l'effet Guttman de ce plan factoriel. Les autres classes de la partition sont, par contre, interprétables sur le deuxième plan factoriel.

Classe 1 (30%) : La polyvalence

L'effectif de cette classe est de 124 et elle contient 30% de la population. Cette classe est la plus importante de cette partition en 5 classes. Elle représente la population des enquêtés qui utilisent presque toutes les méthodes statistiques sauf le contrôle de qualité.

Une large panoplie de méthodes

Ces personnes utilisent souvent l'analyse de la variance (55% des personnes de cette classe contre 27% des personnes enquêtées), la régression (64% ; 35%), l'analyse factorielle et classification (40% ; 25%), les statistiques descriptives (40% ; 27%). Puis elles utilisent souvent ou occasionnellement l'analyse discriminante (pour "souvent" nous avons 27% ; 17% et pour "occasionnellement" nous avons 53% ; 39%), les tests non paramétriques ("souvent" : 32% ; 19%, "occasionnellement" : 53% ; 37%), les séries chronologiques (11% ; 6% et 50% ; 25%) et le plan d'expérience (18% ; 12% et 52% ; 27%). Enfin elles utilisent "occasionnellement" les méthodes log-linéaires (53% ; 26%) et "jamais" le contrôle de qualité (69% ; 59%).

La présence d'un nombre élevé de réponses positives à l'utilisation des méthodes statistiques (85% des personnes de cette classe utilisent plus de 7 techniques différentes contre 46% des personnes enquêtées) confirme une pratique statistique très large.

Très grande activité associative, une grande variété de logiciels et de matériels utilisés

Le matériel de ces utilisateurs est essentiellement, comme toutes les personnes interrogées, le PC mais ici ces personnes emploient aussi le Mini (29% ; 18%) et le Mac (34% ; 25%).

Ces utilisateurs sont membres de l'ASU (35% ; 25%) et participent aux congrès de cette association (37% ; 26%). Ils sont responsables du choix des logiciels pour leur société (29% ; 20%) et ont une formation BAC+5 (35% ; 27%).

Ils utilisent dans la même proportion que l'ensemble de la population le logiciel SAS mais SPAD.N (35% ; 22%) est beaucoup plus proportionnellement utilisé que dans les autres classes. Les logiciels Statgraphics (20% ; 14%), SPSS (10% ; 6%), Systat (9% ; 5%), Stat-ITCF (25% ; 18%) et BMDP (10% ; 6%) le sont aussi. Cette grande variété de logiciels démontre qu'il n'existe pas de logiciel universel et qu'une pratique large des statistiques entraîne le choix d'un nombre important de logiciels.

Classe 2 (25%) : Les très spécialisés

Cette classe est importante (103 réponses) car elle contient 30% de la population. Elle représente les personnes n'utilisant jamais un ensemble important de méthodes d'analyse de données. Les principales méthodes non utilisées sont les méthodes log-linéaire (83% ; 38%), le plan d'expérience (74% ; 31%), les séries chronologiques (82% ; 42%) et le contrôle de qualité (89% ; 56%). Très occasionnellement elles utilisent les méthodes de régression (62% ; 35%) ou d'analyse de la variance (43% ; 30%). 63% des personnes de

cette classe utilisent toujours moins 5 techniques différentes contre 16% des personnes enquêtées.

Peu de types de matériel utilisés et activité récente en statistique

Ces utilisateurs ne font pas souvent appel à l'informatique car la modalité "Jamais" est présente pour le PC (13% ; 6%), pour la station de travail (52% ; 32%) et pour le Mac (60% ; 42%). Ils ont moins de 6 ans d'utilisation des logiciels statistiques (67% ; 56%) ce qui peut expliquer ces réponses. Ils ne participent pas au congrès ASU (86% ; 74%).

Nous avons réalisé également un partitionnement en 7 classes de la population. Ceci à induit un découpage de cette classe 2 en deux sous-classes, ce qui va nous permettre de mieux interpréter les choses :

Classe 2_1 (9%) : Les très spécialisés utilisant l'analyse de la variance, les tests non paramétriques et les méthodes log-linéaires.

Cette classe est composée des personnes ayant une pratique importante de certaines méthodes statistiques. Ces personnes utilisent très souvent l'analyse de la variance (50% ; 27%), les tests non paramétriques (21% ; 9%) et méthodes log-linéaires (12% ; 4%). Cependant ils n'utilisent jamais l'analyse factorielle et classification (100% ; 9%), l'analyse discriminante (82% ; 20%). La panoplie des méthodes utilisées est très restreinte mais leur utilisation est intensive.

Le principal secteur d'activité est l'industrie pharmaceutique : 44% ; 11%.

Ces personnes n'ont pas les logiciels ADDAD et SPAD.N et ne sont pas responsables du choix des logiciels pour leur société (94% ; 80%). Ils participent à d'autres congrès que les congrès en statistiques (38% ; 22%) et sont membres de l'ASU (56% ; 29%).

Cette classe peut être interprétée comme un groupe de personnes encore plus spécialisées que celles de la classe 4 bien qu'appartenant au même secteur d'activité.

Classe 2_2 (16%) : Les statisticiens très spécialisés utilisant l'analyse factorielle, la classification et la régression.

Cette classe est composée des personnes ayant une pratique importante (très souvent) de l'analyse factorielle et de la classification (39% ; 25%). Elles n'utilisent jamais ni méthodes log-linéaires (91% ; 38%) ni plans d'expérience (93% ; 31%).

Le secteur d'activité est l'assurance, la finance et la banque

Une partie de cette population a comme domaine d'activité l'assurance, finance et la banque (25% ; 11%). Ils n'utilisent jamais comme matériel informatique les Minis (72% ; 45%) et ne participent pas aux congrès de l'ASU (89% ; 74%).

Classe 3 (24%) : Les non-réponses

Cette classe est importante (98 personnes) et elle représente les personnes qui n'ont pas répondu à la question Q12. Il y a beaucoup de non-réponses dans le questionnaire de ces personnes car on retrouve cette modalité dans l'utilisation du matériel informatique.

Les principales non-réponses sont sur les méthodes log-linéaires (89% ; 24%), le contrôle de qualité (93% ; 27%) et les séries chronologiques (81% ; 22%).

Ils n'emploient aucun logiciel autre que SAS en statistique, c'est peut être lié à la non connaissance de l'informatique. Ils ne participent pas aux congrès de statistiques ASU (91% ; 74%) et autres (95% ; 78%), ni ne sont membres d'une société savante en statistique (88% ; 62%). Ils ont une formation de moins d'un an en statistique (32% ; 15%). 60% des personnes de cette classe ont plus de 6 non-réponses à ces 10 questions.

Classe 4 (17%) : Les spécialisés utilisant l'analyse de la variance, les tests non paramétriques et le plan d'expérience

Cette classe est composée de 71 personnes ayant une pratique importante des méthodes statistiques. La panoplie des méthodes utilisées est un peu plus restreinte que celle de la classe 1 mais beaucoup plus intensive.

L'analyse de la variance (96% ; 27%), les tests non paramétriques (44% ; 9%), le plan d'expérience (41% ; 9%), la régression (38% ; 20%) et les méthodes descriptives (87% ; 55%) sont très souvent utilisées.

Les méthodes factorielles et de classification (55% ; 33%) ainsi que la discrimination (58%; 39%) ne sont utilisées qu'occasionnellement. Les méthodes associées aux séries chronologiques (72% ; 42%) ne sont jamais utilisées.

Le secteur d'activité est la médecine, la pharmacie et le biomédical

Il y a beaucoup d'utilisateurs du monde pharmaceutique (50% ; 14%). Ils sont membres de l'ASU (58% ; 29%) et participent aux congrès (48% ; 26%) comme ceux de la classe 1. Le profil personnel est très souvent le conseil (56% ; 30%), ce sont des utilisateurs de SAS (86% ; 63%). Ils possèdent d'autres logiciels mais ne les utilisent pas ; ceci explique peut être leur spécialisation dans le type d'études qu'ils dirigent.

Classe 5 (5%) : séries chronologiques

C'est une petite classe (20 réponses) et elle est composée des personnes ayant une pratique très importante des séries chronologiques (100% ; 6%), du contrôle de qualité (20% ; 6%), de la régression (50% ; 20%), de l'analyse factorielle et classification (55% ; 25%) et de la

discrimination (30% ; 10%). La pratique de ces techniques était occasionnelle ou nulle dans la classe 4. La panoplie des méthodes pratiquées est très complémentaire à celles utilisées par les personnes de la classe 4 précédente.

Le profil personnel et le domaine d'activité est la recherche et l'enseignement (35% ; 15%). Ils n'ont pas de comportements particuliers vis à vis des moyens informatiques et des logiciels utilisés.

5.3 Conclusion sur les liens entre les techniques et les logiciels

Les deux études présentées aux paragraphes 4. et 5. montrent une très grande cohérence entre les familles de techniques utilisées.

Les techniques descriptives et les méthodes de régression sont utilisées dans toutes les pratiques statistiques.

Les méthodes d'analyse factorielles, de classification, de discrimination forment un groupe de méthodes qui sont utilisées dans les secteurs de la recherche, de l'assurance et de la médecine. L'autre groupe représente les méthodes d'analyse de variance, les plans d'expérience, les tests non paramétriques et les méthodes log-linéaires. Ces méthodes sont surtout utilisées dans le secteur de l'industrie pharmaceutique. La pratique du contrôle de qualité et des séries chronologiques est très ponctuelle et non liée à un secteur d'activité.

La classe 1 associée aux techniques statistiques est caractérisée par la polyvalence et représente 30% de la population. Dans cette classe beaucoup de logiciels sont cités car les personnes de cette classe pratiquent plus de 7 méthodes statistiques différentes et l'analyse de ces logiciels réalisée au paragraphe 4 montre qu'un logiciel, en moyenne, ne couvre que 4 types de techniques statistiques. Ceci montre qu'aucun logiciel de statistique n'est perçu comme universel par ces utilisateurs.

Les autres classes de cette partition sont beaucoup plus spécialisées. Les classes associées à la pratique de l'analyse factorielle, de la classification et de la discrimination utilisent les logiciels SPAD.N et ADDAD. Les classes associées à la pratique de l'analyse de variance, des plans d'expérience, des tests non paramétriques et des méthodes log-linéaires utilisent les logiciels SAS, BMDP, Systat, SPSS et Statgraphics.

En conclusion ces deux grands groupes de méthodes se retrouvent aussi bien du côté des utilisateurs que du côté des logiciels. Ceci peut montrer l'influence des logiciels sur la méthodologie employée par les utilisateurs.

6. La question ouverte

Après avoir recueilli les opinions concernant les qualités et les défauts des logiciels utilisés, on a tenté de faire s'exprimer les souhaits des enquêtés à travers une question ouverte. Celle-ci était libellée de la manière suivante:

"Dans ce questionnaire, vous vous êtes exprimé sur le contenu des logiciels de statistique et leurs qualités ou défauts. Certains points importants pour vous n'ont peut être pas été évoqués. Pouvez-vous conclure en exprimant ce que serait pour vous, d'une façon générale, un bon logiciel de statistique?"

Près de la moitié des répondants ont laissé cette réponse en blanc (216 réponses exprimées sur 416 répondants). Ce fort taux de non-réponse, sans être en-dehors des normes usuellement rencontrées, ne montre pas un enthousiasme exagéré pour cette question.

FREQ		FREQ		FREQ	
27	aide	26	graphique	11	outil
5	aide en ligne	42	graphiques	10	outils
5	aide à l interprétation	21	interface	12	permettre
23	analyse	12	interprétation	31	possibilité
20	analyses	15	en français	17	possibilités
10	autres logiciels	14	erreurs	11	procédures
6	bases de données	14	exemples	23	programmation
19	bon logiciel de statistique	25	facile	11	programmes
13	bon logiciel statistique	11	facilement	21	qualité
13	bonne documentation	10	facilité	28	résultats
10	commande	10	fichiers	10	saisie
10	complet	19	français	77	SAS
21	convivial	26	gestion	10	sgbd
19	convivialité	6	gestion de données	14	simple
5	de bonne qualité	12	gestion des données	28	sorties
15	doc	26	graphique	10	sorties graphiques
53	documentation	42	graphiques	21	stat
7	documentation en français	21	interface	85	statistique
89	données	12	interprétation	69	statistiques
15	en français	7	langage de commande	12	techniques
14	erreurs	6	langage de programmation	23	traitement
14	exemples	10	manipulation	12	traitements
25	facile	11	manque	32	utilisateur
11	facilement	11	menu	19	utilisateurs
10	facilité	10	méthode	50	utilisation
10	fichiers	48	méthodes	13	utiliser
19	français	15	méthodes statistiques	10	utilisées
26	gestion				
6	gestion de données				
12	gestion des données				

Table 1. Vocabulaire des réponses

Le corpus, formé par les 216 réponses exprimées, a une longueur de 7376 mots et est formé par 294 mots distincts.

La table 1 reproduit les mots pleins les plus fréquents (employés au moins 10 fois) ainsi qu'une partie des segments répétés détectés dans le corpus, ces segments ont été choisis afin de préciser le contexte des mots. Seuls les 152 mots prononcés au moins 10 fois sont conservés, mais y compris les mots-outils.

La lecture de ce glossaire, enrichie de celle des concordances de nombreux mots, permet de dégager les principaux thèmes de réponse, que nous présentons ci-dessous, en les illustrant par des extraits de réponse.

Un point important apparaît être la documentation (53 citations de "documentation" et 15 citations de "doc" dans les 216 réponses), avec une grande demande d'aide statistique, de documentation conçue comme une formation à la statistique, offrant une aide méthodologique, en particulier au moyen d'exemples. Il semble nécessaire de disposer d'aide (et de garde-fous) à l'interprétation.

La documentation doit être de qualité, si possible en français (français 19 citations). Une aide en ligne, à l'écran est appréciée.

"Un point très important me paraît être la documentation statistique des logiciels"
"(...) documentation illustrée par des exemples"
"(....) documentation en français".

- Les méthodes statistiques sont souvent mentionnées (méthodes est cité 48 fois, méthode 10 fois et méthodes statistiques 15 fois). "Une bonne documentation sur les méthodes" est demandée, contenant plus d'information sur les méthodes employées - et éventuellement des précisions sur les algorithmes mis en oeuvre pour les implanter. Certains demandent la possibilité d'effectuer des analyses statistiques nombreuses et diversifiées, et de pouvoir, éventuellement, utiliser les méthodes récentes.

La qualité des techniques et des résultats est jugée importante.

"Le plus difficile n'est pas l'utilisation d'un logiciel mais (....) l'interprétation des résultats qu'il fournit."
"(....) erreurs énormes de la part des utilisateurs n'ayant pas forcément de connaissances en statistique"
"(....) intégration des méthodes récentes".

La convivialité (convivial, 21, convivialité 19, simple 14, facile 25, facilement 11, facilité 10 citations respectivement) est un thème très récurrent. On recherche un logiciel simple à utiliser, convivial, d'apprentissage rapide et facile, qui offre un traitement clair des erreurs.

On voudrait aussi pouvoir disposer d'une connexion facile avec d'autres logiciels et/ou avec des bases de données, et pouvoir gérer et manipuler les données facilement.

"Un bon logiciel est performant, convivial"
"Un bon logiciel en statistique est, pour moi, un logiciel simple d'utilisation"
"Un bon logiciel devrait être spécialisé, mais devrait communiquer très facilement avec d'autres logiciels spécialisés"

Des thèmes plus secondaires, mais néanmoins d'une fréquence non négligeable sont:

-La présentation soignée des sorties (sorties 28 citations), en particulier la possibilité de sorties graphiques (sorties graphiques 10 citations), est recherchée.

-La possibilité d'introduire ses propres routines dans le logiciel, et de programmer des calculs complémentaires et/ou des enchaînements spécifiques (programmation 23, programmes 11 citations respectivement).

-La problématique logiciel complet / logiciel spécialisé est évoquée.

"(...) graphiques de bonne qualité"
"(...) avec facilité de programmation"
"(...) possibilité de programmation de calculs complémentaires"
"(...) peut-être un logiciel qui ferait tout, ou alors de nombreuses interfaces"
"Encore faut-il distinguer entre les logiciels qui font "tout" (ou presque) et les logiciels spécifiques"

Les réponses aux questions fermées du logiciel permettent de mettre en relation les caractéristiques des utilisateurs et les mots employés dans les réponses. Pour cela, on regroupe les réponses des utilisateurs en fonction de leur profil personnel d'activité (Conseil et études, Enseignement et formation, Recherche en statistique, Recherche dans un autre domaine - choix multiple possible), de leur formation en statistique (diplôme spécifique en statistique ou non), du nombre d'années d'utilisation des méthodes statistiques et du nombre d'années d'utilisation d'au moins un logiciel de statistique (inférieurs ou supérieurs à 5 ans). Une analyse de correspondances multiples suivie d'une classification des individus à partir de leurs coordonnées factorielles permet d'obtenir 6

classes ou situations-types [20]: 3 classes de statisticiens, 2 classes d'utilisateurs non statisticiens, et une classe de non-réponses.

Classe 1(19% des utilisateurs): statisticiens jeunes, activité en conseil et études (principalement marketing-pub, conseil-services mais aussi dans l'industrie pharmaceutique et la banque). Ce sont des utilisateurs jeunes, avec une expérience assez courte dans l'emploi des logiciels et des méthodes. Formation longue ou courte en statistique.

Classe 2 (17%): statisticiens, dans le secteur conseil et études, plus expérimentés que les membres de la classe 1. Ces utilisateurs travaillent souvent dans l'industrie pharmaceutique ou en médecine. Davantage de diplômes longs en statistique que dans la classe précédente.

Classe 3 (18%): chercheurs en statistique et/ou enseignants, les membres de cette classe ont un diplôme spécifique long. Ils ont, en général, une très bonne expérience des méthodes et des logiciels de statistique.

Classe 4 (6%): cette classe est caractérisée par les non-réponses.

Classe 5 (15%): utilisateurs non statisticiens, peu expérimentés quant aux méthodes et logiciels statistiques. Ces utilisateurs ont reçu une formation complémentaire en statistique, assez souvent courte.

Classe 6 (25%): utilisateurs non statisticiens, mais expérimentés et plus longuement formés à la statistique que les membres de la classe 5.

Les réponses individuelles sont regroupées selon l'appartenance des individus à l'une des 6 classes. La table lexicale agrégée, qui contient la fréquence d'emploi des mots dans chacune des classes, est soumise à l'analyse de correspondances simples. La classe 4 est considérée supplémentaire. Pour avoir une bonne représentation des 5 classes, il faut conserver les trois premiers axes factoriels (figures 1 et 2). Sur chacun des plans présentés ici, seules figurent les classes bien représentées sur ce plan.

Le premier axe oppose les deux classes de non statisticiens (classe 5 et 6) selon leur expérience et formation en statistique. C'est un axe "expérience", celle-ci étant surtout importante pour les non statisticiens. La projection des modalités des variables "Nombre d'années d'utilisation d'un logiciel de statistique" (LOG1: moins de 2 ans, LOG2: de 2 à 5 ans, LOG3: de 5 à 10 ans, LOG4: plus de 10 ans) et "Nombre d'années d'utilisation des méthodes statistiques" (MET1: moins de 2 ans, MET2: de 2 à 5 ans, MET3: de 5 à 10 ans, MET4: plus de 10 ans) confirme la progression du vocabulaire selon l'expérience.

A l'extrême droite de l'axe, on demande de l'aide, aide à l'écran, aide à l'interprétation, mais aussi un appui quant à l'aspect statistique, en particulier dans la documentation au moyen d'exemples. "Bon logiciel statistique - convivial, très appliqué, explicite, bon détecteur d'erreur, bonne documentation".

A l'autre extrême de ce premier axe, on se préoccupe davantage de la qualité des méthodes et des facilités pour la manipulation des données. "Un bon logiciel de stat doit d'abord être sûr sur le plan des lois et bases statistiques utilisées. Il doit indiquer les raisonnements ou lois utilisés pour chaque cas d'analyse. Il doit permettre de tenir compte des effectifs déséquilibrés (...). Il doit pouvoir recevoir les données issues de tableurs "classiques" (...) sans manipulations trop lourdes.".

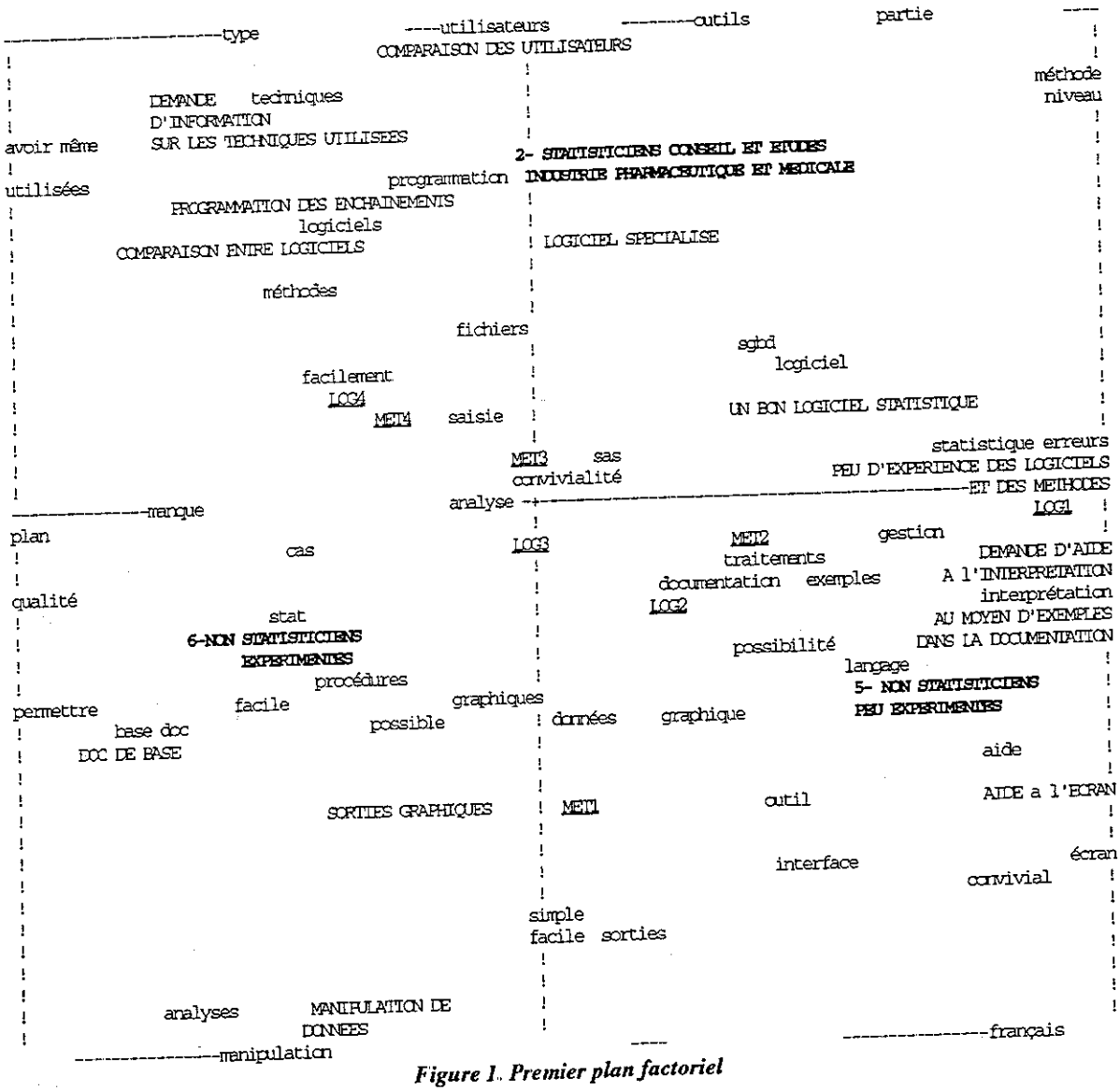


Figure 1. Premier plan factoriel

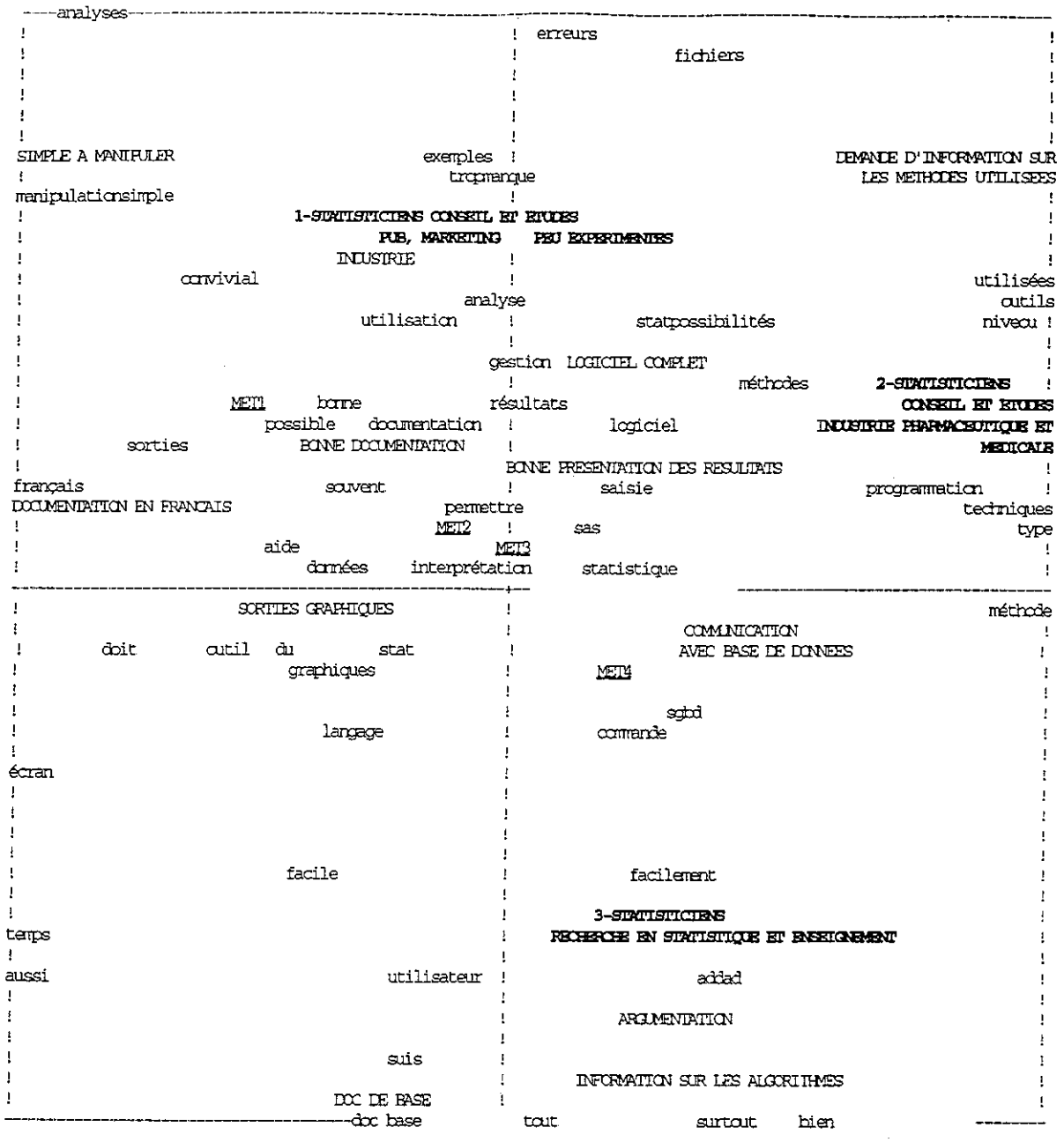


Figure 2. Deuxième plan factoriel

Le deuxième axe oppose la classe 2 aux classes 5 et 6, oppose les statisticiens expérimentés de l'industrie aux utilisateurs de la statistique dans leurs recherches. Les statisticiens de la classe 2 comparent les logiciels et les utilisateurs. Les plus expérimentés réclament une bonne information sur les techniques utilisées. " (....) Il y a des logiciels généraux et des logiciels spécialisés. Un logiciel peut être bien s'il est spécialisé - il ne peut pas y avoir de logiciel général, universel, qui soit universellement bien! Un bon logiciel devrait être spécialisé mais communiquer très facilement avec les autres logiciels spécialisés.".

Le troisième axe est un axe dédié aux statisticiens. Il oppose les statisticiens de l'industrie (classe 1 et 2) aux statisticiens-chercheurs en statistique (souvent enseignants), et de façon très marquée les statisticiens peu expérimentés aux statisticiens-chercheurs en statistique. Les premiers demandent énormément de choses (logiciel complet, de manipulation simple, permettant une bonne manipulation de fichiers....). "Logiciel complet, rigoureux, documentation claire, conviviale".

Les seconds, statisticiens expérimentés, préoccupés par leurs propres recherches, ne se distinguent pratiquement que par leur argumentation ("parce que", "surtout") : "(....) méthodes statistiques plus complexes", "intervention possible à tout moment pour modifier et surtout voir ce qui se passe".

Il est intéressant de noter qu'au fur et à mesure qu'augmente le niveau de formation en statistique et l'expérience dans l'utilisation des logiciels, la demande est plus focalisée sur les méthodes statistiques et sur leur implantation. Le domaine d'activité -et les problèmes abordés- font privilégier différents aspects des logiciels.

La lecture des réponses les plus caractéristiques de chacune des classes d'utilisateurs complète l'information fournie par l'analyse des correspondances. Les réponses sont ordonnées en fonction des valeurs-test associées aux mots qu'elles contiennent. Ces valeurs-test [20] mesurent le degré de caractérisation de chacun des mots pour chacune des classes. Les 4 réponses les plus caractéristiques de chacune des différentes classes d'utilisateurs sont présentées à la table 2.

CLASSE 1: STATISTICIENS PEU EXPERIMENTES, CONSEIL / INDUSTRIE:

1 logiciel complet, rigoureux.. documentation claire. convivial..

2 programme communiquant avec un tableur, en interface graphique, pourrait devenir le logiciel no 1 en matière de séries temporelles

3 - gestion des données facile et souple
- utilisation en interactif et en différé
- documentation claire et précise

CLASSE 2: STATISTICIENS EXPERIMENTES, CONSEIL / INDUSTRIE

1 je suis très satisfaite de pcsm.. ce logiciel demanderait à être amélioré par rapport à la "lourdeur" dans son utilisation, mais peut-être est-ce inhérent à ce type de logiciels sur pc ?

2 certains logiciels sont plutôt de type "pédagogique" d'autres plutôt du type "professionnels" un bon logiciel est un logiciel professionnel ayant des qualités pédagogiques.. il y a des logiciels généraux et des logiciels spécialisés.. un logiciel peut être bien s'il est spécialisé - il ne peut pas y avoir de logiciel général, universel, qui soit universellement bien ! un bon logiciel devrait être spécialisé mais communiquer très facilement avec les autres logiciels spécialisés

3 - il n'y a pas une seule remarque sur les techniques d'analyse exploitations dans ce questionnaire..
- un logiciel de statistique qui se voudrait "universel" me parait être que mauvais car il ne pourrait être adapté de la meilleure façon aux différentes techniques..

CLASSE 3: STATISTICIENS EXPERIMENTES, RECHERCHE ET ENSEIGNEMENT

1 bon logiciel statistique :
- meilleure convivialité
- documentation (en plusieurs langues) plus abordable
- assistance plus à portée de l'utilisateur

2 - un logiciel à 2 niveaux..

1er niveau : très accessible, proposant les méthodes statistiques usuelles permettant de traiter une enquête de a à z (le plus difficile dans une enquête étant de passer d'un logiciel à un autre)

2ème niveau : méthodes statistiques plus complexes mais surtout utilisées dans des cas particuliers.. le tout bien expliqué (à l'écran et ou à l'aide d'une doc) "j'aime bien le spss/pc+, peut être parce que c'est le premier que j'ai connu"

CLASSE 5: NON STATISTICIENS, PEU EXPERIMENTES

1 bon logiciel statistique - convivial, très appliqué, explicite, bon détecteur d'erreur, bonne documentation

2 - convivial
- rapide
- facile d'utilisation
- bonne documentation méthodologique
- interface avec d'autres logiciels
- prix modéré

3 sas avec en plus la gestion de données comme dans lotus et avec plus de facilités d'échanges de données

CLASSE 6: NON STATISTICIENS, EXPERIMENTES..

1 qualité techniques des traitements et convivialité

2 un bon logiciel de stat doit d'abord être sûr sur le plan des lois et bases statistiques utilisées.. il doit indiquer les raisonnements ou lois utilisés pour chaque cas d'analyse.. il doit permettre de tenir compte des effectifs déséquilibrés et le cas contraire l'indiquer.. il doit permettre de savoir à tout moment sur quel type de fichier on travaille et d'éviter d'en sortir sans sauvegarde (défaut sas). il doit pouvoir recevoir les données issus de tableurs "classiques" : excel, lotus, par ex.. sans manipulations trop lourdes. il doit permettre des analyses graphiques avec sorties faciles sur imprimantes..

3 1- facile à utiliser dès 10% de connaissance de ses possibilités de base et de son langage
2- documentation complète mais non pléthorique (contre-exemple : sas) organisée de façon thématique et non alphabétique (l'index doit suffire pour l'entrée alphabétique dans la doc.)

Table 2.. Réponses caractéristiques des classes d'utilisateurs

Annexe 1

liste des logiciels cités en partie 1 (Q17 et Q18)
triés par nombre de citations : logiciels utilisés

logiciel	Q17	Q18
SAS	261	23
Addad	95	12
SPAD-SPADN	90	13
Stat-ITCF	77	30
Statgraphics	60	14
S-Plus	47	16
SPSS	26	10
BMDP	24	9
StatView	22	4
Systat	21	8
Word	18	1
Eole 3	16	0
Excel	15	3
StatXact	13	2
Modulad	12	4
Quattro Pro	12	1
RS1-RSD	12	5
PCSM	11	4
Chadoc	10	3
SPADI	10	1
Biomeco	9	2
Amance	8	3
Data desk	8	0
Genstat	8	1
Lotus	8	3
Sicla	8	0
GLIM	6	6
NAG	6	3
dBase	5	1
EDA	5	3
Epiinfo	5	0
EyeLID	5	1
LEAS	5	5
Minitab	5	1
RATS	5	3
DS3	4	0
Harvard	4	0
Multiplan	4	2
NCSS	4	0
NLI Clotilde	4	1
Qalstat	4	0
Sphinx	4	1
TSP	4	2
Bidoui	3	1
Chart	3	3
Cricket Graph	3	0
CSS	3	2
JMP	3	0
Knowledge Seeker	3	0
Mandrake	3	1
Matlab	3	0

logiciel	Q17	Q18
PAC	3	0
Pest	3	1
Simca-PLS	3	0
StatWorks	3	1
Destin	2	0
ESC	2	0
Factoriel	2	0
Gauss	2	2
Geopack	2	0
IMSL	2	1
Intera	2	2
Lisa	2	0
Lisp-Stat	2	0
Nomad	2	0
PC-Tools	2	0
Power	2	0
Quantum	2	0
Question	2	0
Siphar	2	0
Super Anova	2	0
Troll	2	0
Windows	2	0
Anamul	1	0
Blue Pack	1	0
Canoco	1	0
CI	1	0
Climed	1	0
Cosi	1	0
CS-NL	1	0
CSIAI	1	0
D-Expert	1	0
DCF	1	0
Decisionnel Graphique	1	0
Desir	1	0
Disqual	1	0
Epilog	1	0
Eureka	1	0
Fastat	1	0
FOCA	1	0
Forecast	1	0
FoxPro	1	0
Freelance	1	0
Furelast	1	0
GEDE	1	0
HAUS	1	0
IBF	1	0
Item	1	0
Journal	1	0
K-Man	1	0
Leda	1	0
Limdep	1	0

logiciel	Q17	Q18
Lisrel	1	0
LTSM	1	0
LVPLS	1	1
MacDraw	1	0
MacPen	1	0
Mathematica	1	0
Modalisa	1	0
Nemrod	1	0
NLG	1	0
Norton	1	0
Open Access	1	0
Opep	1	0
Oracle	1	0
Osiris	1	0
Paradox	1	2
PASS	1	0
PC-Image	1	0
PC-Shell	1	0
PC-Xpress	1	0
Persee	1	0
Pez	1	0
Questidep	1	0
Radling	1	0
Ramis	1	0
Reflex	1	0
SCA	1	0
Select	1	0
Selescor	1	0
Sharam	1	0
Soritec	1	0
Stat-PC	1	0
Stata	1	0
Statlab	1	0
Symphony	1	1
TSA	1	1
TSE	1	0
Turbo C	1	0
Uniras	1	0
VAR3	1	1
WingZ	1	0
Works	1	1
Xstat	1	0
Ade	0	0
Adeco	0	1
Clustan	0	1
DCS	0	0
Flash	0	2
Ingres	0	0
Magic Plan	0	1
Tri-Deux	0	2

Annexe 2

tableaux de synthèse de l'opinion des utilisateurs
basés sur les réponses à la partie 2 du questionnaire

techniques statistiques utilisées												
logiciels	N	stat. descriptive	régression	plan expérience	analyse variance	série chrono	contrôle qualité	analyse factorielle	stat. non param.	analyse discrim.	modèle log-linéaire	autre
Addad	83	+	++		+	+	+	++++	+	+++		oui
BMDP	17	+++	++++		+++	++		+	+++	+++	+++	
Eole	17	++++	++++		++++	+	+	++++	+++	+++		oui
RS	12	++++	++++	+++	+++		++		++			oui
S-Plus	41	++++	++++	+	++	++	+	+++	++	+	+	oui
Statgraphics	49	++++	++++	++	++++	++	++	++	+++	++	+	oui
Stat-ITCF	59	++++	+++	++	++++	+	+	++++	++	+++	+	
Statview	17	++++	++++	++	++++	+		++	+++	+	+	
StatXact	10	+		+					++++			
Systat	16	++++	++++	++								oui
SAS	253	++++	++++	++	++++	+	+	+++	+++	+++	++	oui
SPAD-N	72	++	+	+	+	+		++++		++	+	oui
SPSS	17	++++	++++	+	++++	++		++	++++	+++	++	

tableau 1 **utilisation des techniques statistiques**

Matériels informatiques et Modes de travail									
logiciels	N	Mac	PC	station de travail	mini	site central	langage de commande	menu ou interface	macro ou programme
Addad	83	+	++++	+	+	++	++++	+	+
BMDP	17		++	+	++	++++	++++	+	+
Eole	17		++++			+	++	++++	+++
RS	12		++	++	++	+++	++++	+++	++
S-Plus	41		+	++++	+	+	++++	+	+++
Statgraphics	49	+	++++			+	+	++++	+
Stat-ITCF	59	+	++++		+	+	+	++++	+
Statview	17	++++					+	++++	
StatXact	10		++++				++	++++	
Systat	16	++	++++	+			++++	+++	+
SAS	253	+	+++	++	+	+++	++++	+	+++
SPAD-N	72	+	+++	+	+	++	++++	++	+
SPSS	17	+	++++	+	+	+++	++++	+++	+

tableau 2 **utilisation des logiciels par machine et modes de travail**

plus de 60%	++++
de 41 à 60%	+++
de 21 à 40%	++
de 1 à 40 %	+
non utilisé	*rien*

Opinion sur les logiciels									
logiciels	N	confiance résultats	confiance développement	réaction à l'erreur	gestion des données	édition des données	échange autres logiciels	échange autres SGBD	temps d'apprentissage
Addad	83	+++	+	-	-	-	-	-	- 1 mois
BMDP	17	+++	++	++	+	-	-	-	- 1 mois
Eole	17	+++	++	+++	+++	++	++	++	- 2 mois
RS	12	+	-	++	++	+++	++	-	- 1 mois
S-Plus	41	+++	++	+	+++	++	+	-	- 1 mois
Statgraphics	49	+	-	+	-	+	+	+	- 1 mois
Stat-ITCF	59	+	-	-	-	+	+	+	- 1 semaine
Statview	17	++	+	++	++	++	+++	+	- 1 semaine
StatXact	10	+++	+	+	-	+	-	-	- 1 semaine
Systat	16	+++	++	+	+	+	+	-	1 semaine
SAS	253	+++	+	+	++	+	++	+	2 mois
SPAD-N	72	+++	+	-	-	-	-	-	- 1 mois
SPSS	17	++	+	+	++	+	+	+	1 mois

tableau 3 importance d'utilisation des techniques statistiques

très confiant/très satisfaisant	+++
confiant/satisfaisant	++
ssez confiant/assez satisfaisant	+
peu confiant/peu satisfaisant	-

Services autour du logiciel									
logiciels	N	documentation	doc. utilisation	doc. statistique	aide à l'écran	assistance téléphonique	cours du fournisseur	groupe utilisateurs	particip. groupes utilisateurs
Addad	83	indispensable	++	++		+++	+++	oui	faible
BMDP	17	indispensable	+++	+++			+	oui	
Eole	17	recommandée	++	+	+	+++	+++	oui	faible
RS	12	recommandée	+++	++	++	+	+	oui	
S-Plus	41	indispensable	+++	++	++			oui	faible
Statgraphics	49	indispensable	++	+	+	++		oui	
Stat-ITCF	59	utile	++	++	+		+	oui	
Statview	17	utile	++	+					
StatXact	10	recommandée	+++	+++		+			
Systat	16	indispensable	+++	+++	++			oui	
SAS	253	indispensable	++	++	+	+	++	oui	forte
SPAD-N	72	indispensable	++	++		++	+++	oui	faible
SPSS	17	indispensable	+++	++	+	-		oui	

tableau 4 importance d'utilisation des techniques statistiques

très bon	+++
bon	++
moyen	+
faible	-
inexistant	*rien*

Références

1. Drouet d'Aubigny C., Drouet d'Aubigny G. *Le choix d'un logiciel de traitements statistiques des données sur micro-ordinateur : les logiciels généralistes* .Actes des 3 èmes journées logiciel d'AUMER. Vannes, pp.1-70. 1987.

2. Schwaller V. *Analyse comparative des logiciels statistiques du Centre de Calcul de Strasbourg du point de vue des méthodes d'analyse factorielle*. Rapport de stage de maîtrise. U.L.P. Centre de Calcul CNRS Strasbourg.

3. Poix J. *Analyse comparative des logiciels statistiques du Centre de Calcul de Strasbourg, du point de vue des méthodes de Classification des données*. Rapport de stage de maîtrise U.L.P. Centre de Calcul du CNRS Strasbourg.

4. Bigot. H. *Comparaison de logiciels*. XXIVe Journées ASU Bruxelles 1992. Groupe logiciel de l'ASU (synthèse des rapports (2) et (3)).

5. FRIDLUND A. *Statistiques et preuve par 9* - INFO-PC 47, Déc. 1988/Janv. 1989.

6. *le comparatif, les logiciels de statistiques* - l'ordinateur individuel n° 5, Mars 1990, pages 101 à 113.

7. Rapport STATILOGIE - Présentation de tableaux comparatifs.

8. THERNAU M.- *The American Statistician* - S-Plus vol. 44 n° 3 pp.239-241 1990

9. *Les logiciels de statistiques* - SVM Macintosh n° 11 juin/juillet 1990.

10. *Les statisticiels* - ICONES n° 21 - février/mars 1990.

11. *Liste des produits statistiques* - MacGuide Magazine aut-hiv. 1988.

12. *Statistics on the Macintosh* - BYTE juillet 1987.

13. *Le comparatif, comprendre et analyser les données*. Info-PC pp121- 151.Avril 93.

14. SAS Institute S.A. Domaine de Grégy - BP 5 - 77166 GREGY SUR YERRES.

15. SPAD.N, CISIA 1, avenue Herbillon 94160 SAINT MANDE.

16. SICLA; CISIA 1, avenue Herbillon 94160 SAINT MANDE.

17. EXCEL Microsoft, 18, Avenue du Québec, Courtaboeuf 1,91957 Les Ulis Cedex.

18 Bécue M. (1991) - *Análisis Estadística de Datos Textuales. Métodos de Análisis y Algoritmos*. CISIA, Paris.

19 Benzécri J.P. (1981) - *Pratique de l'Analyse des Données*, tome 3, Linguistique & Lexicologie, Dunod, Paris.

20 Lebart L., Salem A. (1989) - *Analyse Statistique des Données Textuelles*, Dunod, Paris.

21 Lebart L., Morineau A., Bécue M., (avec la coll. de P. Pleuvret et L. Haeusler) (1989) - *SPAD.T, Système Portable pour l'Analyse des Données Textuelles*. Manuel de Référence. CISIA, Paris.