MINISTERE DE L'AMENAGEMENT, DE L'EQUIPEMENT, DU LOGEMENT
ET DU TOURISME

COMMISSARIAT GENERAL DU PLAN

PRINCIPES ET MODE D'EMPLOI DE LA TECHNIQUE D'EXPLOITATION EN ARBRE

utilisée pour analyser :

- les retards scolaires à l'école primaire dans le district scolaire de Denain
  (enquête réalisée par les services de M. HANTUTE)

- l'affectation des élèves entre les différentes sections des élèves des classes de 3ème et de préparation au C.A.P. dans l'arrondissement de Valenciennes
  (enquête réalisée par les centres d'information et d'orientation des districts de Valenciennes, Saint-Amand et Denain)

P. CALAME/JB
n° 450-1
E 791 b

Février 1973

CERAU Société Nouvelle
Département Socio-Economie

I N T R O D U C T I O N

Nous avons été amenés à analyser simultanément deux enquêtes.

La première, réalisée par M. HANTUTE, portait sur les élèves d'école primaire du district de Denain. Elle a été réalisée en 1969. M. HANTUTE s'interrogeait sur l'ampleur des retards scolaires dans son district et sur les moyens à mettre en oeuvre, notamment sur le plan pédagogique, pour les réduire. A cette fin, il fit procéder par les instituteurs du district à un relevé exhaustif des élèves de leurs classes. Pour chaque élève étaient mentionnés l'âge, la profession du père, la profession de la mère, le nombre d'enfants dans la famille, l'appréciation de l'instituteur sur le milieu familial de l'élève et, le cas échéant, la raison par laquelle l'instituteur expliquait le retard. Par exemple, les instituteurs avaient évoqué des raisons telles que : maladie nerveuse, difficultés de santé, origine étrangère ou troubles affectifs. Dans l'esprit de M. HANTUTE les différentes informations rassemblées ainsi sur chaque élève étaient susceptibles de permettre une meilleure compréhension des retards scolaires.

La seconde enquête a été réalisée par les Centres d'Information et d'Orientation Professionnelle de l'Arrondissement de Valenciennes. Elle a porté sur tous les élèves des classes de 3ème, sur les élèves en troisième année de C.A.P. et sur les élèves de préparation au B.E.P. et au C.E.P.. Ces élèves ont eu à remplir un questionnaire détaillé. Grâce à ces questionnaires, on sait, par exemple, pour chaque élève, la section à laquelle il appartient (section 1 préparant à l'enseignement supérieur long, section 2 préparant à l'enseignement supérieur court, classe dite de transition que nous appelerons section 3, classe de C.E.P. - section 4 -, classe de C.A.P. de mécanique et de traitement des métaux - sections 5 et 6 - et préparation au B.E.P. - section 7 -), le travail du père et de la mère, le nombre d'enfants dans la famille, la nationalité, les pays de naissance du père et de la mère, les projets que formulent les parents pour l'avenir de leurs enfants, etc... Avec toutes ces informations, on peut chercher à comprendre ce qui explique la présence d'un élève donné dans une section donnée ou ce qui explique son retard scolaire.

En somme, ces deux enquêtes posent des problèmes identiques : mettre en relation certaines caractéristiques des performances scolaires d'un élève (son retard scolaire dans la première enquête et la section à laquelle il appartient dans la seconde) avec des caractéristiques sociales familiales, voire psychologiques ou physiologiques de l'élève.

Nous avons utilisé pour les deux enquêtes une méthode spécialement mise au point pour les besoins de la cause, méthode que nous appellerons dans la suite "méthode d'exploitation en arbre".

Vous trouverez ci-joint une liasse de six pages montrant l'application de la méthode à l'analyse des retards scolaires dans le district de Denain, "toutes classes confondues" c'est-à-dire lorsque sont étudiés simultanément les élèves de toutes les classes depuis le cours préparatoire jusqu'au CM2.

Dans le cadre des deux enquêtes précitées nous avons réalisé sept liasses semblables. Les trois premières concernent les retards scolaires à l'école primaire : retard toutes classes confondues, retard au cours préparatoire, retard en CM2. Les quatre dernières sont relatives à l'affectation des élèves en 3ème du type I, II ou pratique et en classe préparatoire au C.A.P.. Ces liasses paraissent à première vue fort rébarbatives.

A chaque page correspond un niveau numéroté de 1 à 6. Ainsi la première page porte la mention niveau 1, la seconde porte la mention niveau 2, etc ... Chaque page est couverte de chiffres dont il importe de comprendre la signification précise de manière à ce que chacun puisse utiliser ces chiffres pour alimenter sa réflexion personnelle en s'appuyant sur sa propre expérience.

Ce chapître a pour objet d'expliquer "le mode d'emploi" c'est-à-dire les objectifs et les méthodes qui ont servi à mettre au point cette analyse.

§ 1 - MISE EN RELATION ET EXPLICATION

Dans la suite de ce texte nous puiserons les exemples concrets dans la liasse jointe.

Dans le district de Denain, en 1969, 11,5 % des enfants de cadres supérieurs étaient en retard à l'école. Par contre, les deux tiers des mineurs l'étaient. Spontanément, on en conclut que le retard scolaire, dépend, dans des proportions très importantes, du milieu social de l'enfant. Cette impression se confirme lorsque l'on compare le pourcentage d'enfants en retard dans les autres catégories sociales. On constate alors que ce pourcentage croît lorsque l'on descend dans la "hiérarchie sociale". 11,5 % de retard chez les enfants de cadres supérieurs, 20,9 % chez ceux de cadres moyens, 34,4 % chez ceux de commerçants, d'artisans ou d'employés, 39,2 % chez ceux d'ouvriers qualifiés, 49,7 % chez ceux d'ouvriers spécialisés, 66,6 %, enfin, chez ceux de mineurs.

Lorsqu'on énonce ces faits on met en relation la classe sociale des parents et le pourcentage d'enfants en retard et on raisonne de la manière suivante. Supposons que le retard scolaire ne soit pas lié à la classe sociale mais à des causes tout à fait différentes, génétiques par exemple, distribuées aléatoirement - c'est-à-dire au hasard - dans la population. Lorsque les effectifs d'élèves correspondant à chaque classe sociale augmentent, la loi des grands nombres veut que le pourcentage d'enfants en retard dans chaque classe sociale tende vers un même nombre. S'il n'en est pas ainsi c'est donc bien a contrario qu'il y a une relation entre retard et classe sociale. Dans le fond, lorsqu'on raisonne ainsi on fait bien l'hypothèse que la relation observée serait vraie dans d'autres lieux, c'est-à-dire qu'une analyse semblable menée à Lille ou à Nice donnerait des écarts semblables en ce qui concerne les retards scolaires entre les fils de cadres supérieurs ou d'ouvriers. Mais, mettre en relation deux variables, par exemple la classe sociale et le retard scolaire, ne veut pas dire pour autant "expliquer le retard scolaire". Par quels mécanismes se fait-il que la classe sociale influence les résultats scolaires toujours de la même manière ? On pourrait a priori imaginer un très grand nombre de mécanismes profondément différents. Prenons quelques exemples :

- On peut imaginer que les enfants des différentes classes sociales ne fréquentent pas les écoles maternelles avec la même fréquence ; si les enfants qui ne l'ont pas fréquentée se trouvent handicapés au moment de l'entrée en classe préparatoire, ce handicap et le retard scolaire qui s'en suivrait rendraient compte des différences observées.

- On sait par d'autres études que l'univers quotidien d'un enfant d'ouvrier, son horizon, ses préoccupations, son langage en un mot toute la culture véhiculée par le milieu familial, sont déconnectés de l'univers, du langage et des exigences de l'école. Cette déconnection n'existe pas dans le cas d'enfants de cadres moyens ou supérieurs.

- On peut penser que l'enfant de parents peu "cultivés" reçoit de la part de son milieu familial moins de soutien mais aussi moins d'encouragement dans son effort scolaire, la relation entre effort scolaire et réussite future, voire les perspectives de promotion sociale pouvant n'être pas perçues.

- Le pourcentage d'étrangers varie considérablement d'une classe sociale à l'autre. Si les étrangers sont fort handicapés dans leur apprentissage du français, ces handicaps peuvent en partie rendre compte des différences observées globalement entre classes sociales.

- De la même manière le pourcentage de familles nombreuses est plus élevé chez les mineurs, par exemple, que chez les employés. Qui dit famille nombreuse, dit moins d'argent à consacrer à chaque enfant, moins d'attentions portées à chacun d'eux, voire une attitude assez passive à l'égard de l'avenir. Dans ces conditions c'est parce que les familles peu nombreuses sont plus fréquentes chez les ouvriers et chez les mineurs que les retards scolaires s'y rencontreraient aussi plus fréquemment.

- Enfin, les conditions de logement et de nutrition plus mauvaises pourraient constituer une cause importante de retard dans certaines classes sociales.

Cette liste de mécanismes déjà fort longue est loin d'être exhaustive : on pourrait en imaginer bien d'autres, depuis une "explication génétique" les enfants de travailleurs manuels étant pour une raison ou pour une autre moins "doués" que les autres, jusqu'à des explications psychologiques, les instituteurs ayant une certaine tendance à plus pousser les enfants les plus proches de leur propre mode de vie et les enfants les mieux préparés aux exigences scolaires.

Il n'est pas gratuit de s'interroger sur les mécanismes effectifs. En effet, ils conditionnent étroitement les moyens d'action. S'il s'agit de réduire la fréquence de retard chez les enfants de mineurs, par exemple, il n'est pas indifférent de savoir par quels mécanismes précis ces retards se produisent et donc sur quel mécanisme il faut avant tout agir : scolarisation plus intensive en maternelle, mise au point de méthodes spéciales d'apprentissage de la langue, meilleure information des parents sur l'importance de l'effort scolaire, possibilité de relais pour les mères de famille, mise à la disposition des enfants de locaux adaptés pour faire les devoirs, etc...

Mais quels sont les moyens, une fois la relation entre classe sociale et retard scolaire clairement établie, de détecter de manière aussi précise que possible les mécanismes réellement à l'oeuvre ?

Deux méthodes complémentaires doivent être mises en oeuvre. Un exemple fera bien comprendre la première : si on veut vérifier que le handicap des enfants de mineurs face à l'école primaire résulte d'une moindre scolarisation en maternelle, on comparera deux groupes d'enfants de mineurs, l'un ayant fréquenté l'école maternelle, l'autre ne l'ayant pas fait. Si les premiers sont moins fréquemment en retard que les seconds, on peut en déduire que la scolarisation en maternelle réduit, pour cette catégorie sociale tout au moins, le handicap culturel. Si les pourcentages d'enfants en retard sont sensiblement les mêmes dans les deux groupes, on en concluera que l'absence de scolarisation en maternelle ne suffit pas à "expliquer" les handicaps des mineurs. La seconde méthode consiste à interroger des personnes, enseignants ou non, qui sont en contact quotidien avec les enfants, à partir des chiffres établis et des différences constatées. Comment, eux, expliquent-ils ces résultats ? en sont-ils étonnés ? pourquoi ? etc...

Loin de s'opposer, les deux méthodes se fécondent l'une l'autre : la technique chiffrée ne parvient jamais à nous "faire comprendre" comment les choses se passent, elle nous fait seulement "toucher du doigt" des relations aussi peu contestables que possible ; inversement, les "explications" des personnes compétentes appellent, autant que possible, des vérifications chiffrées. L'étude chiffrée ne permet jamais de démontrer de manière définitive la validité d'une explication ; elle met seulement sur la voie et permet surtout <u>d'éliminer les explications fausses.</u> Pour reprendre l'exemple précédent, l'existence d'une relation entre retard scolaire et scolarisation en maternelle chez les enfants de mineurs ne permettrait pas d'affirmer de manière définitive que c'est la faible scolarisation pré-scolaire du groupe qui explique ses échecs ultérieurs ; mais inversement, l'absence de relation permettrait <u>d'affirmer</u> que l'explication du handicap des enfants de mineurs doit être cherchée ailleurs.

§ 2 - VARIABLES LIEES - CONTROLE DES VARIABLES - HIERARCHIE DES VARIABLES -

Dans l'enquête de M. HANTUTE nous disposons, outre le retard scolaire, de cinq informations sur chaque enfant. Rappelons-les : profession du père, d'où l'on a dérivé la classe sociale en répartissant les différentes professions énoncées en onze catégories (C.S.P. : catégorie socio-professionnelle), profession de la mère d'où l'on a tiré un seul renseignement (la mère exerce-t-elle ou non une activité professionnelle), nombre d'enfants, jugement sur le milieu social, appréciation des causes de retard. A ces cinq informations on a ajouté une autre : d'après le nom de famille de l'élève on a caractérisé grossièrement son origine nationale en faisant trois catégories : les noms français ou belges, les noms d'origine polonaise, et les noms d'origine méditerranéenne ou maghrébine.

On aurait pu de la sorte faire une série de mises en relation : relation entre retard scolaire et classe sociale, relation entre retard scolaire et activité professionnelle de la mère, relation entre retard scolaire et milieu familial, etc... Mais toutes ces mises en relation successives ne nous auraient pas appris grand chose. En effet, toutes ces informations sont liées entre elles : pour ne prendre qu'un exemple, la classe sociale est liée simultanément à la nationalité, au taux d'activité de la femme, au nombre moyen d'enfants dans les familles, aux conditions de logement donc à la fréquence des accidents de santé, etc... L'analyse de l'enquête nous a montré - ce qui a priori n'était pas évident - que l'appréciation de l'instituteur sur le milieu familial était liée à la fois à la classe sociale et au nombre d'enfants.

A partir du moment où toutes ces variables sont liées, il y a gros à parier que l'observation d'une relation entre la classe sociale et le retard scolaire s'accompagne d'une relation identique entre le retard scolaire et les autres variables. Toutes ces relations ne nous apprennent pas grand chose de plus. En particulier il ne faudrait pas confondre les causes et les effets. Cette confusion entre causes et effets fait l'objet d'une littérature sociologique très abondante et ce n'est pas ici le lieu de la développer une fois de plus. Comment savoir si la relation entre nombre d'enfants dans la famille et retard scolaire n'est pas factice ? Pour s'en assurer il suffit de regarder si, à classe sociale donnée, on observe toujours cette relation, c'est-à-dire si, par exemple, pour les enfants d'ouvriers le pourcentage de retard est plus élevé dans les familles nombreuses que dans les familles peu nombreuses. En d'autres termes, pour progresser dans la compréhension des phénomènes, il faut introduire successivement chacune des variables selon un certain ordre et regarder à chaque fois si l'introduction d'une nouvelle variable, lorsque toutes les autres sont fixées, fait encore apparaître une relation entre cette variable et le retard scolaire. Ainsi, le nombre d'enfants a été la quatrième variable introduite, après la classe sociale, la nationalité et le travail de la mère. Nous savons que le nombre d'enfants est lié à ces trois variables à la fois, ce dont on peut s'assurer d'ailleurs dans les tableaux manifestant les relations entre ces variables. Contrôler les trois premières variables et introduire le nombre d'enfants consiste alors à savoir si dans un groupe d'élèves de même classe sociale, de même nationalité et de même activité de la mère, le pourcentage d'enfants en retard varie en fonction du nombre d'enfants dans la famille. Nous constatons par exemple que 30, 9 % des enfants d'ouvriers qualifiés français, dont la mère ne travaille pas, et appartenant à des familles de moins de trois enfants, sont en retard. Ce pourcentage s'élève à 47, 2 % pour les enfants d'ouvriers qualifiés français dont la femme ne travaille pas, appartenant à des familles de quatre ou cinq enfants et à 65,3 % pour les enfants d'ouvriers qualifiés français dont la femme ne travaille pas, appartenant à des familles de plus de cinq enfants. Dans ces conditions on peut dire que la relation entre retard scolaire et nombre d'enfants "tient" remarquablement, au moins en ce qui concerne ce groupe particulier, lorsque l'on fixe classe sociale, nationalité et travail de la mère.

Ce mode d'analyse consiste donc à contrôler systématiquement les variables introduites antérieurement, c'est-à-dire à ne faire des comparaisons de pourcentages qu'à l'intérieur des groupes déjà définis en fixant chacune de ces variables.

Dans ce type d'analyse le problème crucial consiste à fixer convenablement l'ordre d'introduction des variables. Cet ordre a une logique certaine. Par exemple, si des phénomènes psychologiques et sociaux font que les ouvriers ont plus d'enfants que les cadres moyens, on constate une relation entre classe sociale et nombre d'enfants. Cette relation ne peut pas être interprétée en sens inverse : on ne peut pas dire que c'est parce qu'un homme à beaucoup d'enfants qu'il devient ouvrier ; c'est parce qu'il est ouvrier qu'il a beaucoup d'enfants (en raison de mécanismes à expliciter naturellement). On doit donc introduire systématiquement les variables les plus importantes, les plus déterminantes en premier lieu.

Ainsi, la classe sociale est-elle introduite d'abord. Ensuite vient la nationalité car on sait que les taux d'activité féminins et le nombre d'enfants varient avec la nationalité, mais on ne peut pas soutenir qu'une personne, parce qu'elle a beaucoup d'enfants devient étrangère. C'est parce qu'elle est étrangère que (toujours par des mécanismes à préciser) elle a tendance à avoir beaucoup d'enfants.

Après la nationalité, on pouvait hésiter à introduire soit l'activité professionnelle de la mère, soit le nombre d'enfants. En effet, on ne voit aucune raison déterminante qui milite pour un ordre plutôt que pour un autre : d'une manière générale, taux d'activité de la mère et nombre d'enfants seraient plutôt deux effets d'une même cause : ouverture sur l'extérieur, désir de promotion sociale, information sur les méthodes contraceptives, désir "d'arriver", mode de vie normalisé par le groupe social, etc... Nous avons choisi d'introduire l'activité professionnelle de la mère en premier, mais l'examen des analyses en arbre montre que ce choix n'était probablement pas le bon et, tirant les conséquences de ce premier essai, nous avons inversé l'ordre pour l'analyse de la seconde enquête.

L'appréciation par l'instituteur du milieu familial de l'élève et les raisons apportées au retard furent introduites en dernier. Pourquoi ? Parce que, en procédant autrement, on eût risqué de s'enfermer dans un cercle vicieux : du retard scolaire et du jugement de l'instituteur sur le milieu familial, quelle est la cause, quel est l'effet ? Autant cette question était gratuite en ce qui concerne les quatre variables précédentes, autant elle s'impose en ce qui concerne celle là. On peut fort bien imaginer qu'un instituteur essayant de comprendre un retard ait tendance à porter un jugement sur le milieu familial au vu de ce retard et décrète ce milieu défavorable précisément parce qu'il y a retard.

Cette relation est encore plus évidente en ce qui concerne les causes du retard : l'instituteur par définition n'a mentionné ces causes que lorsqu'il y avait effectivement retard. Il était donc intéressant de voir si la relation, qui existe par construction, entre jugement sur le milieu social ou appréciation des causes du retard scolaire et fréquence du retard "tenait" encore lorsque l'on contrôlait classe sociale, nationalité, activité professionnelle de la mère et nombre d'enfants. Par contre, il eût été parfaitement illégitime d'introduire les deux dernières variables avant les quatre précédentes.

§ 3 - LES DIFFICULTES D'UNE ANALYSE EN ARBRE

La méthode ainsi exposée, d'une simplicité enfantine, serait entièrement satisfaisante si elle ne se heurtait à de graves problèmes pratiques.

En effet, lorsque l'on introduit la classe sociale, on découpe déjà la population en onze groupes correspondant à chaque catégorie socio-professionnelle (C.S.P.) introduite : agriculteurs, cadres supérieurs, cadres moyens, commerçants-artisans, employés, ouvriers qualifiés, ouvriers spécialisés et manoeuvres, mineurs, autres personnels de service, retraités et pensionnés, décédés enfin.

Lorsqu'ensuite, contrôlant la classe sociale, on introduit la nationalité en faisant comme on l'a dit trois types de nationalités (français et belges, polonais, autres nationalités) on découpe chacun des onze groupes initiaux en trois nouveaux sous-groupes. On aura donc à faire, pour chaque C.S.P., trois tableaux différents représentant les pourcentages d'enfants en retard pour chaque nationalité. Dès l'introduction de cette seconde variable, on devra donc manier trente trois pourcentages différents qu'il faudra comparer entre eux.

Si l'on introduit ensuite le travail de la mère, on va de nouveau diviser chaque groupe en deux, les femmes qui travaillent et les femmes qui ne travaillent pas, on aura donc soixante six groupes, etc... Et quand on en arrivera au sixième niveau on sera arrivé à constituer plus de 2 500 groupes, ce qui pose deux problèmes : d'une part, un problème d'impression et de lecture, que va-t-on faire de la comparaison entre 2 500 valeurs et, d'autre part, un problème encore plus fondamental d'effectif.

En effet, au terme de ce découpage de la population en 2 500 sous-groupes, on aura réparti les 9 565 élèves concernés par l'enquête dans 2 500 cases dont beaucoup seront finalement vides et dont beaucoup d'autres ne comporteront, en tout et pour tout, que un, deux, trois ou quatre enfants. Que signifie, dans ces conditions, la comparaison des pourcentages obtenus ? S'il n'y a qu'un enfant par groupe, ce pourcentage ne peut être que un ou zéro : ou bien l'enfant est en retard ou il ne l'est pas. Il n'y a pas de demi-mesure.

La méthode que nous avons mise au point vise à contourner ces deux obstacles sans se priver de l'avantage considérable que constitue le contrôle successif de chaque variable.

Elle est basée sur les deux principes suivants :

- il ne faut imprimer et analyser réellement que les groupes d'effectifs les plus importants ;

- il faut cependant construire des indices synthétiques caractérisant l'influence de chacune des variables introduites en tenant compte de l'ensemble des groupes constitués et pas seulement des plus nombreux.

Dans ce paragraphe nous allons essentiellement exposer la mise en oeuvre du premier principe et apprendre à lire les tableaux présentés dans chaque niveau des liasses. Les paragraphes 4 et 5 seront consacrés à la mise en oeuvre du second principe et au développement de la notion de "pouvoir explicatif" d'une variable.

3.1. - Caractéristiques des groupes les plus nombreux

Bien qu'il y ait 9 565 élèves à répartir dans plus de 2 500 cases, soit moins de quatre élèves par case en moyenne, l'effectif de certaines cases reste important. Par exemple, le groupe "enfants d'ouvriers qualifiés français dont la femme ne travaille pas, appartenant à des familles de moins de quatre enfants, pour lesquels l'instituteur a jugé que le milieu familial était favorable et n'a avancé aucune raison de retard" a un effectif de 647 élèves. Le même groupe, mais avec des enfants d'ouvriers spécialisés, compte 446 élèves. Il s'agit donc de groupes très importants qu'il est indispensable de sélectionner et d'analyser complètement.

On a donc décidé d'imprimer pour chaque niveau, c'est-à-dire après chaque introduction d'une nouvelle variable, les trente groupes d'effectifs les plus importants. Ce sont ces groupes qui figurent sur chaque page d'une liasse.

Ainsi, si nous reprenons la liasse "analyse des retards scolaires en école primaire, toutes classes confondues", on constate que sur la première page figurent onze colonnes, une par C.S.P.. En effet, à ce premier niveau, il n'y a que onze groupes. Mais dès que l'on aborde le niveau 2, c'est-à-dire la page suivante, on remarque que la page est divisée en deux : dans la partie haute de la page quinze groupes, chaque groupe étant désigné par une colonne de six chiffres au dessus des astérisques, et sur la deuxième partie de la page quinze autres groupes désignés de la même manière. Donc au total trente groupes.

Apprenons à lire la désignation de chacun des groupes : la colonne de six chiffres qui le désigne nous donne la modalité de chaque variable correspondante. Dans la colonne de gauche de la page on voit en effet six lignes : la première est notée CSP, première variable introduite ; la seconde NAT, c'est-à-dire nationalité ; la troisième TMERE, c'est-à-dire travail de la mère ; la quatrième NBENF, c'est-à-dire le nombre d'enfants ; la cinquième MILIEU, c'est-à-dire l'appréciation de l'instituteur sur le milieu familial ; et la dernière RAISONS, c'est-à-dire l'appréciation par l'instituteur des raisons du retard.

Prenons par exemple la première colonne de chiffres. Dans la ligne CSP on lit le chiffre 6, le groupe considéré correspondant correspond donc à des fils d'ouvriers qualifiés. En dessous de ce chiffre, à la deuxième ligne, on trouve le chiffre 1, ce qui signifie nationalité, modalité 1, c'est-à-dire Français et Belges. Les quatres chiffres en dessous sont des 0, ce qui signifie que les variables travail de la mère, nombre d'enfants, milieu familial et raisons du retard n'ont pas encore été étudiées.

En dessous de la ligne d'astérisques figurent huit lignes de nombres. Ces nombres nous donnent toutes les indications qui nous intéressent concernant le groupe considéré. Il est donc indispensable de bien comprendre la signification de chaque ligne.

La première ligne est désignée dans la colonne de gauche par le symbole NFI, qui se lit N Fi, ce qui signifie "effectif correspondant au groupe considéré". Ainsi, l'effectif pour le premier groupe décrit, "enfants d'ouvriers qualifiés français", est de 2 221. Pour le second groupe considéré, c'est-à-dire pour la deuxième colonne, "fils d'ouvriers spécialisés (CSP 7) de nationalité française (modalité 1)", la valeur de NFI, N Fi, est de 2 185, c'est-à-dire qu'il y a 2 185 élèves, enfants d'ouvriers spécialisés, à nom français. Dans la troisième colonne désignée par [7.3], c'est-à-dire "ouvriers spécialisés (CSP 7), noms d'origine méditerranéenne ou maghrébine (modalité 3)", l'effectif NFI est 574, etc... On notera que les groupes sont rangés par effectifs décroissants.

Ainsi, toujours en suivant la première partie de la page on remarque que le groupe le plus important a un effectif de 2 221, tandis que le quinzième groupe désigné par la colonne repère [10.3], "pensionnés à nom d'origine méditerranéenne ou maghrébine", l'effectif n'est plus que de 147.

Si on regarde enfin la seconde partie de la page on trouve dans la colonne de droite le trentième groupe par ordre d'importance : il n'a plus qu'un effectif de 9 enfants.

On peut donc dire que l'on a, pour le niveau 2, recensé tous les groupes de taille importante, ce qui est naturel puisque l'on a en fait imprimé trente groupes sur les trente trois au total.

Apprenons maintenant à lire les autres nombres qui figurent dans une colonne, en reprenant la première colonne en haut de la page, groupe désigné dans la colonne repère par [6.1]. En dessous de l'effectif NFI = 2 221, on rencontre une deuxième ligne désignée par TFI, qui se lit T Fi ; la valeur correspondante est de 0.760. Cette valeur veut dire que parmi les ouvriers qualifiés, 760 pour 1 000 ont un nom d'origine française ou belge. La valeur de TFI correspondant à la seconde colonne est 0.744, ce qui signifie que pour 1 000 ouvriers spécialisés (CSP = 7) 744 ont un nom d'origine française ou belge. Regardons maintenant la valeur de TFI pour la cinquième colonne. On lit TFI = 0.875. Le groupe correspondant est désigné par la colonne repère [3.1.0.0.0.0]. Il s'agit donc du groupe des cadres moyens (CSP = 3) ayant un nom d'origine française ou belge (modalité 1). Ceci veut donc dire que sur 1 000 cadres moyens 875 ont un nom d'origine française ou belge. Si l'on regarde la sixième colonne, on trouve TFI = 0.902, ce qui signifie que pour 1 000 employés (CSP = 5), 902 ont un nom d'origine française ou belge.

La lecture de la seconde ligne TFI nous donne donc des renseignements sur la manière dont un groupe se subdivise en sous-groupes lorsqu'on introduit une nouvelle variable.

Le rapide exercice auquel nous venons de procéder, confirme par exemple l'intuition de départ, à savoir qu'il y a plus de noms étrangers chez les ouvriers que chez les cadres moyens ou les employés.

Passons maintenant à la troisième ligne, désignée par UFI, qui se lit U Fi. Nous trouvons dans la première colonne la valeur 0.232, qui signifie que pour 1 000 élèves, 232 sont des enfants d'ouvriers qualifiés français, soit 23,2 % du total. Cette troisième ligne, moins utile que la seconde, nous renseigne sur l'importance relative du groupe analysé par rapport à l'effectif total des élèves. 0.232 est par définition égal à 2 221 (effectif du groupe) divisé par 9 565 (effectif total).

La quatrième ligne notée PFI (P Fi) est la plus importante. PFI désigne en effet le pourcentage d'enfants en retard dans le groupe considéré. Ainsi, dans la première colonne nous lisons PFI = 0.402, ce qui se lit : sur 1 000 élèves fils d'ouvriers qualifiés français, 402 sont en retard, soit un pourcentage de 40,2. Dans la deuxième colonne (ouvriers spécialisés français), PFI = 0.493 : le pourcentage d'enfants en retard est donc de 49,3.

Toujours dans la première partie de la page, la septième colonne est désignée par la colonne repère [6.3.0.0.0.0], c'est-à-dire : 6, ouvriers qualifiés, 3, noms d'origine méditerranéenne ou maghrébine. Dans cette colonne PFI = 0.428. Cela siginifie que 42,8 % des enfants d'ouvriers qualifiés méditerranéens ou maghrébins sont en retard, soit un pourcentage légèrement supérieur à celui des enfants d'ouvriers qualifiés français (40,2 % d'enfants en retard). La première réaction lorsqu'on lit un tableau de ce genre, c'est-à-dire lorsque l'on a introduit une nouvelle variable consiste à comparer les PFI entre les différents sous-groupes issus d'un même groupe. Pour simplifier le langage, on appelera ces sous-groupes "les fils d'un même père". Ainsi lorsque l'on introduit une nouvelle variable qui a trois modalités, par exemple le nombre d'enfants, on divise chaque groupe initial en trois sous-groupes, c'est-à-dire que l'on donne trois fils à chaque père. Poursuivant cette analogie démographique, on appellera "fils ainé" celui qui a le triste privilège d'avoir le plus fort pourcentage d'enfants en retard, et au contraire "benjamin" celui pour lequel le pourcentage d'enfants en retard est le plus faible. De la même manière on parlera du "fils le plus grand" pour désigner celui où l'effectif est le plus nombreux.

Les quatre dernières lignes de nombres correspondant à chaque colonne servent à faciliter la comparaison entre "fils du même père". Ceci s'avère d'autant plus nécessaire qu'il arrive qu'un fils plus grand que les autres soit représenté dans le tableau sans que les autres fils, plus petits, le soit. Cela arrivera chaque fois que les fils, correspondant à une des modalités de la variable introduite, sont beaucoup plus petits que les autres. Par exemple, le taux d'activité féminin étant souvent inférieur à 10 %, notre méthode qui consiste à sélectionner les trente plus grands groupes a tendance à éliminer systématiquement les groupes où la femme travaille.

La cinquième ligne de nombres est intitulée "RANG" ; elle désigne ce que nous appelerons le rang de naissance de chaque fils, c'est-à-dire de chaque groupe constitué après introduction de la dernière variable, parmi ses frères. Ainsi le "fils ainé", c'est-à-dire le groupe dans lequel le pourcentage d'enfants en retard est plus élevé que celui des autres "fils du même père", portera le rang 1. Le second portera le rang 2 et le troisième le rang 3. A titre d'exemple, reprenons dans le niveau 2, c'est-à-dire après introduction de la nationalité, trois fils du groupe "ouvriers qualifiés" (CSP 6). Ces trois fils figurent respectivement dans la colonne 1, pour le fils [6.1], dans la colonne 7 pour le fils [6.3] et à la colonne 11 pour le fils [6.2]. C'est le fils [6.3] qui porte le rang 1 car il a le pourcentage PFI le plus élevé, 42,8 %. Ensuite vient le fils [6.1] qui porte le rang 2 avec un pourcentage de retard PFI de 40,2 %, enfin le "benjamin", [6.2], porte le rang 3 avec un pourcentage de retard de 26,3 %. <u>On remarquera que l'ordre de classement est celui des pourcentages d'enfants en retard et non celui des effectifs</u>. Ainsi, le fils [6.1], le plus grand quant à l'effectif, ne porte que le rang 2 car le pourcentage de retards est plus faible que dans le groupe [6.3]. Concrètement, cet exemple signifie que parmi les enfants d'ouvriers qualifiés, ce sont les enfants à nom méditerranéen ou maghrébin qui sont le plus fréquemment en retard (42,8 %), les enfants à nom français ou belges viennent immédiatement ensuite (40,2 %), tandis que les enfants à nom polonais s'avèrent de bien meilleurs élèves (26,3 % seulement d'enfants en retard). Ainsi, la lecture du rang correspondant à une colonne donnée permet de se faire immédiatement une idée de la place occupée par un groupe étudié parmi ses frères.

Les trois autres lignes vont avoir des usages semblables.

La ligne notée "GDM" donne l'écart entre le pourcentage d'enfants en retard dans le groupe considéré et ce même pourcentage chez son frère ainé. Ainsi, dans le groupe [6.1], à la première colonne du tableau, on lit GDM = 0.026, ce qui signifie que l'écart entre le pourcentage PFI de ce groupe et le pourcentage PFI du "frère ainé" est de 26 pour 1 000. Effectivement nous avons vu que dans ce cas le "fils ainé" était [6.3], à la colonne 7, qui porte le rang 1. La valeur PFI pour ce groupe est 0.428 ; l'écart de PFI entre le groupe [6.3] et le groupe [6.1] est : 0.428 - 0.402, c'est-à-dire précisément 0.026.

Donc la seule lecture du nombre GDM de la colonne 1, nous permettait sans avoir à chercher les autres frères du groupe dans le tableau, de voir que l'écart entre PFI du groupe [6.1] et PFI pour l' "ainé" était un écart faible : en clair, les enfants d'ouvriers qualifiés à nom méditerranéen ou maghrébin sont certes légèrement plus en retard mais dans des proportions très faibles : 2,6 % d'écart seulement.

La ligne PDM nous donne, au contraire, l'écart entre la valeur PFI du groupe considéré et la valeur PFI de son "benjamin". Toujours pour le même groupe [6.1] nous avons une valeur PDM = 0.139. C'est-à-dire la différence entre PFI du groupe [6.1] (0.402) et PFI du groupe [6.2] (0.263). La lecture de cette valeur PDM permet donc, sans chercher dans le tableau, de savoir que le groupe [6.1] a un frère pour lequel les retardsscolaires étaient beaucoup moins fréquents. On notera que lorsque l'on a à faire à un groupe de rang 2 les valeurs GDM et PDM permettent de rétablir immédiatement la valeur de PFI pour l' "ainé" et la valeur de PFI pour le "benjamin". Au contraire lorsque l'on a un groupe de rang 1 ou de rang 3 on ne peut pas à partir des seuls indices GDM et PDM calculer la valeur de PFI pour le fils de rang 2. Regardons par exemple dans la troisième colonne le groupe [7.3]. Il porte le rang 1 avec un PFI égal à 0.547 (en clair, 54,7 % des enfants d'ouvriers spécialisés à nom méditerranéen ou maghrébin sont en retard). Comme il s'agit là du "fils ainé" la valeur GDM est égale à 0. La valeur PDM est égale à 0.171. Ce qui signifie que le "benjamin" des trois fils du groupe "ouvriers spécialisés" a un PFI égal à 0.547 - 0.171, soit 0.376. Cherchons dans le tableau ce "benjamin" ; nous le trouvons à la colonne [1.3] ; il s'agit de nouveau des enfants à nom polonais, groupe comportant un effectif de 178 élèves et portant naturellement le rang 3. Inversement si on lit dans la colonne 4 le groupe [10.1], c'est-à-dire groupe d'enfants dont les parents sont pensionnés ou non actifs et portent un nom français, ce groupe de 525 élèves a un PFI de 0.621 et porte le rang 3. Il s'agit donc du "benjamin"; dans ce cas, PDM = 0.0. Par contre, GDM = 0.066 ; ce qui signifie que l'"ainé des fils" du groupe 10 à un PFI égal à 0.621 + 0.066 = 0.687. Ce groupe figure effectivement à la quinzième colonne de notre tableau.

Enfin, la dernière ligne, notée "EFI" (E Fi) donne la différence des PFI entre le "fils ainé" et le "benjamin" des frères du groupe considéré. Cette valeur EFI est tout à fait importante pour comprendre dès une première lecture du tableau l'influence de la dernière variable introduite. En effet, si cette variable n'avait aucune influence sur les retards scolaires, les pourcentages d'enfants en retard correspondant aux différentes modalités de cette variable seraient égaux. Dans tous les groupe d'une certaine importance (de manière à éliminer les effets du hasard susceptibles de jouer dans de petits effectifs) la valeur EFI devrait être systématiquement voisine de 0. Au contraire, si la variable introduite a une influence importante, il va s'établir de grandes différences entre les PFI des "fils d'un même père" suivant la modalité de la variable. Par définition EFI est le même pour tous les "fils d'un même père".

Comparons EFI dans les différentes colonnes du tableau niveau 2. Pour le groupe 6 (1ère, 7ème et 11ème colonne), EFI = 0.165, c'est-à-dire que la différence des PFI de [6.2]et de [6.3] est de 16,5 %. Au contraire, chez les mineurs (CSP = 8), qui sont représentés dans la colonne 8 8.1 , la colonne 10 8.3 et la colonne 19 8.2 , EFI = 0.020. En clair, selon la nationalité, le pourcentage d'enfants en retard ne varie que de 2 % chez les mineurs. Ce qui signifie que dans cette classe sociale la nationalité n'a que peu d'influence sur les retards scolaires : il s'agit d'un groupe où de toute manière le pourcentage de retard est excessivement élevé (de 65 à 67 % d'enfants en retard). Pour interpréter EFI, on doit toujours prendre garde que les effectifs soient suffisants chez tous les "fils du même père". Regardons par exemple le cas des cadres supérieurs (CSP 2). Les trois fils figurent à la colonne 12 2.1 , qui a un effectif de 208 élèves, à la colonne 29 2.2 qui a un effectif de 9 élèves et à la colonne 30 qui a elle aussi un effectif de 9 élèves. Pour les cadres supérieurs la valeur EFI est très élevée : 0.556. Ceci tient au fait que le "benjamin", [2.2], a un PFI = 0.0 (aucun des 9 enfants de cadres supérieurs à nom d'origine polonaise est en retard) alors que le PFI du groupe [2.3] est de 0.556 (ce qui signifie que sur les 9 enfants de cadres supérieurs à nom d'origine méditerranéenne ou maghrébine 5 sont en retard). On ne peut conclure, de si petits effectifs, que la nationalité joue un rôle déterminant dans les résultats scolaires des enfants de cadres supérieurs.

3.2. - Définition des rangs pondérés

Nous venons de voir en détail la manière dont avaient été imprimés, à chaque niveau, c'est-à-dire après chaque introduction d'une nouvelle variable, les 30 groupes d'effectif le plus important. Mais cette sélection des plus grands groupes risque de faire perdre beaucoup de pertinence et de généralité à l'analyse. Au niveau 2, cela n'était pas grave. En effet, on avait imprimé 30 groupes sur les 33. Ces 30 groupes réunis constituaient un effectif de 9 550 élèves, c'est-à-dire que les trois groupes non imprimés, à eux tous, ne représentaient que 15 élèves. Par contre, il en va tout autrement lorsque le nombre de cases possibles augmente. Ainsi, au niveau 4, les 30 premiers groupes ne représentent plus à eux tous qu'un effectif de 7 708 élèves. En n'imprimant que 30 groupes on a "perdu" plus de 1 850 élèves. Il a donc fallu concevoir pour chaque niveau des données plus synthétiques, traduisant l'influence de la variable sur l'ensemble des élèves. Ces indices synthétiques ont été conçus de manière à répondre à deux questions fondamentales :

- la variable nouvellement introduite a-t-elle un "effet" important sur les retard scolaires ? C'est-à-dire "explique-t-elle" beaucoup les retards ?

- la variable nouvellement introduite a-t-elle toujours le même effet ? Ainsi, dans l'exemple précédent, nous avons constaté, en lisant en détail les tableaux au niveau 2, que les enfants d'origine polonaise réussissaient mieux que les autres tant chez les ouvriers qualifiés que chez les ouvriers spécialisés. De la même manière, on constatera que les enfants de famille peu nombreuse réussissent mieux dans ces mêmes groupes que les enfants de famille nombreuse (le lecteur, à titre d'exercice, le vérifiera sur le niveau 4). On se pose alors la question suivante : si on imprimait tous les groupes, constaterait-on une grande régularité dans le phénomène ou bien y aurait-il des groupes dans lesquels on observerait le phénomène inverse, les enfants polonais réussissant moins bien que les autres ou les enfants de famille nombreuse réussissant mieux que les autres, par exemple ?

La première question pose un problème de fond et nous consacrerons le paragraphe suivant à y répondre. Par contre, la seconde question est assez simple.

Pour caractériser la régularité de l'influence d'une variable, on raisonne de la manière suivante : numérotons par 1, 2 et 3, les différentes modalités d'une variable à introduire. Supposons que la modalité 1 corresponde, en général, à une influence défavorable sur les résultats scolaires. C'est-à-dire que les "fils de chaque père" portant la modalité 1 seront bien souvent les aînés, le pourcentage d'enfants en retard y étant maximum. A ce moment là, presque toujours, ils porteront le rang 1. Au contraire, si, d'un groupe à l'autre, la modalité 1 de la variable correspond à des effets contradictoires, les fils portant cette modalité 1 auront tantôt le rang 1, tantôt le rang 2, tantôt le rang 3. On a construit des indices synthétiques, notés K1, K2, K3, qui caractérisent le "rang pondéré" des modalités 1, 2 et 3 de la variable nouvellement introduite. S'il y a très peu d'inversions d'un groupe à l'autre, en ce qui concerne l'effet des modalités 1, 2 et 3, ou si ces inversions ne concernent que des groupes d'effectif très faible, ces rangs pondérés auront des valeurs voisines des nombres entiers 1, 2 et 3, l'ordre de ces nombres dépendant des effets des modalités 1, 2 et 3. Exemple : reprenons le niveau 2 sur lequel nous avons déjà travaillé. Sur la ligne en haut de la page nous lisons une série d'indices. Un premier indice, DH = 0.962, sera commenté ensuite. Nous lisons ensuite K1 = 2.038, c'est-à-dire une valeur très proche de l'entier 2 ; K2 = 2.807, c'est-à-dire une valeur assez voisine de 3 et enfin, K3 = 1.225, c'est-à-dire une valeur assez voisine de 1. En d'autres termes, le rang pondéré de la modalité 1 (nom d'origine française) est 2.038, etc... Ce qui signifie que les fils correspondant à la modalité 1 de la variable nationalité ont presque toujours le rang 2, c'est-à-dire s'intercalent presque toujours entre un "aîné" qui porte la modalité 3 (puisque K3 est voisin de 1) et un "benjamin" correspondant à la modalité 2 de la variable nationalité, puisque K2 est voisin de 3. Cela montre que la hiérarchie des résultats scolaires, constatée pour les ouvriers qualifiés et les ouvriers spécialisés, est bien une hiérarchie générale : assez systématiquement les enfants à nom polonais réussissent mieux que les enfants à nom français qui réussissent eux-mêmes mieux que les enfants à nom méditerranéen ou maghrébin.

§ 4 - POUVOIR EXPLICATIF D'UNE VARIABLE : DEFINITION ET CALCUL

Nous venons de voir l'intérêt qu'il y avait à caractériser l'influence d'une variable, c'est-à-dire à répondre à la question : telle variable (CSP, nationalité, nombre d'enfants, etc...) a-t-elle ou non une grande "influence" sur les retards scolaires ? L'analyse du tableau a apporté une première réponse : l'étude des EFI donne en effet une indication précieuse. Intuitivement, si les EFI sont grands en général, on a envie de dire que la variable introduite a une grande influence. Mais nous avons vu également que le maniement des EFI s'avérait doublement délicat, d'une part parce que tous les groupes ne sont pas représentés sur le tableau, d'autre part parce que certains EFI sont calculés à partir de groupes d'effectif restreint. Ainsi, dans l'exemple utilisé, EFI était important pour les cadres supérieurs lorsqu'on introduisait la nationalité, mais cela ne voulait pas dire que la nationalité permettait de comprendre convenablement le retard scolaire des enfants de cadres supérieurs.

En raison de ces difficultés, nous avons éprouvé la nécessité de définir de manière plus rigoureuse la notion de pouvoir explicatif d'une variable. Ce paragraphe comporte, de ce fait, un aspect plus ardu que les autres. Le lecteur devra néanmoins s'efforcer, même si ses connaissances mathématiques sont faibles, d'en lire au moins la partie relative à la notion "d'explication complète". Il en saura alors assez pour comprendre de manière intuitive comment nous avons ensuite raisonné.

4.1. - La notion d'explication complète

Un enfant est en retard ou il ne l'est pas. On ne peut donc pas vraiment parler de la probabilité pour qu'il soit en retard ou pour qu'il ne le soit pas. C'est seulement lorsqu'on considère un groupe d'enfants présentant une certaine homogénéité, que, en calculant le nombre d'enfants en retard dans ce groupe, on peut associer au groupe et à chaque membre de ce groupe le pourcentage de retard. C'est ce pourcentage que l'on a tendance à appeler : probabilité d'un enfant du groupe d'être en retard.

Que signifierait "expliquer complètement" le retard scolaire ? Cela signifierait qu'on a identifié un ensemble de causes telles que chaque fois que ces causes existent il y a retard scolaire et inversement que lorsque ces causes n'existent pas il n'y a pas retard scolaire. Supposons identifiées ces causes. On peut alors diviser la population d'enfants en groupes, certains groupes étant composés exclusivement d'enfants où les causes de retard existent, les autres étant exclusivement composés d'enfants pour lesquels ces causes n'existent pas. Par définition, dans les premiers groupes, tous les enfants seront en retard, dans les autres aucun enfant ne sera en retard. La probabilité associée aux enfants du premier groupe sera donc 1 et la probabilité associée aux enfants des autres groupes sera 0.

Supposons maintenant qu'on divise la même population d'élèves en groupes selon des modalités qui n'ont a priori rien à voir avec le retard scolaire, couleur des cheveux, ou forme des oreilles par exemple. On dit que l'on procède à une partition de la population. Chaque fois que l'on effectue une telle partition, on peut calculer dans chaque groupe constitué le pourcentage d'enfants en retard, c'est-à-dire la probabilité pour qu'un enfant soit en retard lorsqu'il appartient à ce groupe. Si vraiment les groupes ont été constitués au hasard, c'est-à-dire sans utiliser de critères liés à priori au retard scolaire, et s'ils ont des effectifs suffisamment importants, le pourcentage d'enfants en retard dans chaque groupe sera sensiblement le même c'est-à-dire que la partition conduira à associer à chaque enfant une probabilité de retard voisine de la probabilité moyenne pour l'ensemble de la population. On sera donc très loin de la partition effectuée à partir des causes complètes du retard, partition qui conduisait à n'affecter à chaque élève que des probabilités 0 ou 1 d'être en retard. La partition effectuée au hasard sera donc très éloignée de l'"explication complète".

Quand, intuitivement, on dit qu'une variable explique beaucoup de choses ou qu'elle entretient une relation intense avec le retard scolaire, on veut dire simplement qu'en découpant la population en fonction des modalités de cette variable on se rapproche beaucoup de l'explication complète.

C'est cette notion de distance à l'explication complète que l'on a fondé la définition de pouvoir explicatif d'une variable.

4.2. - Partitions associées à un arbre

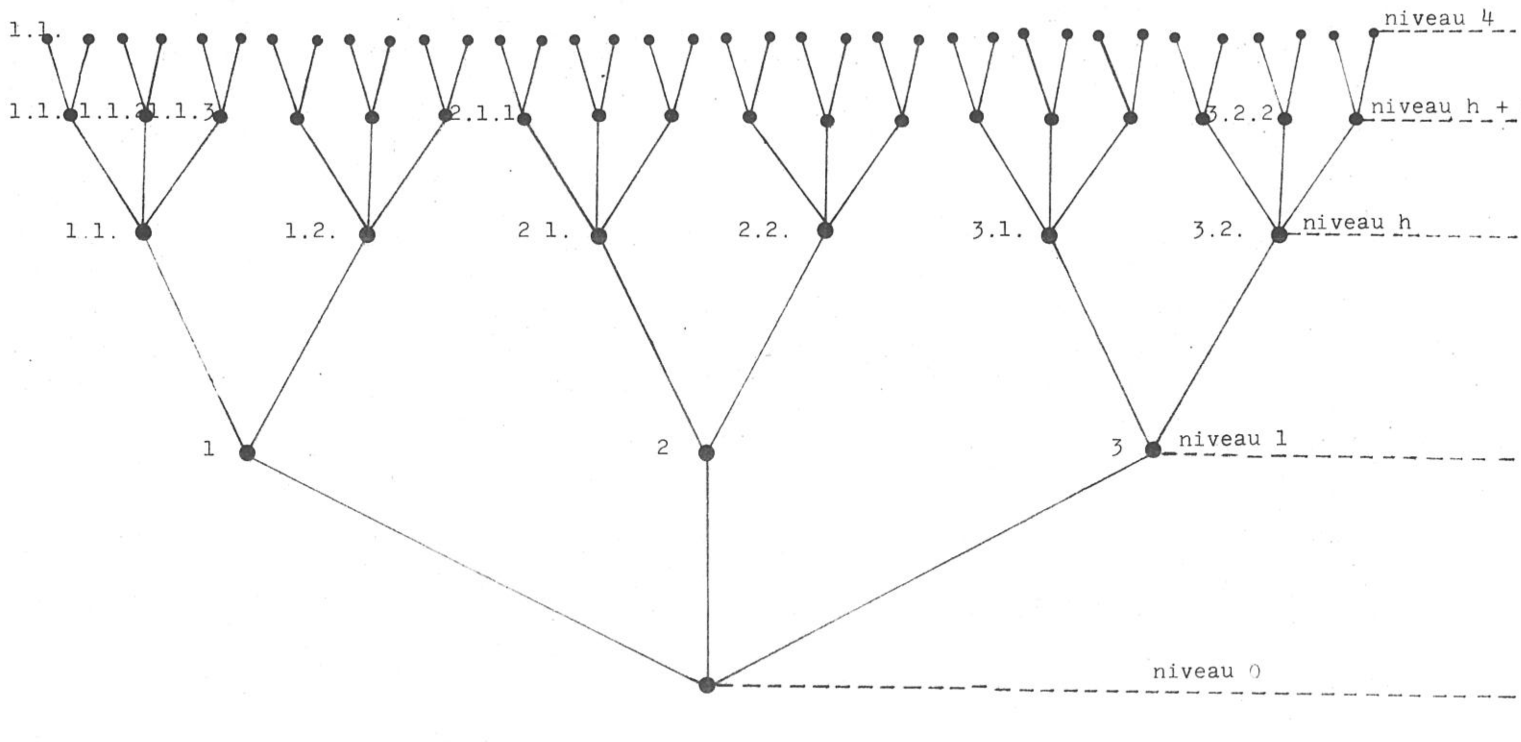
Dans l'analyse, six variables ont été introduites successivement. Chaque nouvelle variable introduite conduit à découper les anciens groupes en autant de sous groupes qu'il y a de modalités de la variable.

Ainsi, en introduisant la nationalité, qui comportait trois modalités (noms français ou belges, noms polonais, noms méditerranéens ou maghrébins) on a découpé chacun des onze groupes définis par la C.S.P. du père en trois sous groupes.

Par analogie avec un arbre généalogique nous avons donc parlé de groupe père et de groupe fils. Chaque nouvelle variable introduite correspond à une nouvelle génération.

On peut dessiner un arbre qui comporte autant de niveaux qu'il y a de variables, plus un, le niveau zéro, dans lequel toute la population est réunie dans un même groupe.

Voici un exemple d'arbre simple, à cinq niveaux :



Il correspond au cas où on aurait introduit successivement quatre variables ayant respectivement 3, 2, 3 et 2 modalités.

Chaque "noeud" de l'arbre correspond à un groupe et chaque niveau de l'arbre à une partition de la population.

Au niveau 1, la partition est faite en trois groupes (1, 2 et 3), au niveau 2 en six groupes (1.1, 1.2 ; 2.1, 2.2 ; 3.1, 3.2), etc... au niveau 4 en trente six groupes, soit 3 X 2 X 3 X 2.

A chaque arbre on peut donc associer autant de partitions de la population qu'il y a de niveaux dans l'arbre.

Pour chacune de ces partitions, on peut associer à chaque élève le pourcentage d'enfants en retard dans son groupe (chaque élève appartenant par définition à un groupe et à un groupe seulement) et la probabilité d'être en retard, nombre compris entre 0 et 1 qui est par définition le rapport entre nombre d'enfants en retard dans le groupe et effectifs du groupe.

Dire qu'une partition est très éloignée de l'explication complète revient à dire que les nombres ainsi associés à chaque élève sont loin d'être tous voisins de 0 ou de 1. C'est cette distance que l'on veut définir.

4.3. - Distance entre deux partitions

A chaque partition de la population correspond, pour chaque élève, λ, une valeur p_λ appelée "probabilité d'être en retard".

Soient deux partitions G et G' de la même population. Elles associent respectivement à l'élève λ les probabilités p_λ et p'_λ.

La "distance" entre les deux partitions, du point de vue des retards scolaires, est intuitivement liée aux différences entre p_λ et p'_λ pour chacun des élèves.

C'est bien en raison du fait que les probabilités p_λ associées à une partition sont très différentes des valeurs 0 ou 1 que l'on dit intuitivement que cette partition est très éloignée d'une "explication complète".

On définira donc la distance entre deux partitions G et G', notée D(G, G') par l'expression :

$$D(G, G') = \sqrt{\frac{\sum_{\lambda=1}^{n} (p_\lambda - p'_\lambda)^2}{n}}$$

Cette définition "colle" bien avec la définition intuitive et possède bien les propriétés classiques d'une <u>fonction distance.</u>

4.4. - <u>Distance d'une partition à l'explication complète</u>

Chaque élève étant en retard ou ne l'étant pas, on peut associer à ceux qui sont en retard la probabilité 1 et à ceux qui ne le sont pas la probabilité 0.

Appelons, pour nous conformer à l'habitude en mathématique, δ_λ ce nombre 0 ou 1 associé à l'élève λ. C'est ce nombre qui lui serait effectivement associé dans une partition qui expliquerait complètement le retard. Soit L une telle partition.

On peut alors calculer la distance de toute partition G à cette partition L, cette distance étant appelée "distance à l'explication complète".

Par définition : $$D(G, L) = \sqrt{\frac{\sum_{\lambda=1}^{n} (p_\lambda - \delta_\lambda)^2}{n}}$$

Il se trouve que dans ce cas le calcul de D (G, L) est très simple. La partition G "découpe" la population en k groupes. Appelons φ (lettre grecque qui se lit Fi) l'un de ces groupes, n_φ le nombre d'élèves appartenant à ce groupe.

On peut écrire : $D\ (G,\ L) = \sqrt{\dfrac{\sum\limits_{\varphi=1}^{k}\left[\sum\limits_{j=1}^{n\varphi}(p_{j,\varphi} - \delta_{j\varphi})^2\right]}{n}}$

Cette expression est la même que la précédente, on a seulement regroupé dans la somme les termes correspondant à chacun des groupes.

Calculons : $\sum\limits_{j=1}^{n\varphi}(p_{j,\varphi} - \delta_{j,\varphi})^2 = \sum\limits_{j=1}^{n\varphi}(p_{j,\varphi}^{\ 2} - 2\ p_{j\varphi}\delta_{j,\varphi} + \delta_{j,\varphi}^{\ 2})$

Revenons à la définition de $p_{j\varphi}$: c'est le rapport entre le nombre d'enfants en retard dans le groupe et l'effectif du groupe, n_φ.
Pour tous les élèves d'un même groupe, $p_{j\varphi}$ est le même et est égal à p_φ, "probabilité d'être en retard" dans le groupe (c'est le fameux P Fi que nous avons vu apparaître dans le tableau, au cours du paragraphe 3).

Toujours par définition, $p_{j\varphi} = p_\varphi = \sum\limits_{j=1}^{n\varphi}\delta_{j\varphi}/n_\varphi$ puisque $\delta_{j\varphi}$ prend la valeur 1 pour un élève en retard et 0 autrement.

On peut donc écrire ($\delta_{j\varphi}^{\ 2} = \delta_{j\varphi}$ car $1^2 = 1$ et $0^2 = 0$) :

$$\sum_{j=1}^{n\varphi}(p_{j\varphi} - \delta_{j\varphi})^2 = n_\varphi p_\varphi^{\ 2} - 2\,p_\varphi\Big(\sum_{j=1}^{n\varphi}\delta_{j\varphi}\Big) + \sum_{1}^{n\varphi}\delta_{j\varphi}^{\ 2} = n_\varphi p_\varphi^{\ 2} - 2\,n_\varphi p_\varphi^{\ 2} + n_\varphi p_\varphi$$

Soit en définitive : $\sum\limits_{j=1}^{n\varphi}(p_{j\varphi} - \delta_{j\varphi})^2 = n\varphi\ (p_\varphi - p_\varphi^2) = n_\varphi p_\varphi(1 - p_\varphi)$

D'où :

$$\boxed{\sqrt{\frac{\sum\limits_{\varphi=1}^{k} n_\varphi p_\varphi(1 - p_\varphi)}{n}} = D\ (G, L)}$$

Ce qui signifie que la connaissance des n_φ (les N Fi du tableau) et des p_φ (les P Fi du tableau) permet de déterminer la distance d'une partition à l'explication complète.

4.5. - Pouvoir explicatif d'un niveau

On part du niveau 0 : aucune variable n'est introduite, toute la population des élèves est réunie dans un même groupe. Le rapport entre nombre d'enfants en retard et nombre d'enfants total, p_o, définit la probabilité moyenne d'être en retard dans le District de Denain.

A ce niveau 0 correspond une partition G_o composée d'un seul groupe (l'ensemble de la population). La distance à l'explication complète s'écrit :

$$D\,(G_o, L) = \sqrt{p_o(1 - p_o)}$$

Lorsqu'on introduit successivement les variables on se rapproche, plus au moins vite, de l'explication complète. On ne l'atteint jamais. Regardons pour nous en convaincre le tableau "niveau 6" des retards scolaires à l'école primaire, toutes classes confondues. Si on était proche de l'explication complète, on n'aurait pour les P Fi que des 0 ou des 1. Ce n'est pas le cas. Lorsque les valeurs de P Fi sont supérieures à 0.5 elles sont effectivement très voisines de 1 (0.978 à la 17ème colonne; 0.987 à la 25ème colonne, 1.000 à la 29ème). Ceci s'explique facilement : lorsque les instituteurs ont décrit une cause de retard c'est qu'il s'agissait d'un élève en retard.

Par contre il existe des élèves en retard pour lesquels l'instituteur n'a apporté aucune raison particulière au retard sinon des raisons déjà prises en compte dans les niveaux précédents (origine étrangère, famille nombreuse ou milieu familial défavorable par exemple). Dans les groupes où aucune raison particulière n'a été avancée pour expliquer le retard, il y a des élèves en retard et d'autres qui ne le sont pas. Prenons par exemple le cas d'élèves d'origine étrangère. Certains sont en retard et l'instituteur a mis comme cause de retard, "étranger". C'est une explication possible (handicap culturel, difficulté de langage, etc...). Mais les enfants d'origine étrangère ne sont pas tous en retard, soit parce qu'ils ne subissent pas tous le même handicap du fait de leur origine (certains sont nés en France, d'autres viennent d'y arriver), soit parce que certains le surmontent mieux que d'autres.

L'origine étrangère constitue donc non une cause absolue de retard ou de non retard mais un handicap qui "prédispose plus au retard".

Supposons que l'on sache définir dans l'absolu "l'intelligence" d'un élève et que la distribution des intelligences soit la même dans la population française et la population étrangère.

Dans la population française, les 30 % moins doués seront en retard scolaire, mais, en raison du handicap de la langue, les 45 % moins doués de la population étrangère le seront.

Introduire d'autres variables explicatives que "l'intelligence" pour comprendre les retards scolaires c'est admettre que d'autres facteurs viennent fortement infléchir les résultats scolaires, constituant soit de des atouts soit des handicaps.

Plus les variables introduites sont "explicatives" et plus elles réduisent la portée d'une explication des retards scolaires par la diversité des dons innés.

Définir le pouvoir explicatif du niveau h consiste à dire "de combien on s'est approché de l'explication complète, depuis le niveau 0".

Soit G_h la partition associée au niveau h. Le pouvoir explicatif du niveau h sera donc défini par :

$$\mathcal{P}_h = \frac{D(G_o, L) - D(G_h, L)}{D(G_o, L)} = 1 - \frac{D(G_h, L)}{D(G_o, L)}$$

soit :

$$\mathcal{P}_h = \frac{\sqrt{\dfrac{\sum_{\varphi=1}^{k} n_\varphi p_\varphi (1 - p_\varphi)}{n}} - \sqrt{p_o (1 - p_o)}}{\sqrt{p_o (1 - p_o)}}$$

Dans les tableaux, on a porté non pas $\mathcal{P}_h$ mais DH, "distance relative à l'explication complète". Par définition,

$$DH = \frac{D(G_h, L)}{D(G_o, L)}$$

Donc $\mathcal{P}_h = 1 - DH$

A titre d'exemple, lisons les valeurs DH pour l'arbre des retards scolaires dans les écoles primaires du District de Denain, toutes classes confondues.

En haut de chaque page, à gauche de la première ligne, on lit DH immédiatement en dessous du numéro de niveau.

On trouve les valeurs suivantes :

DH (niveau 1) = 0.966
DH (niveau 2) = 0.962
DH (niveau 3) = 0.961
DH (niveau 4) = 0.934
DH (niveau 5) = 0.891
DH (niveau 6) = 0.754

D'un niveau à l'autre, on se rapproche de 0 tout en étant loin de l'atteindre (ce qui traduit le fait que nous avons identifié des "prédispositions au retard" plutôt que des causes absolues de retard ou de non retard).

4.6. - Pouvoir explicatif d'une variable

Lorsqu'on introduit la variable h, passant ainsi du niveau h - 1 au niveau h, on améliore le pouvoir explicatif du niveau : DH (niveau h) est plus petit que DH (niveau h - 1).

Le pouvoir explicatif de la variable h est donc défini par :

$$\boxed{\gamma(h) = [DH\ (\text{niveau } h - 1) - DH\ (\text{niveau } h)] \times 1\,000}$$

L'étude des retards scolaires à l'école primaire, toutes classes confondues, permet de dresser le tableau suivant :

Ordre d'introduction de la variable	Nom de la variable	Pouvoir explicatif
1	Catégorie socio-professionnelle du père	34
2	Nationalité	04
3	Activité professionnelle de la mère	01
4	Nombre d'enfants	28
5	Milieu familial	43
6	Causes du retard	137

Ce tableau permet de comparer directement l'influence des différentes variables introduites sur le retard scolaire.

Comme pour toute mesure, il faut apprendre à apprécier les ordres de grandeur. Le pouvoir explicatif de 34, dévolu à la catégorie socio-professionnelle traduit des écarts de très grande ampleur puisque 11,5 % des enfants de cadres supérieurs sont en retard contre 66,6 % des enfants de mineurs.

Les variables 5 (milieu familial) et 6 (causes du retard) apparaissent avec un pouvoir explicatif considérable : respectivement 43 et 137. Mais nous avons vu au début du chapitre qu'il fallait faire très attention à leur signification : c'est le retard scolaire qui induit en partie le jugement de l'instituteur ; il est donc hasardeux d'expliquer ensuite le retard par ce jugement !

Ensuite vient la variable 4, "nombre d'enfants", avec un pouvoir explicatif de 28. Le nombre d'enfants dans la famille conditionne donc très fortement les résultats scolaires. Résultat d'autant plus remarquable qu'on a raisonné à classe sociale, à nationalité, et à travail de la mère fixés.

Les deux autres variables introduites, nationalité et travail de la mère, pèsent d'un poids plus faible, pour deux raisons : d'une part parce que les pourcentages ne varient pas dans des proportions considérables d'une modalité à l'autre de la variable, d'autre part parce que les effectifs correspondant à chaque modalité sont très inégaux : les français et les femmes qui ne travaillent pas sont largement majoritaires dans le District de Denain. Or le calcul de D (G, L) et du pouvoir explicatif prend en compte à la fois l'effectif n_{φ} et le pourcentage p_{φ} de chaque groupe.

Ceci traduit bien une idée intuitive : même si les enfants étrangers étaient très en retard, cela n'expliquerait pas les différences entre classes sociales car les étrangers ne représentent jamais plus de 25 % des effectifs d'une classe sociale et souvent bien moins.

Remarque fondamentale : Le pouvoir explicatif d'une variable dépend de son ordre d'introduction. On en trouve une illustration dans le cas de Denain : l'activité de la mère a été introduite avant le nombre d'enfants et cela conduit à la conclusion que les enfants dont la mère travaille sont moins en retard que les enfants dont la mère ne travaille pas.

Mais l'activité de la mère est très liée au nombre d'enfants : au delà de trois enfants très peu de mères travaillent.

Or la réussite scolaire dépend étroitement du nombre d'enfants dans la famille : les enfants de familles peu nombreuses réussissent beaucoup mieux que les autres. Si on introduit le nombre d'enfants avant le travail de la mère, l'influence du travail de la mère s'annule : son pouvoir explicatif est proche de 0.

Cela ne veut pas dire nécessairement que l'ordre initial était mal choisi. Il faut seulement bien prendre conscience que le pouvoir explicatif de chaque variable dépend de son ordre d'introduction.

§ 5 - DETERMINATION DES MODALITES D'UNE VARIABLE

5.1. - Position du problème

Une femme travaille ou ne travaille pas. La variable "travail de la mère" n'a que deux modalités. Par contre on a décidé de façon relativement gratuite que la variable "catégorie socio-professionnelle" aurait onze modalités et que la variable "nombre d'enfants" en aurait trois.

En prenant un plus grand nombre de modalités pour une variable on améliore la précision de l'analyse : il n'est pas inutile de savoir que les ouvriers qualifiés de la sidérurgie ont des enfants plus souvent en retard que ceux de la métallurgie ou que les élèves issus de familles de six enfants sont nettement plus en retard que ceux de familles de cinq enfants. Mais cette précision deviendrait vite illusoire car le maintien de modalités nombreuses pour chaque variable multiplierait le nombre de groupes et réduirait l'effectif de chacun en rendant la comparaison entre pourcentages difficile sur le plan pratique et peu significative sur le plan théorique.

Pour réduire le nombre de modalités de chaque variable il faut effectuer des regroupements. Par exemple, on a décrété, en ce qui concerne la nationalité, que les noms belges "iraient avec" les noms français et les noms espagnols avec les noms maghrébins. En ce qui concerne le nombre d'enfants on a décrété que les familles de un, deux ou trois enfants formeraient la modalité 1, les familles de quatre, cinq, six enfants la modalité 2 et les familles de sept enfants et plus la modalité 3.

Quelle logique a présidé à ces regroupements ?

Pour exploiter l'enquête réalisée par M. HANTUTE sur les élèves d'école primaire on a procédé de manière intuitive, en considérant qu'il y avait plus de "proximité" entre belges et français qu'entre français et algériens, par exemple.

Par contre, lorsqu'il s'est agi d'appliquer la même méthode à l'enquête des Centres d'Information et d'Orientation il a paru nécessaire d'améliorer les procédures en mettant au point une technique plus rigoureuse de regroupement des modalités des variables introduites.

Le problème du choix des modalités à regrouper se posait pour deux variables : pays de naissance du père et nombre d'enfants dans la famille (1).

On voulait, pour pouvoir utiliser le même programme informatique que dans l'enquête de M. HANTUTE, réduire à trois le nombre de modalités de chaque variable.

Il s'agissait donc de constituer pour les pays de naissance et pour la taille des familles trois groupes "aussi homogènes que possible".

Que recouvre la notion intuitive d'homogénéïté ?

Pour les pays de naissance, le seul critère à prendre en compte est les résultats scolaires. Un regroupement du type : [Français et Algériens], [Belges, Italiens et Portugais] et [Polonais, Marocains, Espagnols] nous parait curieux parce qu'il nous semble qu'il y a plus de "proximité" entre Algériens et Marocains par exemple qu'entre Français et Algériens. Mais après tout, si, du point de vue des résultats scolaires, Français et Algériens ont des comportements semblables, Algériens et Marocains des comportements différents, pourquoi mettre ensemble Algériens et Marocains sous prétexte que leurs pays sont limitrophes ? Par contre, en ce qui concerne le nombre d'enfants, un regroupement du type : [familles de 1 ou de 3 enfants] ; [familles de 2, 5 ou plus de 9 enfants] ; [famille de 4, 6, 7, 8, 9 enfants] n'a guère de sens. On ne saurait comment interpréter les différences entre les modalités de la variable.

Les regroupements doivent donc "mettre ensemble" des groupes ayant des comportements semblables face à l'appareil scolaire, tout en gardant un sens.

La règle choisie est la suivante : "le meilleur regroupement est celui qui réduit le moins le pouvoir explicatif de la variable introduite, à condition que le regroupement garde un sens". La notion de "perte minimum du pouvoir explicatif", est la meilleure traduction logique de l'idée intuitive d'homogénéïté.

(1) Au départ, il avait été prévu également d'introduire la nationalité des élèves. Mais seuls 5 % des élèves sont étrangers : les effectifs de chaque nationalité sont trop faibles pour que les différences observées soient significatives. Par contre, pour 18 % des élèves le père est né hors de France.

5.2. - Procédure adoptée

Il s'agit de mesurer la perte de pouvoir explicatif lorsqu'on regroupe deux modalités d'une variable. Le pouvoir explicatif de la variable h est donné par la formule :

$$\gamma(h) = \left[\frac{D\ (G_{h-1},\ L)}{D\ (G_o,\ L)} - \frac{D\ (G_h,\ L)}{D\ (G_o,\ L)} \right] \times 1\ 000$$

Soit lorsque toutes les variables sont fixées jusqu'au rang h - 1 :

$$\gamma(h) = a \sqrt{\sum_{\varphi=1}^{k} n_\varphi p_\varphi (1 - p_\varphi)} + b$$

a étant une constante négative et b une constante positive.

$\gamma(h)$ varie donc en sens inverse de la fonction

$$y = + \sum_{\varphi=1}^{k} n_\varphi p_\varphi (1 - p_\varphi)$$

Dans cette fonction on peut regrouper tous les groupes frères entre eux.

Soit q le nombre initial de modalités de la variable. Au niveau h - 1 il y avait donc $\frac{k}{q}$ groupes qui sont les $\frac{k}{q}$ pères des groupes actuels. Numérotons ces groupes de 1 à $\frac{k}{q}$ en les caractérisant par l'indice i.

Chaque groupe fils est désigné par deux indices : l'indice i qui caractérise son père et l'indice j qui caractérise la modalité de la variable h qui lui a donné naissance.

On peut alors écrire :

$$y = \sum_{i=1}^{k/q} \sum_{i=1}^{q} n_{ij} \, p_{ij} \, (1 - p_{ij})$$

Les modalités j et k de la variable h à regrouper sont celles dont le regroupement <u>augmente le moins</u> la valeur de y.

Commençons à raisonner sur un cas simple. On a un groupe d'effectif N et de probabilité associée P.

Ce groupe se divise en deux sous-groupes d'effectif et de probabilité respectifs n_1, p_1 et n_2, p_2.

Par définition on a :

$$\begin{cases} n_1 + n_2 = N \\ n_1 \, p_1 + n_2 \, p_2 = NP \end{cases}$$

Comment varie la fonction $y = \sum np\,(1 - p)$ lorsque le groupe initial éclate en deux ? C'est à dire combien vaut la différence :

$$np\,(1 - p) - \left[n_1 \, p_1 \, (1 - p_1) + n_2 \, p_2 \, (1 - p_2) \right] = \Delta y$$

Un calcul simple montre que l'on a : $\boxed{\Delta y = \frac{n_1 \, n_2}{n_1 + n_2} \, (p_1 - p_2)^2}$

Inversement, quand on regroupe deux groupes initialement séparés, y augmente de la même valeur.

Lorsqu'on regroupe deux modalités j et k d'une variable on effectue, <u>dans chaque ensemble de frères</u> (caractérisés par l'indice i du père), une fusion de cette sorte et on augmente donc y de :

$$\frac{n_{ij} \, n_{ik}}{n_{ij} + n_{ik}} \, (p_{ij} - p_{ik})^2$$

En définitive, le regroupement des modalités j et k conduit à augmenter y de la valeur :

$$\boxed{\rho_{jk} = \sum_{i=1}^{k/q} \frac{n_{ij}\ n_{ik}}{n_{ij} + n_{ik}} \quad (p_{ij} - p_{ik})^2}$$

S'il s'agit de regrouper seulement deux modalités de la variable h, le regroupement à effectuer est celui pour lequel ρ_{jk} est minimum.

On calcule donc tous les ρ_{jk} (soit $\frac{q\ (q - 1)}{2}$ valeurs) et on choisit de regrouper le couple (j, k) rendant ρ_{jk} minimum.

Dans l'analyse effectuée, il fallait faire plus d'un regroupement puisqu'on devait passer de variables à neuf modalités à des variables à trois modalités. Plutôt que d'effectuer une démarche itérative trop coûteuse, on a préféré faire sortir à chaque fois le tableau des ρ_{jk} et on a pris les décisions de regroupement en tenant compte des ρ_{jk} les plus petits (qui définissent les regroupements les plus naturels), des ρ_{jk} les plus grands (qui définissent les modalités qu'il ne faut pas regrouper) et des impossibilités (on ne peut pas regrouper deux et quatre enfants en laissant trois de côté). La technique s'est avérée tout à fait opératoire.

5. 3. - Application

L'exemple le plus frappant d'application concerne le nombre d'enfants dans la famille. Dans l'enquête C.I.O. ce n'est pas le retard scolaire que l'on cherchait à expliquer mais l'affectation d'un élève à une section : pourcentage d'élèves d'un groupe donné en section 1, 2, pratique ou C.A.P..

A chaque pourcentage correspond un arbre distinct. Il y a ainsi quatre arbres, chacun servant à décrire les variations du pourcentage d'élèves d'un certain type dans une section donnée.

On pouvait ne pas regrouper de la même manière les différentes modalités d'une même variable, d'un arbre à l'autre.

Au contraire des différences de regroupement peuvent s'avérer très significatives. Comparons par exemple les regroupements effectués pour le nombre d'enfants dans l'analyse du pourcentage d'élèves en section 1 et dans celle du pourcentage d'élèves en section pratique.

Modalités regroupées de la variable nombre d'enfants	Regroupements effectués	
	Etude du pourcentage d'élèves en section 1	Etude du pourcentage d'élèves en section pratique
1	1 ou 2 enfants par famille	1, 2, 3 ou 4 enfants par famille
2	3 ou 4 enfants par famille	5 ou 6 enfants par famille
3	5 enfants et plus	7 enfants et plus

En section 1 une opposition forte se manifeste déjà entre un ou deux enfants et trois ou quatre enfants. Alors qu'en section pratique, qui recueille les élèves les moins "doués," les enfants de famille nombreuse sont majoritaires et le premier clivage s'établit au delà de quatre enfants.

La relation entre nombre d'enfants et résultat scolaire est tellement intense que les clivages selon la taille de la famille se trouvent déplacés vers les familles nombreuses lorsqu'on passe d'une "bonne" section à une "mauvaise" section.

NIVEAU 1 : Analyse des retards scolaires de l'école primaire dans le district de Denain - Toutes classes confondues
****

Catégorie Socio Professionnelle des Parents

DH = 0.966 SOMME = 9565

	Agriculteurs	C. Sup	C. Moyens	(Commerçants Artisans	Employés	OP	OS	Mineurs	P. Service	Absents Invalides Décédés	Pensionnés Retraités
CSP	1	2	3	4	5	6	7	8	9	10	11
NFI	142	226	578	427	550	2921	2937	739	95	720	230
TFI	0.015	0.024	0.060	0.045	0.058	0.305	0.307	0.077	0.010	0.075	0.024
PFI	0.282	0.115	0.209	0.344	0.344	0.392	0.497	0.666	0.411	0.635	0.630

Rangs pondérés des modalités de chaque variable			Pouvoir explicatif des variables		
Modalité 1	Modalité 2	Modalité 3	Numéro du niveau	Nom de la variable	Pouvoir Explicatif
X	X	X	1	C. S. P	34
2,04	2,81	1,22	2	Origine Nationale	04
1,22	1,88	X	3	Travail de la mère	01
2,84	1,94	1,08	4	Nombre d'enfants	27
2,83	2,01	1,13	5	Jugement sur le milieu familial	43
2,78	1,27	1,66	6	Causes du retard	137
			total 1 à 4	X	76
			total 1 à 6	X	236

NIVEAU 2
****

Nationalité (origine nationale)

DH = 0.962 K1 = 2.038 K2 = 2.807 K3 = 1.225 SOM30 = 9550 SOM90 = 9565

CSP	6	7	7	10	3	5	6	8	4	8	6	2	7	11	10
NAT	1	1	3	1	1	1	3	1	1	3	2	1	2	1	3
TMERE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NBENF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MILIEU	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RAISONS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

****

NFI	2221	2185	574	525	506	496	411	382	347	297	289	208	178	172	147
TFI	0.760	0.744	0.195	0.729	0.875	0.902	0.141	0.517	0.813	0.402	0.099	0.920	0.061	0.748	0.204
UFI	0.232	0.228	0.060	0.055	0.053	0.052	0.043	0.040	0.036	0.031	0.030	0.022	0.019	0.018	0.015
PFI	0.402	0.493	0.547	0.621	0.204	0.343	0.428	0.670	0.311	0.663	0.263	0.101	0.376	0.640	0.687
RANG	2	2	1	3	2	2	1	1	2	2	3	2	3	2	1
GDM	0.026	0.054	0.0	0.066	0.106	0.157	0.0	0.0	0.237	0.007	0.165	0.455	0.171	0.001	0.0
PDM	0.139	0.117	0.171	0.0	0.037	0.109	0.165	0.020	0.033	0.013	0.0	0.101	0.0	0.113	0.066
EFI	0.165	0.171	0.171	0.066	0.143	0.267	0.165	0.020	0.271	0.020	0.165	0.556	0.171	0.115	0.066

CSP	1	9	4	8	10	3	11	3	5	5	11	4	9	2	2
NAT	1	1	3	2	2	2	3	3	2	3	2	2.	2	2	3
TMERE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NBENF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MILIEU	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RAISONS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

****

NFI	135	77	62	60	48	42	39	30	30	24	19	18	10	9	9
TFI	0.951	0.811	0.145	0.081	0.067	0.073	0.170	0.052	0.055	0.044	0.083	0.042	0.105	0.040	0.040
UFI	0.014	0.008	0.006	0.006	0.005	0.004	0.004	0.003	0.003	0.003	0.002	0.002	0.001	0.001	0.001
PFI	0.267	0.403	0.548	0.650	0.625	0.310	0.641	0.167	0.233	0.500	0.526	0.278	0.400	0.0	0.556
RANG	3	2	1	3	2	1	1	3	3	1	3	3	3	3	1
GDM	0.733	0.097	0.0	0.020	0.062	0.0	0.0	0.143	0.267	0.0	0.115	0.271	0.100	0.556	0.0
PDM	0.0	0.003	0.271	0.0	0.004	0.143	0.115	0.0	0.0	0.267	0.0	0.0	0.0	0.0	0.556
EFI	0.733	0.100	0.271	0.020	0.066	0.143	0.115	0.143	0.267	0.267	0.115	0.271	0.100	0.556	0.556

NIVEAU 3
****

Travail de la mère

DH = 0.961 K1 = 1.221 K2 = 1.876 K3 = 0.0 SOM30 = 9246 SOM90 = 9550

CSP	7	6	7	5	3	6	8	10	8	6	4	7	6	10	2
NAT	1	1	3	1	1	3	1	1	3	2	1	2	1	1	1
TMERE	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1
NBENF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MILIEU	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RAISONS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

****

NFI	2065	2057	555	432	406	395	369	362	294	272	254	173	164	163	161
TFI	0.945	0.926	0.967	0.871	0.802	0.961	0.966	0.690	0.990	0.941	0.732	0.972	0.074	0.310	0.774
UFI	0.216	0.215	0.058	0.045	0.042	0.041	0.039	0.038	0.031	0.028	0.027	0.018	0.017	0.017	0.017
PFI	0.499	0.406	0.548	0.359	0.222	0.425	0.669	0.646	0.660	0.257	0.319	0.376	0.348	0.564	0.118
RANG	1	1	1	1	1	2	2	1	2	2	1	2	2	2	1
GDM	0.0	0.0	0.0	0.0	0.0	0.075	0.023	0.0	0.340	0.096	0.0	0.024	0.059	0.082	0.0
PDM	0.499	0.406	0.548	0.359	0.222	0.425	0.669	0.646	0.660	0.257	0.319	0.376	0.348	0.564	0.118
EFI	~~0.499~~	~~0.406~~	~~0.548~~	~~0.359~~	~~0.222~~	~~0.500~~	~~0.692~~	~~0.646~~	~~1.000~~	~~0.353~~	~~0.319~~	~~0.400~~	~~0.406~~	~~0.646~~	~~0.118~~

CSP	11	10	7	1	3	4	9	5	8	4	2	10	3	11	11
NAT	1	3	1	1	1	1	1	1	2	3	1	2	2	3	1
TMERE	1	1	2	1	2	2	1	2	1	1	2	1	1	1	2
NBENF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MILIEU	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RAISONS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

****

NFI	138	124	120	113	100	93	66	64	59	52	47	44	35	35	34
TFI	0.802	0.844	0.055	0.837	0.198	0.268	0.857	0.129	0.983	0.839	0.226	0.917	0.833	0.897	0.198
UFI	0.014	0.013	0.013	0.012	0.010	0.010	0.007	0.007	0.006	0.005	0.005	0.005	0.004	0.004	0.004
PFI	0.667	0.685	0.400	0.265	0.130	0.290	0.409	0.234	0.644	0.558	0.043	0.636	0.286	0.657	0.529
RANG	1	2	2	2	2	2	1	2	2	1	2	1	2	1	2
GDM	0.0	0.010	0.099	0.007	0.092	0.029	0.0	0.124	0.356	0.0	0.075	0.0	0.143	0.0	0.137
PDM	0.667	0.685	0.400	0.265	0.130	0.290	0.409	0.234	0.644	0.558	0.043	0.636	0.286	0.657	0.529
~~EFI~~	~~0.667~~	~~0.696~~	~~0.499~~	~~0.273~~	~~0.222~~	~~0.319~~	~~0.409~~	~~0.359~~	~~1.000~~	~~0.558~~	~~0.118~~	~~0.636~~	~~0.429~~	~~0.657~~	~~0.667~~

NIVEAU 4
****

Nombre d' enfants dans la Famille

DH = 0.934 K1 = 2.837 K2 = 1.945 K3 = 1.077 SOM30 = 7708 SOM90 = 9246

CSP	6	7	7	6	7	3	7	5	6	6	6	10	8	4	7
NAT	1	1	1	1	1	1	3	1	2	1	3	1	1	1	3
TMERE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NBENF	1	1	2	2	3	1	2	1	1	3	2	2	2	1	3
MILIEU	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RAISONS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NFI	1046	873	826	815	366	273	271	265	202	196	186	182	175	158	153
TFI	0.509	0.423	0.400	0.396	0.177	0.672	0.488	0.613	0.743	0.095	0.471	0.503	0.474	0.622	0.276
UFI	0.109	0.091	0.086	0.085	0.038	0.029	0.028	0.028	0.021	0.020	0.019	0.019	0.018	0.017	0.016
PFI	0.309	0.357	0.561	0.472	0.697	0.179	0.531	0.253	0.203	0.653	0.452	0.643	0.634	0.234	0.608
RANG	3	3	2	2	1	3	2	3	3	1	2	2	2	3	1
GDM	0.344	0.339	0.136	0.181	0.0	0.463	0.076	0.377	0.233	0.0	0.182	0.101	0.143	0.225	0.0
PDM	0.0	0.0	0.203	0.164	0.339	0.0	0.020	0.0	0.0	0.344	0.143	0.167	0.059	0.0	0.096
EFI	0.344	0.339	0.339	0.344	0.339	0.463	0.096	0.377	0.233	0.344	0.325	0.267	0.202	0.225	0.096

CSP	6	5	8	6	7	8	3	10	2	8	7	10	7	3	4
NAT	3	1	3	1	3	1	1	1	1	3	2	1	1	1	1
TMERE	1	1	1	2	1	1	1	1	1	1	1	2	2	2	2
NBENF	1	2	2	1	1	3	2	3	1	3	1	1	1	1	1
MILIEU	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RAISONS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NFI	149	140	140	134	131	121	119	117	117	110	102	92	88	85	76
TFI	0.377	0.324	0.476	0.817	0.236	0.328	0.293	0.323	0.727	0.374	0.590	0.564	0.733	0.850	0.817
UFI	0.016	0.015	0.015	0.014	0.014	0.013	0.012	0.012	0.012	0.012	0.011	0.010	0.009	0.009	0.008
PFI	0.309	0.507	0.657	0.336	0.511	0.777	0.269	0.744	0.111	0.673	0.314	0.565	0.341	0.094	0.289
RANG	3	2	2	2	3	1	2	1	3	1	3	2	3	3	2
GDM	0.325	0.122	0.016	0.064	0.096	0.0	0.374	0.0	0.222	0.0	0.186	0.066	0.259	0.906	0.711
PDM	0.0	0.254	0.021	0.336	0.0	0.202	0.089	0.267	0.0	0.036	0.0	0.027	0.0	0.0	0.039
EFI	0.325	0.377	0.036	0.400	0.096	0.202	0.463	0.267	0.222	0.036	0.186	0.093	0.259	0.906	0.750

NIVEAU 5
****

Jugement sur le milieu familial

DH = 0.891 K1 = 2.832 K2 = 2.008 K3 = 1.132 SOM30 = 5684 SOM90 = 7679

CSP	6	7	6	7	6	7	7	6	3	7	5	7	6	6	4
NAT	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
TMERE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NBENF	1	1	2	2	2	2	1	1	1	2	1	3	1	2	1
MILIEU	1	1	1	2	2	1	2	2	1	3	1	3	1	3	1
RAISONS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

****

NFI	723	505	393	309	278	275	267	254	244	239	211	182	166	138	137
TFI	0.691	0.578	0.482	0.374	0.341	0.333	0.306	0.243	0.894	0.289	0.796	0.497	0.822	0.169	0.867
UFI	0.076	0.053	0.041	0.032	0.029	0.029	0.028	0.027	0.026	0.025	0.022	0.019	0.017	0.014	0.014
PFI	0.220	0.228	0.382	0.579	0.518	0.433	0.472	0.461	0.156	0.678	0.204	0.725	0.145	0.623	0.204
RANG	3	3	3	2	2	3	2	2	3	1	3	1	3	1	3
GDM	0.452	0.472	0.242	0.099	0.105	0.245	0.228	0.211	0.844	0.0	0.433	0.0	0.411	0.0	0.296
PDM	0.0	0.0	0.0	0.147	0.136	0.0	0.244	0.241	0.0	0.245	0.0	0.064	0.0	0.242	0.0
EFI	0.452	0.472	0.242	0.245	0.242	0.245	0.472	0.452	0.844	0.245	0.433	0.064	0.411	0.242	0.296

CSP	7	7	2	7	6	6	5	3	7	3	6	6	10	10	8
NAT	1	3	1	1	3	1	1	1	3	1	1	3	1	1	1
TMERE	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1
NBENF	3	2	1	1	1	1	2	2	2	1	3	2	2	3	2
MILIEU	2	2	1	3	1	1	1	1	1	1	2	2	3	3	3
RAISONS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

****

NFI	130	117	110	100	96	95	88	87	86	81	79	76	76	72	70
TFI	0.355	0.432	0.940	0.115	0.644	0.709	0.629	0.731	0.317	0.953	0.403	0.409	0.418	0.615	0.400
UFI	0.014	0.012	0.012	0.011	0.010	0.010	0.009	0.009	0.009	0.009	0.008	0.008	0.008	0.008	0.007
PFI	0.662	0.598	0.073	0.700	0.271	0.242	0.386	0.207	0.349	0.086	0.671	0.513	0.803	0.736	0.786
RANG	3	2	3	1	3	3	3	3	3	2	2	2	1	2	1
GDM	0.064	0.058	0.927	0.0	0.313	0.591	0.341	0.793	0.308	0.164	0.079	0.030	0.0	0.087	0.0
PDM	0.0	0.249	0.0	0.472	0.0	0.0	0.0	0.0	0.0	0.086	0.181	0.212	0.425	0.191	0.452
EFI	0.064	0.308	0.927	0.472	0.313	0.591	0.341	0.793	0.308	0.250	0.260	0.242	0.425	0.278	0.452

NIVEAU 6
****

Causes du retard

DH = 0.754 K1 = 2.780 K2 = 1.273 K3 = 1.659 SOM30 = 4640 . SOM90 = 5684

CSP	6	7	6	3	7	7	6	6	7	5	6	7	4	2	7
NAT	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3
TMERE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NBENF	1	1	2	1	2	2	2	1	1	1	1	2	1	1	2
MILIEU	1	1	1	1	1	2	2	2	2	1	1	3	1	1	2
RAISONS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

****

NFI	647	446	336	230	221	219	212	199	196	181	157	132	121	107	103
TFI	0.895	0.883	0.855	0.943	0.804	0.709	0.763	0.783	0.734	0.858	0.946	0.552	0.883	0.973	0.880
UFI	0.068	0.047	0.035	0.024	0.023	0.023	0.022	0.021	0.021	0.019	0.016	0.014	0.013	0.011	0.011
PFI	0.130	0.137	0.283	0.104	0.303	0.416	0.368	0.327	0.291	0.072	0.096	0.439	0.107	0.047	0.544
RANG	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
GDM	0.870	0.782	0.717	0.896	0.697	0.584	0.632	0.623	0.709	0.928	0.904	0.538	0.893	0.953	0.456
PDM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
EFI	0.870	0.782	0.717	0.896	0.697	0.584	0.632	0.623	0.709	0.928	0.904	0.538	0.893	0.953	0.456

CSP	7	7	6	6	6	3	7	7	3	7	5	6	7	7	6
NAT	1	1	3	1	1	1	3	1	1	1	1	3	1	1	1
TMERE	1	1	1	2	1	1	1	1	2	1	1	1	1	1	1
NBENF	3	2	1	1	2	2	2	3	1	3	2	2	1	2	3
MILIEU	3	3	1	1	3	1	1	2	1	3	1	2	3	2	2
RAISONS	1	3	1	1	1	1	1	1	1	3	1	1	1	3	1

****

NFI	90	89	88	86	83	81	80	78	77	75	71	67	60	58	50
TFI	0.495	0.372	0.917	0.905	0.601	0.931	0.930	0.600	0.951	0.412	0.807	0.882	0.600	0.188	0.633
UFI	0.009	0.009	0.009	0.009	0.009	0.009	0.008	0.008	0.008	0.008	0.007	0.007	0.006	0.006	0.005
PFI	0.467	0.978	0.205	0.163	0.373	0.160	0.300	0.474	0.052	0.987	0.239	0.448	0.517	1.000	0.500
RANG	3	1	3	3	3	3	3	3	3	1	3	3	3	1	3
GDM	0.520	0.0	0.795	0.837	0.627	0.840	0.700	0.526	0.948	0.0	0.761	0.552	0.483	0.0	0.500
PDM	0.0	0.538	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.520	0.0	0.0	0.0	0.584	0.0
EFI	0.520	0.538	0.795	0.837	0.627	0.840	0.700	0.526	0.948	0.520	0.761	0.552	0.483	0.584	0.500