

J. GOUDARD

Y. GRELET

**Les lycéens du second cycle : comparaison
entre filles et garçons**

Les cahiers de l'analyse des données, tome 2, n° 3 (1977),
p. 273-291

http://www.numdam.org/item?id=CAD_1977__2_3_273_0

© Les cahiers de l'analyse des données, Dunod, 1977, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

LES LYCÉENS DU SECOND CYCLE : COMPARAISON ENTRE FILLES ET GARÇONS [LYCÉENS]

d'après J. Goudard ⁽¹⁾ et Y. Grelet ⁽²⁾

1. L'enquête : Pédiatre travaillant à l'Hôpital Herold dans le service de diabétologie que dirige le Professeur H. Lestrade, le Docteur J. Goudard s'interroge sur le "mode de vie et les comportements de l'adolescent par rapport à l'hygiène"... avec en arrière-pensée d'autres questions du type : "Dans quelle mesure l'éducation à laquelle a été soumis, et est encore soumis l'adolescent, aide celui-ci à trouver des règles d'hygiène cohérentes...". Tout en affirmant que le rôle du médecin est d'aider des malades individuels, J. Goudard s'intéresse à l'ensemble des conditions de vie d'une classe d'âge. C'est pourquoi, aidé de quelques collaborateurs il a réalisé une ample et pertinente enquête, interrogeant par questionnaire quelque 1200 jeunes gens de la région parisienne, quant à leur rythme de vie ; activités scolaire et sportive ; loisirs ; alimentation ; vacances et santé ; ... ; sans oublier le logement ; la profession des parents ; etc ; au total 113 questions. Nous ne tenterons pas ici de rendre compte du dépouillement de l'enquête tel que l'ont publié en un volume de deux cents pages J. Goudard et Y. Grelet . Mais à la lecture de ce volume, les différences entre garçons et filles signalées par les auteurs, nous ont paru offrir matière à une étude statistique particulière ; et comme E. Yagolnitzer a récemment fait progresser "la comparaison de deux correspondances entre les mêmes ensembles" (cf ce cahier pp 251-264) c'est dans ce cadre que Y. Grelet a repris l'analyse d'une partie des résultats de l'enquête de J. Goudard. On publie donc ici ces analyses complémentaires, à la fois (§ 4) pour illustrer les recherches de E. Yagolnitzer, et pour faire connaître, ne fût-ce qu'en partie, les intéressantes données recueillies par J. Goudard (§ 3).

2. Les tableaux analysés : Les analyses présentées ici concernent seulement 676 des adolescents interrogés - 381 jeunes gens et 297 jeunes filles - tous élèves des classes de première ou seconde de cinq lycées : Lycée de Chelles, L. Condorcet, L. de Massy, L. Technico-Commercial de Paris 13°, L. de Saint-Ouen. A l'exception du célèbre Lycée Condorcet , réservé quasi exclusivement aux garçons, il s'agit d'établissements mixtes. Quant au questionnaire, les informations relatives au logement ayant été éliminées, on a retenu 33 questions principales (rythme de vie, loisir, alimentation, vacances) et 9 questions supplémentaires (nom du Lycée, catégorie socio-professionnelle du père, intérêt pour la politique, fatigue...) : soit au total 73 + 27 modalités de réponse. Par exemple à la question : Heure de lever des jours de classe, on a deux modalités de réponse LEO (après 6 h 30) et LEI (avant 6 h 30) ; à la question: Temps quotidien du travail scolaire on a 3 modalités (< 30 mn) ; TT2 (de 30 mn à 2 h) et TT3 (2 h 30 <) ; etc. Dans les résultats présentés ici, les individus n'apparaissent pas : seuls ont été analysés des tableaux de cooccurrences de modalités de réponse (ou tableaux de Burt ; cf e.g. [Bin Mult.] ; Ca TII n° 1). Afin de décrire ces tableaux fixons des notations :

(1) Directeur à l'Institut National de la Recherche Agronomique à Tunis.

(2) Ingénieur (C.N.R.S.).

QP : ensemble des 33 questions principales ;

QS : ensemble de 9 questions supplémentaires ;

JP : ensemble des 73 modalités de réponse aux questions de QP ;

JS : ensemble des 27 modalités de réponse aux questions de QS ;

$J, Q : J = JP \cup JS ; Q = QP \cup QS$

$KO_{J \times JP}$: tableau de cooccurrence construit d'après les réponses fournies par les 381 garçons interrogés : on a :

$\forall j \in J, \forall j' \in JP : KO(j, j') =$ nombre de garçons ayant adopté simultanément les modalités j et j' (e.g. $KO(LEO, TT1)$ = nombre de garçons ayant déclaré se lever après 6 h 30 et travailler moins de 30 mn).

Dans ce tableau, les lignes d'indice $j \in JS$, ne fourniront que des éléments supplémentaires.

$f_{J \times JP}^G$: tableau de fréquence (i.e. tableau pour lequel le total des éléments principaux du carré $f_{JP \times JP}^G$, est 1) obtenu en divisant $KO_{J \times JP}$ par le total KO des nombres inscrits dans le carré principal $JP \times JP$. A des irrégularités près, (omissions), les données sont sous forme disjonctive complète ; on a donc $KO \# 381$ (nombre de garçons) $\times$ 33 (nombre de questions) et l'on pose :

$$f^G(j, j') = KO(j, j') / KO$$

$kl_{J \times JP} ; f_{J \times JP}^F$: ces tableaux sont construits à partir des réponses des 297 jeunes filles, comme $KO_{J \times JP}$ et $f_{J \times JP}^G$, l'ont été à partir des réponses des 381 garçons.

$f_{J \times JP}^T$: moyenne des tableaux f^G et f^T :

$$f^T(j, j') = (f^G(j, j') + f^F(j, j')) / 2 ;$$

le tableau f^T est le tableau de fréquence construit d'après les cooccurrences des 676 sujets, à ceci près qu'y est corrigée l'inégalité entre le nombre des jeunes gens et celui des jeunes filles.

$f_{J \times JP}^D$: demi-différence des tableaux f^G et f^T :

$$f^D(j, j') = (f^G(j, j') - f^T(j, j')) / 2 ;$$

on sait (cf [Comparaison]) que E. Yagolnitzer obtient une décomposition de f^D dans l'analyse d'un tableau de correspondance en quatre blocs.

Ceci posé, les tableaux dont nous décrivons l'analyse peuvent être schématisés comme suit :

Sur ces schémas, les éléments supplémentaires sont signalés par des hachures horizontales denses ; les tableaux f^G et f^F sont marqués de hachures obliques, respectivement ascendantes et descendantes de la gauche vers la droite ; le tableau f^T est laissé sans hachures obliques. Quant aux éléments, on a pour les trois blocs de lignes du tableau total utilisé les sigles T_j, G_j, F_j : (e.g. $TLEO, GLEO, FLEO$, pour le Lever après 6 h 30) selon que les fréquences sont calculées sur la population totale (dûment équilibrée, cf *supra*), la population des garçons ou celle des jeunes filles. Dans le tableau en quatre blocs (cf E. Yagolnitzer) on a noté G_j, F_j , ce qui dans l'exposé théorique (cf [Comparaison])

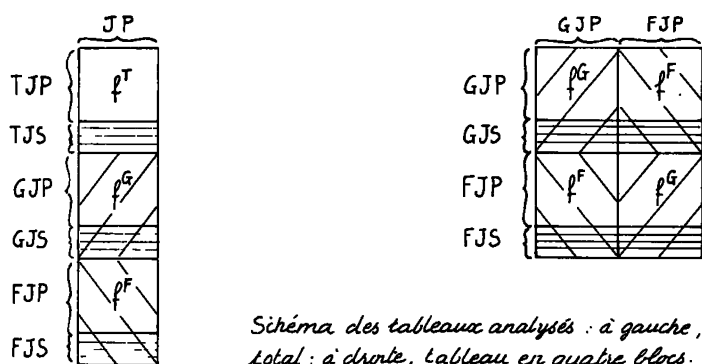


Schéma des tableaux analysés : à gauche, tableau total ; à droite, tableau en quatre blocs.

s'écrit j_0 , j_1 .

Quant à l'intelligence des données, l'analyse du tableau total avec ses nombreuses lignes supplémentaires, fournit l'essentiel de ce que nous exposerons ici (§ 3). Mais l'analyse du tableau en quatre blocs nous intéresse à un autre titre (§ 4). Outre des facteurs ω -symétriques (cf [Comparaisons] § 6) qui ne sont autres que ceux issus du tableau f^T , elle fournit deux facteurs ω -antisymétriques qui se rapportent à la décomposition de f^D , mais ne sont en fait qu'un artéfact, en ce qu'ils expriment la différence entre les marges des deux tableaux f^G et f^F (i.e. les inégalités entre les fréquences f^G et f^F d'une même modalité j , calculées respectivement sur la sous-population des jeunes gens et des jeunes filles). C'est là un exemple de plus du rôle dévolu selon nous aux modèles mathématiques en analyse des données : d'une particularité observée dans les résultats (ici dans le plan 1×2 issu du tableau en quatre blocs) on parvient à rendre raison en schématisant le tableau des données suivant une structure mathématique simple dont on ne pouvait *a priori* affirmer l'importance, ou à laquelle (c'est ici le cas) on n'avait même pas pensé.

3. L'analyse du tableau total : Nous représenterons d'abord le plan 1×2 issu de l'analyse, afin d'interpréter les deux premiers facteurs nettement détachés des suivants quant au taux d'inertie ; les facteurs 3, 4, 5 seront seulement interprétés d'après les principales modalités qui, se signalant par des contributions élevées, constituent et expliquent ces axes. Comme on pouvait l'attendre, l'enquête n'a pas épuisé la diversité de la vie des lycéens ! (pour ne rien dire des jeunes gens du même âge qui pour les trois quarts d'entre eux ne fréquentent pas les Lycées). Les axes 3, 4, 5 sont riches en évocation, mais se ressentent du caractère fragmentaire de l'échantillonnage. Sur tous les axes apparaissent entre jeunes gens et jeunes filles des différences d'inégale amplitude : comme on l'a annoncé, ces différences seront reprises au § 4 d'après le tableau en quatre blocs de E.Y..

N. B. Ici comme toujours dans l'analyse des questionnaires, il convient de distinguer entre le fait d'avoir répondu : "je fume moins de 10 cigarettes par jour" et le fait de fumer une telle quantité de cigarettes. Les réponses ne sont qu'une indication sur les actes qu'elles signifient ; mais elles sont en elles-mêmes un acte qui vaut d'être étudié : il est prudent d'admettre que c'est l'acte de réponse qui est l'objet véritable de l'analyse ; on ne le répètera pas à chaque ligne des interprétations qui suivent, mais il fallait le rappeler ici.

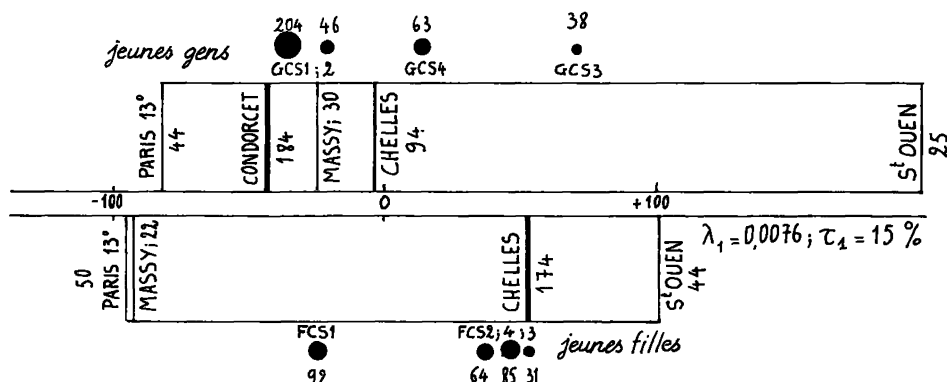
3.1 Le premier axe issu du tableau total : Le premier axe reçoit du côté négatif ses principales contributions : sommeil écourté, repas supprimé, tabac, alcool ; auxquels s'ajoute en éléments supplémentaires la politique : ces traits évoquent une sorte d'adolescent qu'on se représentera sans peine. Peut-être n'est-ce pas un véritable type, mais seulement une silhouette? Nous devons ici l'accepter dans son imprécision, comme nous acceptons du questionnaire les mots de cigarette, d'alcool et de sandwich qui désignent chacun (des produits de coût fort inégal). Les éléments supplémentaires permettent toutefois de situer cette silhouette d'après une fiche d'état civil sommaire.

Garçon ou fille? ; profession des parents ; Lycée fréquenté.

sigle	légende	$F_1(j)$	COR2	CTR	nG	nF
TCOO	couché le samedi avant minuit	120	662	43	261	213
TCO1	couché le samedi après minuit	-275	653	94	120	82
TSO1	temps de sommeil inférieur à 8 h.	-342	487	83	72	44
TRE1	supprime plus de 2f. par s. repas de midi	-226	513	68	112	99
TSO1	mange à midi d'un sand. plus de 2f. par s.	-306	323	49	41	42
TFO2	fume plus de 10 cigarettes par jour	-476	493	79	44	14
TOA4	sort très souvent hors du quartier	-226	374	48	85	66
TAL1	boissons alcoolisées entre les repas	-470	372	61	38	9
TTH1	café ou thé entre les repas	-167	366	41	129	104

Tableau des variables principales contribuant à l'axe 1 : pour chaque modalité j , on lit dans la colonne COR2, la contribution relative de l'axe à celle-ci, i.e. le quotient $F_1(j)^2 / \|Oj\|^2$ qui est aussi le cosinus carré de l'angle formé avec l'axe 1 par le rayon Oj joignant le point j au centre du nuage ; XTR est la contribution de j à λ_1 : $f_j F_1(j)^2 / \lambda_1$, comptée en millièmes (comme F_1 et COR) ; enfin n_G et n_F désignent respectivement le nombre des jeunes gens et des jeunes filles ayant adopté la modalité j .

On trouve du côté négatif de l'axe 1 (tabac, alcool etc : fumées!) plus de garçons que de jeunes filles ; et en moyenne, surtout des enfants de cadres supérieurs (CS1) : si l'on accepte comme borne l'origine de l'axe, on voit d'une part ($F_1 < 0$) : GCS1 et GCS2 avec FCS1 ; d'autre part ($F_1 > 0$) GCS3 et FCS2, 3, 4 ; GCS4 (garçons fils d'ouvriers) étant intermédiaire.



Situations sur l'axe 1: l'épaisseur des traits et la taille des points suggèrent les effectifs dont les valeurs exactes sont également indiquées; ces valeurs exactes ne seront pas répétées sur les autres axes.

Il serait toutefois imprudent d'affirmer l'existence d'un seuil social ; entre CS2 (cadres moyens) et CS3 (employés) pour les garçons ; entre CS1 et CS2 (cadres supérieurs et moyens) pour les jeunes filles : car on a aussi Paris (Paris 13° et Condorcet), écartés sur l'axe 1 négatif or Condorcet va avec CS1 ; et l'examen des listages poursuivi jusqu'au 5° axe montrera que chacune des neuf sous-populations de jeunes gens ou de jeunes filles répartis par Lycée, a un type propre ; en sorte que l'effectif important de l'échantillon ne garantit pas la stabilité des fréquences que l'on tenterait de calculer. Autant et plus qu'une 676 individus, il s'agit ici de 9 grappes, chacune certes hétérogène mais circonscrite socialement ; les catégories CS1 à 4 considérées ici sont de plus d'une inévitable hétérogénéité.

Sur le plan 1 x 2, chargé de nombreux éléments supplémentaires, on a signalé quelques triplets $\{T_j, G_j, F_j\}$; les trois points sont alignés, T_j entre G_j et F_j ; plus proche de celui de ceux-ci qui est le plus lourd (exactement au barycentre de G_j et F_j affectés des masses égales à la fréquence de la modalité j calculée respectivement chez les jeunes gens et les jeunes filles). On remarque par exemple que le point FQAl (ne sort jamais hors du quartier, jeunes filles) pèse trois fois plus que GQAl (idem garçons) ; et de plus s'écarte sur l'axe 1 positif (opposé à tabac) alors que GQAl est presque à l'origine ; toutefois comme cette modalité ne concerne que six garçons, sa place est peu sûre. Du côté négatif on note FAl1 (jeunes filles alcool) et FA01 (j. f. adhère à un mouvement politique) plus légers et plus écartés que les modalités correspondantes des garçons (GAl1 ; GA01) ; etc.

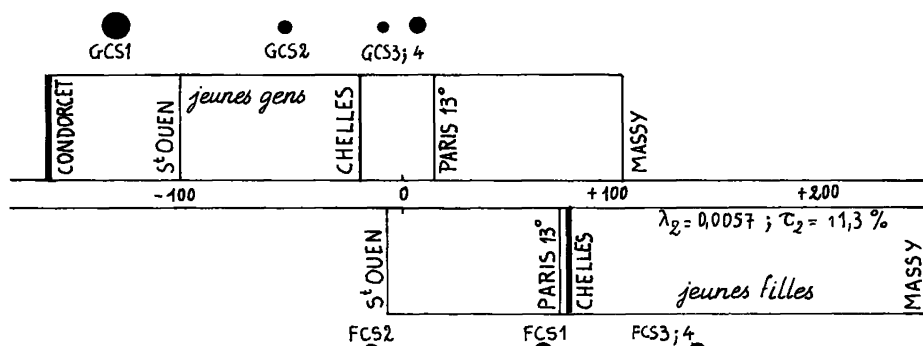
3.2 Le deuxième axe issu du tableau total : Dans le plan 1 x 2, le nuage des modalités est un triangle dont le sommet percute vers l'axe 1 négatif (tabac etc) et la base est inclinée sur l'axe 2, allant du quadrant ($F_1 > 0, F_2 > 0$) vers l'axe 2 négatif : avec là une constellation de jamais ; ici le lait généreux du sport, de la culture et des loisirs ! La différence entre les sexes et l'étalement des catégories socioprofessionnelles atteignent sur cet axe leur amplitude maxima.

Il serait trop schématique de dire que les uns ont tout tandis que les autres n'ont rien : les conceptions diverses de l'éducation se

sigle	légende	$F_2(j)$	COR2	CTR	nG	nF
TVC1	ne part jamais aux petites vacances	+397	506	105	30	48
TCI1	ne va jamais au cinéma	+161	317	49	116	110
TSP1	ne pratique pas le sport	+229	298	55	59	65
THS1	fait plus de 4 h de sport par semaine	-163	352	53	182	68

Tableau des variables principales contribuant à l'axe 2

deviennent ici autant qu'un accès inégal à diverses sources qui ne sont pas toutes limpides... Les diverses causes ne jouant pas toutes dans le même sens pour un sujet donné : la place des jeunes filles de St-Ouen, ou des jeunes filles issues de cadres moyens (FCS2) presque exactement à l'origine résulte sans doute des effets (opposés sur cet axe) d'une éducation à la fois stricte (jamais! $F_2 > 0$) et attentionnée (culture, sports, et loisirs : $F_2 < 0$).



Situations sur l'axe 2 : l'épaisseur des traits et la taille des points suggèrent les effectifs.

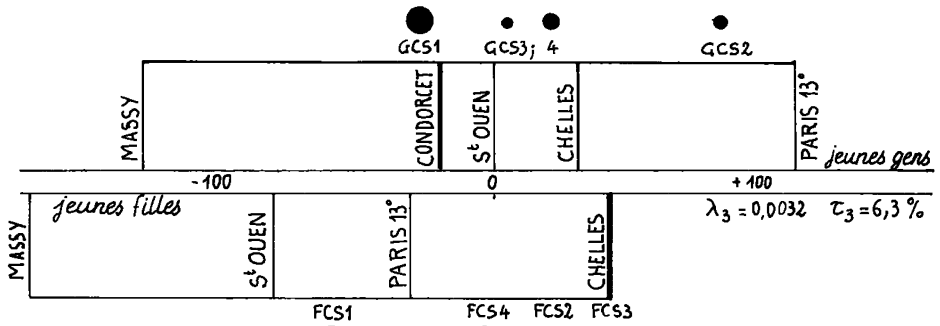
On notera que les quatre modalités relatives au sport : HS1 (sans sport) SPO (moins de 4 h. hebdomadaires), HSO (avec sport), SP1 (plus de 4 h.) forment une échelle qui s'ordonne sur l'axe 2, et réapparaîtra dans un autre contexte sur l'axe 5.

3.3 Le troisième axe issu du tableau total : Un petit nombre de modalités de l'ensemble TJP constitue cet axe et suffit à l'expliquer. Avec la partie négative de l'axe sont associés les longs trajets (la modalité TVO1 qui intéresse 26 jeunes gens et 34 jeunes filles reçoit du 3^e facteur une contribution relative de 43% ; et elle constitue 16,2% de l'inertie λ_3 sur cet axe et par conséquent le recours à

sigle	légende	$F_3(j)$	COR2	CTR	nG	nF
TVO1	effectue un trajet de plus d'une heure	-420	433	162	26	34
TLE1	se lève avant 6 h 30	-495	391	157	18	24
TMO2	utilise plusieurs moyens de transport	-256	207	63	36	28
TSO1	temps de sommeil inférieur à 8 h	-162	110	45	72	44
TBU1	budget insuffisant	+128	195	51	110	98
TFU1	fume moins de 10 cigarettes	+104	150	33	110	91

Tableau des variables principales contribuant à l'axe 3

plusieurs moyens de transport, un lever matinal, un temps de sommeil faible. Du côté positif de l'axe on remarque seulement : budget insuffisant, et fume moins de 10 cigarettes ; il est vraisemblable que ces modalités s'opposent aux longs trajets, parce que les lycéens qui se déplacent beaucoup reçoivent plus d'argent de poche surtout si, comme on le verra, ils appartiennent à des familles assez riches. En effet les éléments supplémentaires permettent de situer géographiquement et socialement les longs trajets. Sur le schéma des lycées et des CSP on voit du côté négatif (longs trajets) MASSY nettement écarté des autres lycées ; et CS1 moins écarté mais opposé aux autres catégories toutes situées du côté positif. Pour expliquer cela on considérera que les longs trajets peuvent avoir deux causes : d'une part l'isolement du lycée au sein d'une vaste zone qu'il est seul à desservir ; d'autre part une attraction particulière qui capte certains élèves les détournant d'un autre lycée plus proche de leur domicile : il n'est pas surprenant que les enfants de cadres supérieurs (CS1) soient le plus susceptibles de répondre à cette attraction et effectuent de longs trajets. Il faut noter encore que



Situations sur l'axe 3 l'épaisseur des traits et la taille des points suggèrent les effectifs.

les longs trajets concernent plutôt les jeunes filles que les garçons : cette différence atteint une amplitude maxima pour le lycée technico-commercial du 13^e arrondissement : on conçoit que par les carrières auxquelles il prépare ce lycée attire surtout les jeunes filles.

3.4 Le quatrième axe issu du tableau total : Comme l'axe 3, l'axe 4 s'explique par un petit nombre de modalités où la télévision prédomine s'opposant à la lecture. Du côté positif - exposition maxima à la télévision - l'axe reçoit ses plus fortes contributions ; du côté négatif on trouve outre les modalités opposées à celles remarquées sur le 4^e axe positif (mais, en vertu du principe du bras de levier, moins écartées que celles-ci parce que plus lourdes) le point TSO3 : temps de sommeil supérieur à 9 h. . L'examen des éléments supplémentaires montre

sigle	légende	F ₄ (j)	COR2	CTR	nG	nF
THT1	plus de 15 h. par sem. dev. la télévision	+269	219	88	42	25
TLI1	lit moins de 5 livres par an	+151	218	73	98	74
TTV1	regarde la télé pendant le repas	+104	227	58	155	131
TLE1	lever avant 6 h 30	+287	131	66	18	24
TSO1	prend plus de 2 f. par s. un sandw. p. repas	+183	116	52	41	42
TSO3	temps de sommeil supérieur à 9 h.	-119	83	33	66	57

dans l'ensemble un décalage entre garçons et jeunes filles compatible avec l'idée reçue que celles-ci lisent plus que ceux-là : le lycée technico-commercial du 13^e arrondissement se signale toutefois par une exposition maxima à la télévision tant chez les garçons que chez les jeunes

filles. Il est intéressant de noter que la différence entre garçons et filles de ce lycée révélée par l'axe 3 n'est pas visible ici : et le point TLY4 (lycée du 13^e, tous élèves) a avec l'axe 4 une corrélation²

(COR²) de .236, à peu près double de celle des points FLY4 et GLY4 (respectivement F et G de ce lycée) car la sommation des deux sous populations (F et G) en neutralisant les caractères opposés propres à celles-ci fait ressortir la contribution relative de l'axe 4, suivant le schéma dessiné ici. Quant aux catégories socioprofessionnelles, on observera

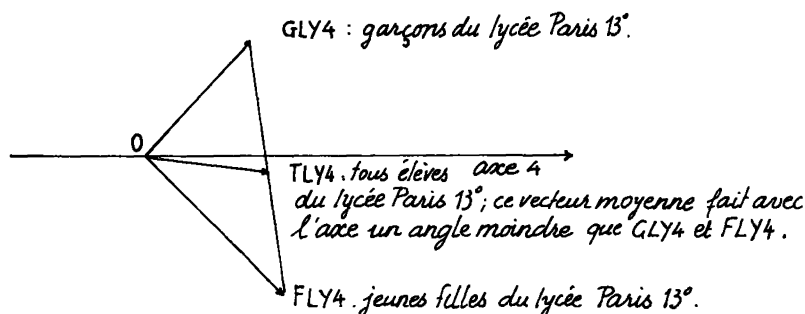
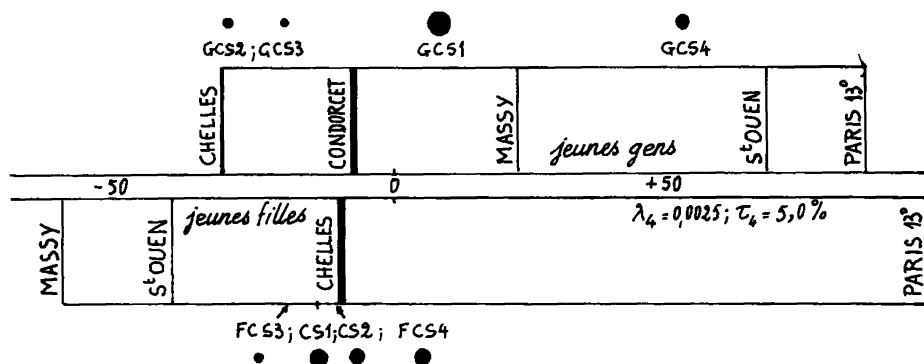


Figure : une modalité T_j peut être plus corréliée à un axe que ne le sont les modalités G_j et F_j .

que parmi les garçons F_4 atteint (télévision sur lecture) son maximum chez les fils d'ouvriers (CS4) et son minimum dans les catégories moyennes (CS2 = cadres moyens ; CS3 = employés) ; les cadres supérieurs (CS1) étant près de l'origine ; chez les jeunes filles, les différences semblent peu significatives. Parmi les autres éléments supplémentaires, GMJ1 (garçon ne va



Situations sur l'axe 4 : l'épaisseur des traits et la taille des points suggèrent les effectifs.

jamais à la maison des jeunes) est du côté de F_4 positif (télévision) avec un COR2 aussi élevé que celui des points THT1, TL11, TTV1 : tandis que FMJ1 (même attitude, jeunes filles) est proche de l'origine et sans corrélation avec l'axe 4. TP01 (ne s'intéresse pas du tout à la politique) est nettement du côté $F_4 > 0$ (télévision) avec COR2 = 230.

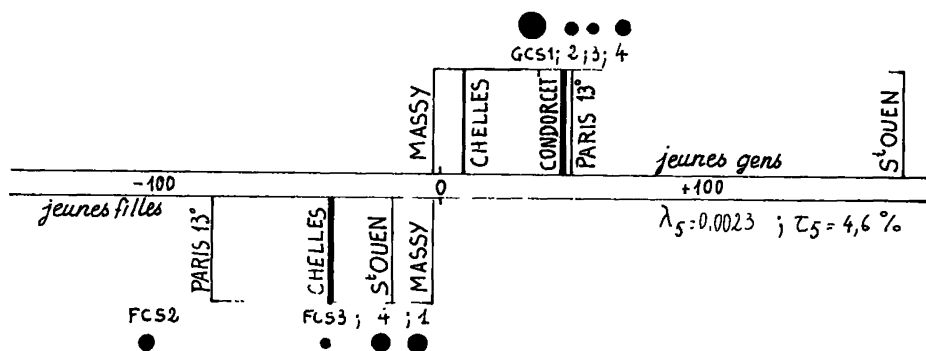
On pourra s'étonner que le rôle de la télévision ne se soit pas signalé avant le 4^e axe (nous le reverrons associé à d'autres traits sur le 5^e axe) : c'est que la télévision est un phénomène domestique ; or l'analyse a révélé d'abord non des caractères de la vie domestique

mais plutôt des axes de sortie : engagement, loisir, ou déplacement ... De plus la télévision est aujourd'hui si généralement répandue, qu'on ne peut la trouver corrélée à une opposition majeure. Cependant la place des élèves du lycée technico-commercial suggère que certains milieux, qu'une hiérarchie sommaire (en quatre CS échelonnées des cadres supérieurs aux ouvriers) ne parvient pas à circonscrire, mais que le choix d'une des professions auxquelles prépare ce lycée caractérise mieux, sont plus que les autres assidus à la télévision. La place de TPO1 suggère de plus que ces milieux s'intéressent fort peu à la politique.

3.5 Le cinquième axe issu du tableau total : Le tableau des points le plus liés à cet axe, découvre du côté positif un rythme de vie tendu, que l'examen des éléments supplémentaires permet de situer parmi les jeunes gens de Saint-Ouen. A l'opposé, des jeunes filles issues notamment

sigle	légende	F ₅ (j)	COR2	CTR	nG	nF
THS1	plus de 4 h. de sport par semaine	+134	237	88	182	68
TTT3	plus de 2,5 h. de travail scolaire p.jour	+175	170	70	61	51
TAL1	prend des boissons alcoolisées ent. repas	+314	166	89	38	9
TBI1	prend de la bière entre les repas	+159	134	57	85	32
THT1	plus de 15 h. /semaine devant la télév.	+195	115	50	42	25
TCA3	va à la campagne 4 ou 5 fois par mois	+153	106	46	73	27
TCI1	ne va pas au cinéma	+ 87	92	35	116	110
TAR1	n'a pas d'argent de poche	+251	88	43	15	18
TQA1	ne sort jamais hors du quartier	+284	80	40	6	17
TSP1	ne fait pas de sport	-192	209	95	59	65
TTT2	0,5 h. à 2 h. de travail scolaire par j.	- 59	273	36	274	221
THS0	fait moins de 3 h. de sport par semaine	- 74	238	49	199	227
TCA2	va à la campagne 2 ou 3 fois par mois	- 73	102	28	153	107

de cadres moyens (FCS2), et le plus nombreuses au lycée technico-commercial plusieurs fois cité déjà. Quelques éléments supplémentaires F₁,



Situations sur l'axe 5 : L'épaisseur des traits et la taille des points suggèrent les effectifs.

reçoivent de l'axe 5 négatif des contributions relatives particulièrement élevées : on a COR2(FTT2) = 436 (temps de travail entre 0,5 et 2h) COR2(FICO) = 392 (va au cinéma) ; COR2(FVCO) = 346 (part aux petites vacances) ; COR2(FFA3) = 303 (fatiguées de temps en temps). C'est assurément un autre rythme de vie ! L'échelle du sport (NON ; < 3 h. ; OUI : 4 h <) déjà sortie sur l'axe 2, apparaît aussi sur l'axe 5 ; mais avec un étalement un peu moindre, et dans un tout autre contexte : sur l'axe 2, sport va avec loisir, lait et s'oppose à jamais ; sur l'axe 5 sport va avec un travail scolaire prolongé, et s'accommode d'un certain nombre de jamais (ni cinéma, ni même sorties) ; la place du vin et du cidre avec le sport ne surprend pas, car dans la tradition française ces boissons sont quotidiennement celles du travailleur.

3.6 La diversité de l'adolescence : Après deux facteurs principaux nettement séparés, l'analyse du tableau total a signalé et situé les longs trajets entre le domicile et le lycée ; puis l'exposition maxima à la télévision. Le facteur de tension venu au cinquième rang signale d'abord un groupe sans doute assez peu nombreux de jeunes gens ; et ensuite à l'opposé, des jeunes filles au profil moins anguleux mais tout aussi nettement dessiné. Pour poursuivre l'analyse, il faudrait cerner ces deux groupes par une classification automatique. On ne le fera pas ici : il nous suffit d'avoir montré qu'au-delà des deux premiers axes (et même sur ceux-ci) on devine des faits particuliers : spécialisation d'un lycée ou seulement d'une de ses classes, homogénéité d'une banlieue ou d'un quartier ; dont la diversité à l'échelle de la seule région parisienne dépasse certainement ce que l'échantillon considéré ici peut embrasser. A la vérité en entreprenant une enquête sur "le mode de vie et les comportements de l'adolescent par rapport à l'hygiène", avec en arrière plan d'autres questions... J. Goudard savait qu'il jetait son filet sur toute la société française ; et qu'il ne capturerait pas tout ce qu'il pouvait convoiter. C'est un des mérites de son étude que de laisser entrevoir, après avoir permis de voir.

Cependant l'analyse a déjà révélé entre jeunes gens et jeunes filles des différences importantes : du point de vue des méthodes il nous reste à retrouver et peut-être à préciser ces différences en analysant le tableau en quatre blocs construit selon E. Yagolnitzer.

4. L'analyse du tableau en quatre blocs : L'exposé de cette analyse comprend trois parties : rappel des résultats de E. Yagolnitzer ([Comparaison]) ; examen du plan 1 x 2 issu de la présente enquête ; explication mathématique des particularités observées.

4.1 Rappel de propriétés de symétrie des facteurs : Les facteurs issus du tableau en quatre blocs se divisent en deux classes :

A) facteurs ω -symétriques : pour un tel facteur F_α l'on a :

$\forall j \in JP : F_\alpha(Gj) = F_\alpha(Fj)$; on notera donc simplement $F_\alpha(j) = F_\alpha(Gj) = F_\alpha(Fj)$ les valeurs d'un tel facteur. Les facteurs ω -symétriques ne sont autres que ceux issus du tableau F^T ; en particulier, si l'on note A l'ensemble de tous ces facteurs, la formule de reconstitution du tableau F^T s'écrit :

$$f_{j,j'}^T = f_j^T f_{j'}^T (1 + \sum \{\lambda_\alpha^{-1/2} F_\alpha(j) F_\alpha(j') | \alpha \in A\}) ;$$

on remarque dans cette formule que (cf [Comparaison] § 5 N.B. in fine) les facteurs F_α sont tous directs, i.e. ont même valeur en j qu'on considère cet élément comme ligne ou comme colonne ($F_\alpha(j) = G_\alpha(j)$) : cette particularité n'est vraie que parce que $f_{JP \times JP}^T$ est un tableau de Burt (on lit en effet dans [Bin. Mult.] § 2 in fine : $\lambda_\alpha' \varphi_\alpha'^J = \varphi_\alpha'^J \circ p_J^J$; i.e. par transition le facteur normalisé φ_α' , ou le facteur F_α' se reproduisent à un coefficient λ_α' près (positif parce qu'il n'est autre que la valeur propre issue de l'analyse du tableau de description logique $I \times J$)).

B) facteurs ω -antisymétriques : pour un tel facteur F_β l'on a :

$\forall j \in JP : F_\beta(Gj) = -F_\beta(Fj)$; on notera $F_\beta(j) = F_\beta(Gj) = -F_\beta(Fj)$ les valeurs d'un tel facteur. Les facteurs ω -antisymétriques fournissent une décomposition du tableau différence f^D au sens suivant :

$$f_{jj}^D = f_j^T f_j^T (\sum \{\lambda_\beta^{-1/2} F_\beta(j) F_\beta(j') \mid \epsilon(\beta) \mid \beta \in B\});$$

où on a noté B l'ensemble de tous les facteurs ω -antisymétriques ; dans cette formule (comme dans la précédente) λ désigne la valeur propre même obtenue en analysant le tableau en quatre blocs ; et $\epsilon(\beta)$ est la parité du facteur ; parité qui - nous en verrons un exemple - n'est pas nécessairement positive à la différence de celle des facteurs ω -symétriques. Il importe de noter que, à la différence des F_α , les F_β ne sont pas de moyenne nulle, mais ils satisfont entre eux aux propriétés d'orthonormalisation suivantes :

$$\forall \beta, \beta' \in B : \sum \{f_j^T F_\beta(j) F_{\beta'}(j) \mid j \in JP\} = \delta_{\beta\beta'} \lambda_\beta$$

4.2 Les résultats de l'analyse du tableau en quatre blocs : Sur le listage demandé par Y. Grelet, on lit les 21 premiers facteurs : tous à l'exception des deux premiers sont ω -symétriques. Précisons que l'on a :

$$\sum \{\lambda_\alpha \mid \alpha \in A\} = 0,0507 = \text{trace de } f^T ;$$

$$\sum \{\lambda_\beta \mid \beta \in B\} = 0,0287 ;$$

$$\sum \{\lambda_\gamma \mid \gamma \in A \cup B\} = 0,0794 = \text{trace du tableau en quatre blocs} ;$$

(nos listages ne donnent pas ces traces ; mais à partir de λ_1 et τ_1 , on calcule : $\text{trace} = 100\lambda_1/\tau_1$; on a ainsi obtenu la trace 0,0794 du tableau en quatre blocs ; la trace 0,0507 du tableau f^T objet de l'analyse du § 3 ; puis par différence la somme des λ_β relatifs aux valeurs propres antisymétriques ; cette somme n'est pas une trace ; la trace de f^D , très faible est donnée par $\sum \{\lambda_\beta \mid \epsilon(\beta) \mid \beta \in B\}$; il faut y tenir compte des parités $\epsilon(\beta)$).

Les premières valeurs propres et les taux sont :

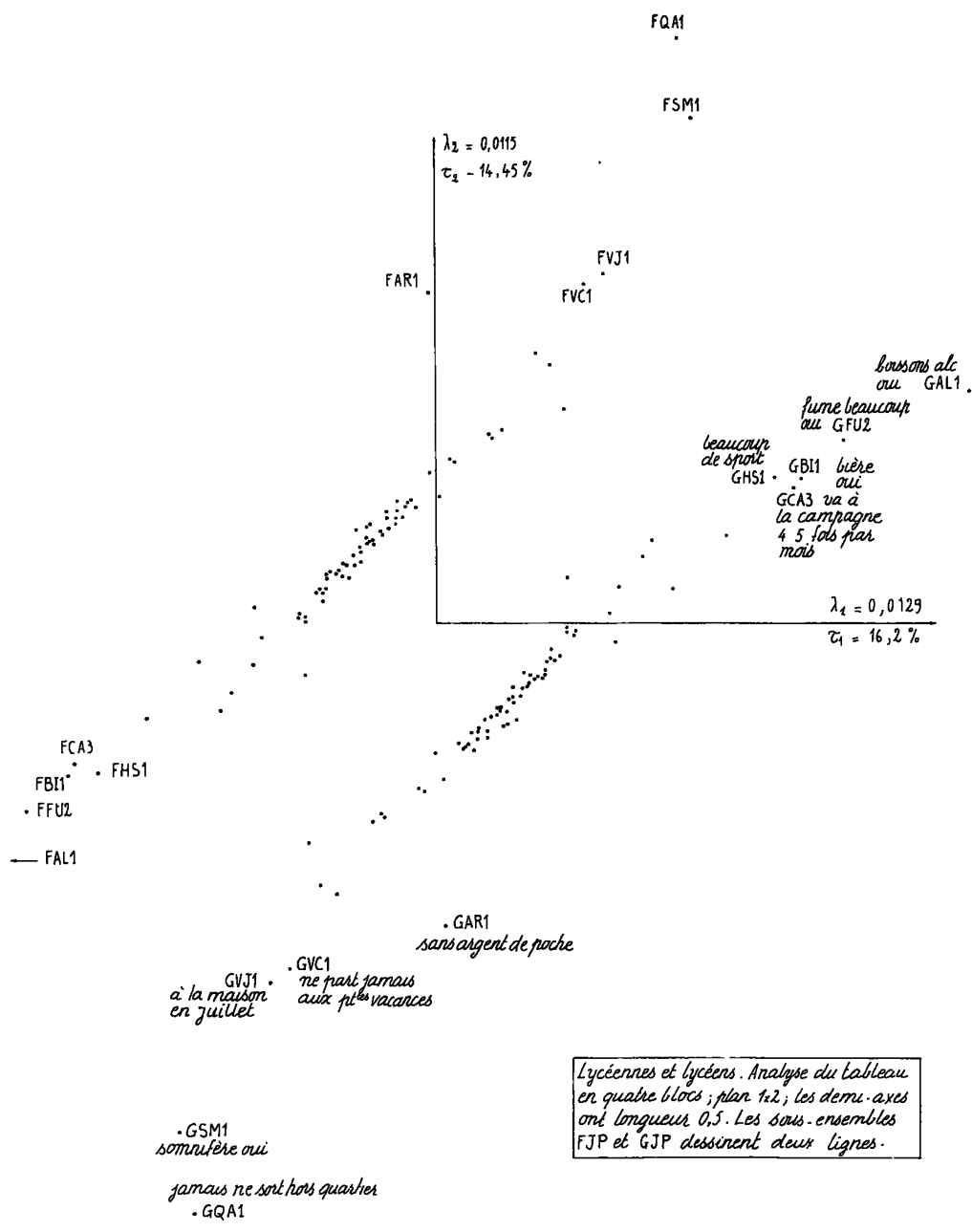
$$\lambda_1 = 0,0129 ; \lambda_2 = 0,0115 ; \lambda_3 = 0,0076 \dots ; \lambda_{21} = 0,00081$$

$$\tau_1 = 16,2\% ; \tau_2 = 14,45\% ; \tau_3 = 9,6\% \dots ; \tau_{21} = 1,02\%$$

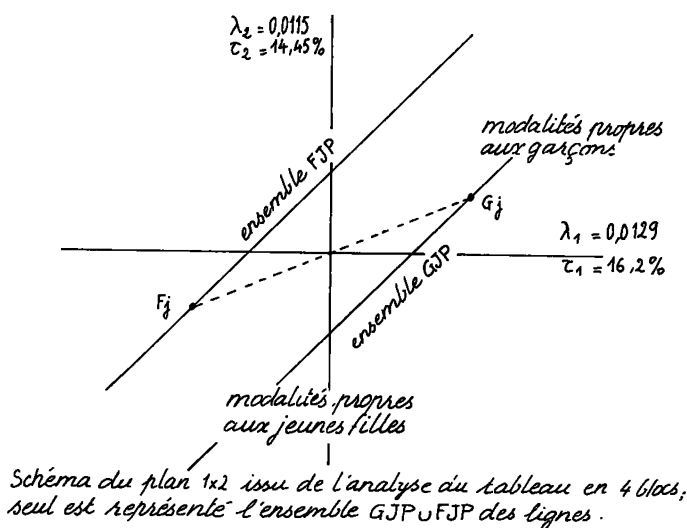
La somme $(\lambda_1 + \lambda_2) = 0,0244$ est à comparer au total 0,0287 des λ_β : on voit que les deux premiers facteurs épuisent presque la contribution des facteurs ω -antisymétriques ; de plus aucun λ_β au delà de λ_2 , ne dépasse 0,00081 valeur de la première v. p. d'un facteur non examiné sur le listage, et donc susceptible d'être ω -antisymétrique. Ainsi de l'analyse du tableau en quatre blocs il ne ressort que deux facteurs relatifs à f^D ; mais dans l'inertie ces deux facteurs ont un poids important : quels sont-ils ? Observons le plan 1×2 , issu de l'analyse du tableau en quatre blocs.

Ce plan est tout à fait insolite : les deux sous-ensembles GJP et FJP de l'ensemble des lignes du tableau en quatre blocs se rangent quasi rigoureusement sur deux droites parallèles à la bissectrice des axes. Ces droites sont symétriques l'une de l'autre par rapport à l'origine, car il résulte de l' ω -antisymétrie de F_1 et F_2 , que les points G_j et F_j afférents à une même modalité j sont des facteurs opposés.

$$F_\beta(G_j) = -F_\beta(F_j) = F_\beta(j).$$



Lycéennes et lycéens. Analyse du tableau en quatre blocs ; plan 1-2 ; les deux axes ont longueur 0,5. Les sous-ensembles FJP et GJP dessinent deux lignes.



Du point de vue de l'inertie on s'étonne que les axes principaux ne soient pas orientés l'un suivant les droites supports de GJP et FJP, l'autre perpendiculairement à celles-ci ; certes $\lambda_1 \neq \lambda_2$... les axes pourraient avoir tourné ; mais comment l'inertie dans la direction des sous-ensembles ne dépasse-t-elle pas nettement l'inertie transversale et ne s'impose-t-elle pas comme premier axe ?

La réponse à cette question est simple : aux extrémités des sous-ensembles FJP et GJP figurent des modalités j légères : bien que l'allongement des deux sous-ensembles dépasse nettement la distance entre les deux parallèles qui les portent l'inertie du nuage $GJP \cup FJP$ est à peu près la même dans toutes les directions du plan.

Il y a plus : ces modalités légères F_j et G_j situées aux extrémités des sous-ensembles, sont aussi celles pour lesquelles les différences relatives entre garçons et filles sont les plus grandes, i.e. pour lesquelles le quotient :

$$d^j = (f_j^D / f_j^T) = ((f_j^G - f_j^F) / (f_j^G + f_j^F))$$

atteint les plus fortes valeurs absolues ; et si l'on parcourt l'ensemble GJP d'une extrémité à l'autre dans le sens du 1^o axe, l'on va des valeurs négatives de d^j (modalités propres aux jeunes filles) aux valeurs positives de d^j (modalités plus fréquentes chez les garçons). On peut proposer une formule approchée telle que :

$$F_1(j) = 0,7 (d^j + 0,11) \quad ; \quad F_2(j) = 0,7 (d^j - 0,11) \quad ;$$

formule qui sera précisée et vérifiée au § 4.3. De plus l'orientation des ensembles GJP, FJP suivant les bissectrices des axes n'est pas fortuite ; bien que $\lambda_1 \neq \lambda_2$ il ne peut y avoir rotation des axes, mélange des facteurs en combinaisons linéaires ; parce que, tandis que F_1 est direct, F_2 est inverse (pour la parité ; i.e. le passage des lignes aux colonnes : $G_2(j) = -F_2(j)$ cf *supra* § 4.1).

Avant d'exposer le modèle mathématique, nous pouvons conclure : dans ([Comparaison] § 4.3) on a prévu que les quatre blocs devaient

avoir même masse (ici $f_{JP,JP}^G$ et $f_{JP,JP}^F$ ont tous deux masse totale 1) ce afin que f^D ne soit pas approximativement proportionnel à f^T : il apparaîtrait ici qu'une simple différence entre les marges des deux tableaux dont on fait la différence, peut dominer la structure de f^D obscurcissant ainsi les éventuelles différences entre les intercorrélations des variables dans G et T. On pourrait tenter d'éviter cela en amenant les tableaux f^G et f^F à avoir mêmes marges que f^T par une transformation itérative telle que la suivante, (où $X=G$ ou F ; et $f^{G(o)} = f^G$; $f^{F(o)} = f^F$) :

$$f_{jj'}^{X(n+1)} = f_{jj'}^{X(n)} \times (f_j^T \times f_{j'}^T) / (f_j^{X(n)} \times f_{j'}^{X(n)})$$

On peut comme l'ont fait J.G. & Y.G. dans leur rapport analyser séparément f^G et f^F et comparer les axes obtenus : mais de telles comparaisons ne sont pas si claires que celles qu'offre la mise en éléments supplémentaires de f^G et f^F sur f^T (cf § 3). Ici cependant on doit quant aux données formuler deux conclusions :

1) la différence entre f_{JJ}^G et f_{JJ}^F ne recèle pas d'autre fait structurel notable que la simple différence entre les lois marginales de ces deux tableaux. En effet, au-delà de λ_1 et λ_2 , il n'y a pas la différence f^D de valeur propre qui dépasse la 19^e v. p. issue de f^T . (Appliqué à f^G et f^T , l'algorithme de correction des marges conduirait donc à $f^{G\infty} \# f^{F\infty} \# f^T$).

2) les différences entre jeunes gens et jeunes filles considérées au § 3 ne résultent pas seulement d'une différence générale dans le genre de vie, l'éducation etc (différence dont cf § 1, la structure est simple ; en ce sens qu'une modification des marges, i.e. des fréquences de modalités j suffit à en rendre compte sans qu'il soit besoin de perturber plus que par cela les interdépendances de ces modalités) elle est aussi le fait de l'échantillonnage : présence du lycée Condorcet, dont les élèves sont des garçons qui se destinent aux études supérieures ; et d'un lycée technico-commercial, qui accueille jeunes gens et jeunes filles, mais que celles-ci apprécient sans doute davantage (cf § 3.3). A l'échelle de la région parisienne, bien d'autres groupes tout aussi caractéristiques devraient être compris dans l'enquête pour attribuer aux divers traits leur juste poids (cf § 3.6).

On ne prétendra donc pas fonder sur les données traitées ici une analyse comparative extrayant du tableau différence f^D des facteurs stables, au-delà des différences marginales apparues sur les deux premiers axes.

4.3 Le modèle mathématique : L'examen des résultats a montré que les facteurs F_1 et F_2 issus du tableau en quatre blocs sont ω -antisymétriques, donc figurent dans la reconstitution du tableau différence f^D ; et de plus sont liés à la différence des marges f^G et f^F plus précisément à la différence relative d^J . On cherchera donc dans f^D un terme principal fonction des d^J ; et ce terme sera identifié à la somme des termes de la reconstitution de f^D afférents au deux premiers facteurs ; d'où une expression de ces facteurs en fonction des différences relatives. Pour plus de généralité et aussi on le verra plus de simplicité nous supposons que les tableaux f^G, f^F, f^T, f^D sont non des tableaux de Burt symétriques mais des tableaux rectangulaires quelconques sur un

produit $I \times J$ puis nous reviendrons au cas de l'exemple.

Pour les tableaux f^G et f^F la formule de reconstitution s'écrit :

$$f_{ij}^G = f_i^G f_j^G (1 + \sum \{\lambda_{\alpha}^{-1/2} F_{\alpha}^G(i) G_{\alpha}^G(j) | \alpha \in A^G\});$$

$$f_{ij}^F = f_i^F f_j^F (1 + \sum \{\lambda_{\alpha}^{-1/2} F_{\alpha}^F(i) F_{\alpha}^F(j) | \alpha \in A^F\});$$

où les indices supérieurs G et F dont sont affectés les facteurs indiquent qu'il s'agit non des facteurs issus de l'analyse f_{IJ}^T mais des facteurs (d'ailleurs peu différents de ceux-ci) issus soit de f_{IJ}^G soit de f_{IJ}^F . L'ordre de grandeur des termes en λ_{α} est donné par $\lambda^{1/2}$ qui dans l'exemple vaut moins de 0,09 ; les termes d'ailleurs diffèrent peu de f^G à f^F en sorte qu'au premier ordre on a :

$$f_{ij}^D = (f_{ij}^G - f_{ij}^F)/2 \# (f_i^G f_j^G - f_i^F f_j^F)/2 = f_{ij}^{D(o)}.$$

Exprimons $f_{ij}^{D(o)}$ en fonction des différences relatives d^i, d^j et des marges de f^T ; il vient :

$$f_i^G = f_i^T (1 + ((f_i^G - f_i^T)/f_i^T)) = f_i^T (1 + ((f_i^G - f_i^F)/2 f_i^T)) ;$$

d'où :

$$f_i^G = f_i^T (1 + d^i) ; f_i^F = f_i^T (1 - d^i) ;$$

et de même pour les marges sur J :

$$f_j^G = f_j^T (1 + d^j) ; f_j^F = f_j^T (1 - d^j) ;$$

d'où pour le terme $f_{ij}^{D(o)}$ la formule :

$$\begin{aligned} f_{ij}^{D(o)} &= f_i^T f_j^T ((1 + d^i)(1 + d^j) - (1 - d^i)(1 - d^j)) \\ &= f_i^T f_j^T (d^i + d^j). \end{aligned}$$

C'est cela qu'on identifie à la somme des deux premiers termes de la formule de reconstitution de f_{IJ}^D en fonction des facteurs issus de son analyse ; on posera donc :

$$f_{ij}^{D(o)} = f_i^T f_j^T (\lambda_1 F_1(i) G_1(j) + \lambda_2^{-1/2} F_2(i) G_2(j)) ;$$

$$F_1(i) = d^i ; G_1(j) = \text{cte} = (\text{var}(d^I))^{1/2} = \lambda_1^{1/2} ;$$

$$F_2(i) = \text{cte} = (\text{var}(d^J))^{1/2} = \lambda_2^{1/2} ; G_2(j) = d^j ;$$

où λ_1, λ_2 sont les variances des fonctions d^I, d^J lesquelles sont comme on le vérifie des fonctions centrées :

$$\begin{aligned} \sum \{f_i^T d^i | i \in I\} &= \sum \{f_i^T (f_i^G - f_i^F)/(2 f_i^T)\} \\ &= \sum \{(f_i^G - f_i^F) | i \in I\} = 0 ; \end{aligned}$$

(car $\Sigma\{f_i^G\} = \Sigma\{f_i^F\} = 1$; et de même pour d^J ; les variances étant :

$$\Sigma\{f_i^T (d^i)^2 \mid i \in I\} = \lambda_1 ;$$

$$\Sigma\{f_j^T (d^j)^2 \mid j \in J\} = \lambda_2 ;$$

(le numérotage λ_1, λ_2 étant choisi sous l'hypothèse que la variance de d^I dépasse celle de d^J ; mais peu importe). On notera que tandis que sur I , F_1 a moyenne nulle, sur J G_1 est constant : cela ne surprendra pas si l'on se souvient que l'analyse de f^D avec des marges fournies par f^T n'aboutit pas nécessairement à des facteurs centrés.

Revenons maintenant au cas du tableau de Burt symétrique considéré dans l'exemple : $I = J = JP$; f^G et f^F symétriques. Alors il n'y a qu'une seule fonction d^{JP} donnant la différence relative des marges. A cette fonction sont bien associés deux facteurs, relatifs à une même valeur propre et qu'on peut noter, comme dans ce cas général $I \neq J$:

$$F'_1(j) = d^j ; \quad G'_1(j) = \sigma = \lambda^{1/2} = (\text{var}(d^J))^{1/2} ;$$

$$F'_2(j) = \sigma ; \quad G'_2(j) = d^j ;$$

cependant cette écriture n'est pas satisfaisante (et c'est pourquoi nous avons mis F'_1, G'_1 etc et non F_1, G_1) car les facteurs ainsi définis ne sont pas symétriques : il faut faire une rotation des axes :

$$F_1 = F'_1 \cos\theta + F'_2 \sin\theta ; \quad G_1 = G'_1 \cos\theta + G'_2 \sin\theta ;$$

$$F_2 = -F'_1 \sin\theta + F'_2 \cos\theta ; \quad G_2 = -G'_1 \sin\theta + G'_2 \cos\theta ,$$

de telle sorte que l'on ait $F_\alpha = \varepsilon G_\alpha$ avec $\varepsilon = +1$ (facteur pair) ou $\varepsilon = -1$ (facteur impair) ; avec une rotation de $\pi/4$ il vient :

$$F_1 = 2^{1/2} (d^j + \sigma) ; \quad G_1 = F_1 ; \text{ facteur direct } ;$$

$$F_2 = 2^{1/2} (-d^j + \sigma) ; \quad G_2 = -F_2 ; \text{ facteur inverse } ;$$

Il importe de remarquer que cette expression en somme et différence qui s'impose dans le cas d'une correspondance symétrique est moins immédiate que celle $F_1 = d^i$; $G_1 = \sigma_1$; $F_2 = \sigma_2$; $G_2 = d^j$ trouvée d'abord : c'est pourquoi nous annonçons au début de ce § que le cas $I \neq J$ est non seulement plus général mais plus simple que le cas symétrique.

Pour s'accorder avec le choix de signe fait par l'ordinateur et conservé au § 4.2 on notera plutôt $F_2 = 2^{1/2} (d^j - \sigma)$; d'où finalement l'expression des facteurs ; expression qui pour l'analyse du tableau en quatre blocs a valeur d'une approximation.

$$F_1(j) \# 2^{-1/2} (d^j + \sigma) ; \quad \lambda_1 \# \sigma^2 ;$$

$$F_2(j) \# 2^{-1/2} (d^j - \sigma) ; \quad \lambda_2 \# \sigma^2 ;$$

Pour apprécier cette approximation on pose :

$$\sigma' = ((\lambda_1 + \lambda_2)/2)^{1/2} = ((0,0129 + 0,0115)/2)^{1/2} = 0,110 ;$$

quant aux d_j ils sont fournis par les données même de l'enquête ; prenons quelques exemples :

$j = \text{LE1} : \text{lever avant 6 h 30} ;$

cette modalité a été adoptée par 18 garçons sur 381 et 24 jeunes filles sur 295 ; d'où :

$$d^{\text{LE1}} = ((18/381) - (24/295)) / ((18/381) + (24/295)) = - . 265$$

d'où pour F_1 et F_2 les valeurs approchées F^{app} qu'on comparera aux valeurs réelles F^{reel} lues sur le listage :

$$F_1^{\text{app}}(\text{LE1}) = - 0,110 ; F_1^{\text{reel}}(\text{LE1}) = - 0,114 ;$$

$$F_2^{\text{app}}(\text{LE1}) = - 0,265 ; F_2^{\text{reel}}(\text{LE1}) = - 0,268 ;$$

Seuls une dizaine de points j sur 73 offrent des écarts plus importants : il s'agit généralement de modalités peu fréquentes, pour lesquelles d_j est élevé et d'ailleurs en butte à des erreurs de fluctuation importantes ; ces points de vue se remarquent sur le plan 1 x 2 par leur mauvais alignement.

Il apparaît donc que les facteurs F_1 et F_2 issus du tableau en quatre blocs de Yagolnitzer, sont bien expliqués par un modèle où interviennent seules les différences relatives de fréquences de modalités de réponses entre jeunes gens et jeunes filles.