PANEL SOCIO-ECONOMIQUE

"LIEWEN ZU LËTZEBUERG"

DOCUMENT PSELL N° 53

DISPOSITIF DES PONDERATIONS INDIVIDUELLES ET DES PONDERATIONS DES MENAGES DE 1985 A 1990

par

Bernard GAILLY

CEPS/Instead
Differdange
Grand-Duché de Luxembourg

1990

DISPOSITIF DES PONDÉRATIONS INDIVIDUELLES ET DES PONDÉRATIONS DES MÉNAGES de 1985 à 1989

B.Gailly
et équipe P.S.E.L.L.

Ce document s'adresse aux utilisateurs des données collectées par le Panel Socio-Economique Luxembourgeois et à tous ceux qui seront appelés à pondérer les échantillons successifs de cette étude.

Il a été réalisé grâce à la collaboration de A.Wagner, S.Breulheid, R. De Wever, P.Hausman, F.Hentges, A.Kerger, G.Schmaus et J.J.Wester. Madame M.Belche en a assuré la dactylographie.

Nos remerciements s'adressent particulièrement à **Greg DUNCAN**. Il n'a ménagé ni son temps, ni sa peine, pour nous introduire dans les arcanes du dispositif de pondération du P.S.I.D. (Ann Arbor, Michigan). A quelques détails près, les techniques de pondération appliquées dans le P.S.E.L.L. sont empruntées au nouveau dispositif adopté par le P.S.I.D.

SOMMAIRE

INTRODUCTION 11

DEUX RAISONS DE PONDÉRER L'ÉCHANTILLON 13

Chapitre 1: Deux raisons de pondérer l'échantillon 15

ORGANISATION DU FICHIER 19

Chapitre 2: Organisation du fichier 21

2.1. Les personnes 22

2.1.1. Les membres et les non-membres d'un échantillon 22

2.1.2. Les membres et les non-membres du panel 23

2.1.3. Eléments des trajectoires individuelles 24

2.1.4. Les variables DIA-gnostic 28

2.2. Les variables de pondération 31

2.2.1. Processus de calcul: étapes et aspects spécifiques 32

2.2.2. Schéma récapitulatif 50

2.2.3. La valeur d'usage des variables 51

2.2.4. Les poids des ménages 52

2.3. Echantillons et variables de pondération 54

ÉCHANTILLON 1985 ET VARIABLES DE PONDÉRATION 55

Chapitre 3: **Échantillon 1985 et variables de pondération** 57
3.1. **L'échantillon initial** 57
3.1.1. Unité de sélection et population de référence 57
3.1.2. Constitution du fichier de référence 58
3.1.3. Sélection de l'échantillon 61
3.1.4. Les unités d'analyse 64
3.1.5. Des biais d'échantillonnage? 68
3.2. **Variables de pondération** 74
3.2.1. Variable instrumentale "MIREV" 74
3.2.2. Variable instrumentale "PONREV" 75
3.2.3. Variable opérationnelle de pondération des ménages en 1985 "MWGT85A" 78
3.2.4. Variable opérationnelle de pondération individuelle en 1985 "WGT85" 79

ANNEXE: Modification de la distribution des individus, en fonction du nombre de titulaires principaux dans le ménage 81

ÉCHANTILLON 1986 ET VARIABLES DE PONDÉRATION 85

Chapitre 4: **Échantillon 1986 et variables de pondération** 87
4.1. **L'échantillon 1986** 88
4.1.1. Les anciens membres du panel 89
4.1.2. Les refus 89
4.1.3. Les émigrés et les décès 90
4.1.4. Les nouveaux membres du panel 90
4.1.5. Les nouveaux non-membres du panel 91
4.2. **Les variables de pondération** 92
4.2.1. Variable instrumentale "TREE" 92
4.2.2. Variable instrumentale "W" 101
4.2.3. Variable instrumentale "W86" 103
4.2.4. Variable opérationnelle "WGT86" 104
4.2.5. Variable instrumentale "W86B" 106
4.2.6. Variable opérationnelle "WGT86TOT" 107
4.2.7. Variable de pondération des ménages en 1986 "MWGT86A" 108

ANNEXE 4A: Taux de survie dans le panel entre 1985 et 1986: contrastes entre les taux de réponses de 18 segments de l'échantillon 111

ANNEXE 4B: Exemple de l'évolution des individus et des ménages 114

ANNEXE 4C: Tableau comparatif des caractéristiques démographiques des membres du panel en 1985 et en 1986 (après pondération) 117

ÉCHANTILLON 1987 ET VARIABLES DE PONDÉRATION 121

Chapitre 5: **Échantillon 1987 et variables de pondération** 123
5.1. **L'échantillon 1987** 124
5.1.1. Les trajectoires individuelles: version courte 124
5.1.2. Les trajectoires individuelles: version complète 128
5.2. **Les variables de pondération** 131
5.2.1. Variable instrumentale TREE87 131
5.2.2. Variable instrumentale RR87 144
5.2.3. Variable instrumentale W87 145
5.2.4. Variable instrumentale W87B 147
5.2.5. Variable opérationnelle WGT87TOT 148
5.2.6. Variable opérationnelle de pondération des ménages en 1987 MWGT87A 150

ANNEXE 5A: Tableau comparatif des caractéristiques démographiques des membres du panel en 1985, en 1986 et en 1987 (après pondération) 153

ÉCHANTILLON 1988 ET VARIABLES DE PONDÉRATION 157

Chapitre 6: **Échantillon 1988 et variables de pondération** 159
6.1. **L'échantillon 1988** 160
6.1.1. Les trajectoires individuelles: version courte 160
6.1.2. Les trajectoires individuelles: version complète 164
6.2. **Les variables de pondération** 168
6.2.1. Variable instrumentale TREE88 168

6.2.2. Variable instrumentale RR88 182
6.2.3. Variable instrumentale W88 184
6.2.4. Variable instrumentale W88B 185
6.2.5. Variable opérationnelle WGT88TOT 187
6.2.6. Variable opérationnelle de pondération des ménages en 1988 MWGT88A 188

ANNEXE 6A: Tableau comparatif des caractéristiques démographiques des membres du panel en 1985, en 1986, en 1987 et en 1988 (après pondération) 192

ÉCHANTILLON 1989 ET VARIABLES DE PONDÉRATION 197

Chapitre 7: **Échantillon 1989 et variables de pondération** 199
7.1. L'échantillon 1989 200
7.1.1. Les trajectoires individuelles: version courte 200
7.1.2. Les trajectoires individuelles: version complète 204
7.2. Les variables de pondération 207
7.2.1. Variable instrumentale TREE89 207
7.2.2. Variable instrumentale RR89 219
7.2.3. Variable instrumentale W89 221
7.2.4. Variable instrumentale W89B 222
7.2.5. Variable opérationnelle WGT89TOT 224
7.2.6. Variable opérationnelle de pondération des ménages en 1988 MWGT89A 225

ANNEXE 7A: Tableau comparatif des caractéristiques démographiques des membres du panel en 1985, en 1986, en 1987, en 1988 et en 1989 (après pondération) 229

Chapitre 8: **Échantillon 1990 et variables de pondération** 235
8.1. **L'échantillon 1990** 236
8.1.1. Les trajectoires individuelles: version courte 236
8.1.2. Les trajectoires individuelles: version complète 240
8.2. **Les variables de pondération** 243
8.2.1. Variable instrumentale TREE90 243
8.2.2. Variable instrumentale RR90 256
8.2.3. Variable instrumentale W90 258
8.2.4. Variable instrumentale W90B 259
8.2.5. Variable opérationnelle WGT90TOT 261
8.2.6. Variable opérationnelle de pondération des ménages en 1988 MWGT90A 262

ANNEXE 8A: Tableau comparatif des caractéristiques démographiques des membres du panel en 1985, en 1986, en 1987, en 1988, en 1989 et en 1990 (après pondération) 266

TABLEAUX DE SYNTHESE DES VARIABLES DE PONDÉRATION ET DES ÉCHANTILLONS PERTINENTS 271

DISPOSITIF DES PONDÉRATIONS INDIVIDUELLES ET DES PONDÉRATIONS DES MÉNAGES de 1985 à 1989

Après un bref rappel des différentes raisons qui justifient l'emploi de facteurs de pondération dans l'analyse des données, ce document présente l'organisation du fichier dans lequel les principales variables de pondération sont répertoriées.

La troisième partie du document est consacrée à la présentation de ces variables. Elles sont décrites systématiquement sous trois aspects: leur fonction, leur contenu et les catégories de personnes éligibles.

Ces variables ont des fonctions différentes dans l'ensemble du système de pondération. Les unes ont un rôle "opérationnel": elles peuvent être utilisées pour corriger les échantillons. Les autres sont des variables "instrumentales" qui ont permis ou permettront de calculer des variables "opérationnelles".

Le "contenu" de ces variables correspond à leur utilité dans le système de pondération; il résulte des opérations qui ont permis de les calculer.

Les catégories de personnes "éligibles" définissent les limites de la pertinence de chaque variable.

Le fichier individuel longitudinal réunit l'ensemble des variables de pondération et l'ensemble des personnes interrogées au moins une fois dans le cadre du panel. La sélection des sous-échantillons détermine le choix des variables de pondération.

Cette version a été structurée de telle façon que les versions suivantes, apportant des éléments de documentation complémentaires ou des informations relatives aux vagues d'enquête successives, ne soient que des développements progressifs d'un même document. Les actualisations périodiques du contenu compléteront systématiquement la documentation présentée dans les éditions précédentes. Les utilisateurs auront ainsi un

accès plus aisé à l'ensemble des informations nécessaires pour pouvoir manier le système des pondérations en fonction de leurs objectifs.

Un tableau synoptique permet de visualiser rapidement les sous-échantillons éligibles et les variables de pondération pertinentes pour chaque échantillon. Il sera mis à jour annuellement. Chaque enquête annuelle modifie l'état de l'échantillon, des nouvelles variables de pondération doivent être calculées et insérées dans le fichier. Le tableau synoptique des sous-échantillons et des variables de pondération suivra donc cette évolution.

DEUX RAISONS DE PONDERER

L'ECHANTILLON

Chapitre 1

DEUX RAISONS DE PONDERER L'ECHANTILLON

Le Panel Socio-Economique Luxembourgeois (PSELL) construit une base de données individuelles temporelles. Ces données décrivent l'état et l'évolution des conditions d'existence des personnes et des ménages résidant au Luxembourg.

La population de référence de l'étude

- inclut

- tous les individus résidant au Luxembourg et bénéficiaires de la Sécurité Sociale ou de la Protection sociale,

- tous les individus membres du ménage de ces bénéficiaires de la Sécurité sociale ou de la Protection sociale à condition qu'ils soient:

  - présents dans le ménage au moment de l'enquête

  - absents temporairement au moment de l'enquête (hospitalisation, prison,...);

- exclut

- les individus bénéficiaires de la Sécurité sociale luxembourgeoise, mais résidant à l'étranger

- les résidents étrangers non assurés à la Sécurité sociale luxembourgeoise et non bénéficiaires de la Protection sociale luxembourgeoise

- les individus installés définitivement dans des ménages collectifs.

Cette base de données offre le choix entre plusieurs unités d'analyse. Les personnes et les ménages seront sans doute les unités les plus couramment utilisées mais la base de données permet de saisir également d'autres unités d'analyse.

Ces unités sont observées annuellement et les données recueillies au cours de chaque vague d'enquête permettent d'effectuer deux types d'analyse. D'une part, les données annuelles peuvent être analysées dans une perspective synchronique; l'analyse vise alors à refléter aussi fidèlement que possible la situation actuelle dans le pays. D'autre part, les échantillons sont liés progressivement les uns aux autres afin de constituer des séries temporelles individuelles. Ce dispositif permet de procéder à des analyses longitudinales. L'objectif est, alors, de progresser dans la connaissance des processus socio-économiques plutôt que de "refléter" la situation générale de la population de référence.

Chacune de ces approches pose des problèmes techniques et méthodologiques particuliers.

L'analyse synchronique pose des problèmes liés à la nature même de ses objectifs. En particulier, elle exige que l'échantillon ne soit pas affecté par des biais systématiques qui invalideraient les estimateurs statistiques et ne permettraient plus d'établir la marge d'erreur qui encadre toute valeur observée ou estimée sur la base de l'échantillon. L'échantillon doit être "représentatif" de la population-cible.

Mais l'apparition de tels biais est pratiquement inévitable lorsque l'observation est diachronique et porte sur un échantillon composé de personnes interrogées à maintes reprises. En effet, il est peu probable que deux propriétés élémentaires de l'échantillon soient respectées d'une période d'observation à l'autre :

- tous les éléments de l'échantillon doivent conserver la même probabilité de sélection d'une observation à l'autre

- tous les sous-ensembles d'éléments de l'échantillon (toutes les catégories de population) doivent conserver la même probabilité de sélection d'une observation à l'autre.

En outre, les catégories de la population que l'on souhaite analyser avec une attention particulière doivent conserver une taille suffisante au fil des années (G. DUNCAN - 1989).

L'usage des pondérations vise à limiter les effets de cette usure de l'échantillon :

- en détectant les sources des biais

- en estimant l'ampleur des biais

- en compensant les déficits qui ont pu être repérés.

Deux types de phénomènes favorisent l'apparition de biais, lorsque le même échantillon est observé à plusieurs reprises:

- les premiers sont le fait de l'échantillonnage:
  - le mode de tirage de l'échantillon
  - les refus de participer à l'enquête
  - l'entrée de nouvelles personnes dans l'échantillon;
- les seconds correspondent à des événements démographiques; ils s'observent également dans la population générale:
  - des naissances
  - des décès
  - des mouvements migratoires.

Tous ces phénomènes contribuent à modifier la structure de l'échantillon. Ils doivent être pris en compte dans le calcul des facteurs de pondération.

Il est utile de décrire, en premier lieu, l'organisation du fichier individuel où les variables de pondération sont et seront répertoriées.

ORGANISATION DU FICHIER

Chapitre 2

ORGANISATION DU FICHIER

Le répertoire des variables de pondération réunit l'ensemble des PERSONNES interrogées au moins une fois au cours du panel et l'ensemble des VARIABLES de pondération.

1. Différentes catégories de personnes peuvent être distinguées en fonction de leur trajectoire au sein du panel. Toutes ces trajectoires peuvent être décrites à partir d'une matrice élémentaire basée sur une double distinction: les membres et les non-membres d'un échantillon annuel, d'une part, les membres et les non-membres du panel, d'autre part.

2. Des distinctions supplémentaires doivent être introduites. Certaines catégories de personnes jouent un rôle particulier dans l'évolution de l'échantillon et dans le calcul des variables de pondération. Elles doivent donc être identifiables dans le fichier.

3. Les variables "DIA-gnostic" permettent précisément de sélectionner ces sous-échantillons et de reconstituer les échantillons qui feront l'objet des analyses.

Chaque fois qu'une nouvelle vague d'enquêtes est introduite dans le fichier, une nouvelle variable "diagnostic" est insérée. Elle permet de partitionner l'ensemble du fichier en sous-échantillons. Ces sous-échantillons correspondent à l'ensemble des trajectoires individuelles observées entre la première vague d'enquête et la dernière vague introduite dans le fichier.

4. Toutes les variables de pondération répertoriées dans le fichier n'ont pas la même fonction. On distinguera des variables "opérationnelles" et des variables "instrumentales" en précisant les indications de leur usage.

2.1. LES PERSONNES

2.1.1. Les membres et les non-membres d'un échantillon

Le fichier total est un fichier évolutif et cumulatif. Chaque année les résultats d'une nouvelle enquête viennent s'ajouter aux données enregistrées au cours des années précédentes. Toutes les personnes présentes dans un échantillon ne sont plus nécessairement présentes dans l'échantillon suivant. Inversement, chaque année, de nouvelles personnes participent à l'étude.

Le fichier est évolutif:

* des **nouvelles personnes entrent** dans l'échantillon et dans le fichier à l'occasion de chaque vague d'enquêtes. Elles ont rejoint, selon des modalités variables, un ménage ou une personne présente dans l'échantillon précédent;

* des **personnes quittent** l'échantillon provisoirement ou définitivement pour des raisons très diverses (absence du ménage, refus de répondre, décès, émigration, ...).

Le fichier est cumulatif:

* il recense systématiquement toutes les personnes contactées, au moins une fois au fil des vagues d'enquêtes.

Par conséquent,

* le nombre de personnes recensées augmente chaque année,

* le nombre de personnes présentes dans le fichier ne diminuera jamais,

* toutes les personnes recensées dans le fichier n'auront pas nécessairement des "valeurs" ou observations enregistrées pour chaque année . Le numéro d'identification de TOUTES les personnes reste néanmoins enregistré parce que certaines personnes peuvent rejoindre l'échantillon après une absence de plus ou moins longue durée.

Le fichier est donc formé de deux grandes catégories de personnes:
- les **membres d'un échantillon**; ils ont été interrogés au cours d'une vague d'enquêtes donnée;
- les **non-membres d'un échantillon**; ils n'ont pas été interrogés au cours de cette vague d'enquêtes.

2.1.2. Les membres et non-membres du panel

Toute personne répertoriée dans le fichier n'est pas nécessairement membre du panel.

Une personne est **membre du panel** lorsqu'elle remplit l'une des deux conditions suivantes.

* Elle appartient à l'échantillon initial tiré en 1985. Les 6110 personnes appartenant à cet échantillon sont des membres du panel.

* Elle est entrée dans un échantillon du panel après la première vague d'enquêtes; elle descend en ligne directe d'un membre du panel (père et/ou mère); elle est entrée simultanément dans l'échantillon et dans la population.

  (Il faut toutefois signaler une exception: quelques personnes sont entrées dans l'échantillon au cours de la deuxième vague d'enquêtes, au titre de fils ou de fille d'un membre du panel. Elles sont rentrées dans leur famille après une brève absence qui correspondait au moment de l'enquête. Elles ont été considérées comme membres du panel bien u'elles ne soient pas entrées simultanément dans l'échantillon et dans la population).

  En 1986, 78 personnes sont entrées dans l'échantillon au titre de membres du panel.

  En 1987, 65 personnes sont entrées dans l'échantillon au titre de membres du panel: elles sont nées pendant l'année qui sépare les deux vagues d'enquête.

Toute autre personne est définie comme **non-membre du panel**. Elle peut être l'ascendant direct, le conjoint, l'oncle ou l'ami d'un membre du panel. Elle entre dans l'échantillon parce qu'elle se joint à un ménage ou à une personne présente au préalable.

Cette définition est très large: elle n'exclut pas qu'une personne non-membre du panel se sépare du ménage ou de la personne qui l'a accueillie et accueille à son tour une nouvelle personne non-membre du panel. Des ménages non-membres du panel peuvent ainsi se former. Ils continueront à participer à l'enquête et seront toujours répertoriés dans le fichier.

Leur rôle dans le calcul des pondérations annuelles devra être précisé par la suite.

2.1.3. Eléments des trajectoires individuelles

Le fichier est organisé à partir de quatre éléments. La combinaison de ces éléments constitue des trajectoires individuelles. Ces trajectoires retracent la carrière des personnes au fil des échantillons successifs. Elles déterminent leur rôle dans la procédure de pondération.

SCHEMA I.

Eléments des trajectoires individuelles

	statut dans l'échantillon	
statut dans le panel	membre	non-membre
membre	A	B
non-membre	C	D

Le schéma I. permet de distinguer quatre catégories de personnes. Celles-ci sont définies en fonction de leur statut dans le panel et dans un échantillon donné.

La trajectoire d'un individu dans le fichier (et dans le panel) correspond à la série des états qu'il occupe successivement dans les échantillons annuels.

Chacun de ces états définit un ensemble de personnes ayant un rôle particulier à jouer dans le calcul des pondérations.

Case A

Cette catégorie rassemble les personnes membres du panel lorsqu'elles sont présentes dans un échantillon.

Lorsqu'elles occupent cette position, elles sont prises en compte dans la procédure de pondération.

Ce sont les seules personnes qui reçoivent une valeur de pondération. En effet, ce sont les seules personnes présentes dans l'échantillon. Leur probabilité de sélection initiale est connue. Il est donc possible de calculer leur nouvelle probabilité de sélection, pour une année donnée: cette nouvelle probabilité dépend du taux de réponse observé cette année-là pour l'ensemble des membres du panel.

Ce taux de réponse est calculé chaque année par rapport à l'ensemble des membres du panel. En d'autres termes, il n'est pas seulement calculé par référence aux membres du panel appartenant à l'échantillon initial; ce taux de réponse prend également en compte les membres du panel entrés dans l'échantillon au cours des vagues d'enquête successives.

* Les taux de réponse doivent être calculés par référence à *l'état de l'échantillon de la première année d'observation*; cette précaution permet de réajuster l'échantillon observé sur le profil de l'échantillon aléatoire initial; elle permet également de prendre en compte chaque année des personnes qui rejoignent l'échantillon après une absence plus ou moins prolongée.

Cette mobilité des personnes ne pourrait plus être prise en compte, si les taux de réponse étaient calculés par référence à l'état de l'échantillon de l'année précédente. Toute personne ayant quitté l'échantillon, à un moment donné, serait définitivement écartée de l'étude, puisqu'il serait impossible de calculer la probabilité de répondre au temps t d'une personne absente au temps t-1.

Les effets liés à l'usure de l'échantillon peuvent être limités, en prenant l'échantillon initial comme base de calcul des taux de réponse.

* L'état de l'échantillon de la première année d'observation doit être modifié chaque année; *les nouveaux membres du panel doivent être insérés dans cet ensemble de référence.*

Une probabilité de sélection peut être attribuée aux nouveaux membres du panel. Leur entrée dans l'échantillon dépend directement de la probabilité de sélection de leurs parents. Les nouveaux-nés reçoivent donc un poids équivalent à la moyenne des poids calculés pour leurs parents au cours de l'année de la naissance. Ils reçoivent donc le même poids que leurs parents lorsque le père et la mère ont des poids identiques. Ils reçoivent le poids de l'ascendant présent dans l'échantillon lorsque l'un des conjoints est absent.

La probabilité de sélection de ces nouveaux membres est ensuite calculée chaque année selon la procédure générale appliquée à l'ensemble des membres du panel.

Cette modification de l'échantillon de référence permet de de prendre en compte l'évolution démographique de la population de référence.

Ces personnes situées dans la case A, ne passeront jamais à l'état C ni à l'état D au cours de leur carrière. Les membres du panel conserveront ce statut définitivement. Par contre, il est possible que certains membres du panel quittent l'échantillon (provisoirement ou définitivement): la seule transition possible conduirait ces personnes vers la case B.

Case B

Cette catégorie rassemble les personnes membres du panel lorsqu'elles sont absentes d'un échantillon annuel.

Trois cas peuvent se présenter.

* **Ces personnes refusent de répondre, sont absentes ou introuvables.** Dans ce premier cas, elles sont prises en compte dans le calcul des taux de réponses qui détermineront toute la procédure de calcul des pondérations. La probabilité de sélection initiale de ces personnes étant connue, il est possible d'évaluer l'impact de leur disparition sur le profil de l'échantillon.

Ces personnes peuvent rejoindre l'échantillon au cours d'une vague d'enquête ultérieure. Elles seront à nouveau membres de l'échantillon et reçevront à nouveau un poids relatif. Elles rejoignent les membres de la case A.

* **Ces personnes ont émigré.** Dans ce deuxième cas, elles ont quitté simultanément la population de référence et l'échantillon. Elles ne sont pas prises en compte dans le calcul des taux de réponses. Leur départ ne modifie pas le rapport entre l'échantillon et la population de référence.

Ces personnes peuvent revenir au Luxembourg et rejoindre l'échantillon après une ou plusieurs années. Dans ce cas, elles rentrent dans la case A. Elles sont à nouveau prises en compte dans le calcul des taux de réponses. Elles reçoivent un poids relatif.

* **Ces personnes sont décédées:** elles ont quitté la population de référence et l'échantillon simultanément. Leur départ n'influence pas la représentativité de l'échantillon. Toutefois, les taux de mortalité cumulés dans l'échantillon et dans la population peuvent diverger sensiblement après quelques années. Dans ce cas, l'échantillon sera ajusté à intervalles réguliers.

Les deux premières catégories de personnes ne passeront jamais vers l'état C ou vers l'état D. Elles resteront définitivement membres du panel. La seule transition possible les conduirait à rejoindre la catégorie A et à réduire le taux d'usure de l'échantillon.

Case C

Cette catégorie correspond aux personnes non-membres du panel, lorsqu'elles sont présentes dans un échantillon.

Toutes ces personnes sont des nouveaux membres. Par définition, aucune d'entr'elles n'appartient au premier échantillon alors qu'elles appartenaient à la population de référence au moment du tirage de l'échantillon initial (hormis les personnes immigrées après 1985 et les enfants nouveaux-nés issus de deux parents membres de l'échantillon mais non-membres du panel).

Leur probabilité de sélection, au cours d'une vague d'enquête, ne peut être établie que sur la base d'hypothèses invérifiables. Elles ne sont donc pas prises en compte dans le calcul des taux de réponses.

Ces personnes reçoivent un poids de '0', lorsqu'elles sont présentes dans l'échantillon.

Cette valeur de pondération n'exerce aucune influence sur le profil de l'échantillon mais elle provoque quatre effets.

* Ces personnes ne sont pas prises en compte lorsque l'analyse porte sur l'échantillon pondéré.

* Ces personnes sont prises en compte lorsque l'analyse porte sur l'échantillon non-pondéré.

* Ces personnes sont prises en compte lorsque l'analyse porte sur des données temporelles, à condition qu'elles appartiennent aux échantillons couverts par la période analysée. (A notre connaissance, les analyses de données individuelles temporelles s'effectuent sur des échantillons non-pondérés).

* Le "poids relatif " de ces personnes intervient d'une part, au moment de calculer le poids relatif des nouveaux membres du panel (moyenne du poids des parents) (2.2.1.7.) et d'autre part, au moment de calculer le poids relatif des ménages (Moyenne des poids des membres du ménage. Voir 2.2.).

La seule transition possible pour les membres de cette catégorie les conduirait dans la case D. Ils ne seront jamais des membres du panel. Ils n'interviendront jamais dans le calcul des taux de réponses annuels. Leur trajectoire individuelle sera entièrement définie par leur stabilité dans la case C ou par des mouvements plus ou moins fréquents entre les états C et D.

Case D

Cette case n'est pas vide. Elle définit le statut des personnes entrées dans l'échantillon au titre de non-membres du panel lorsque ces personnes quittent l'échantillon.

Non-membres du panel, ces personnes ne sont pas prises en compte dans le calcul des taux de réponses annuels.

Absentes de l'échantillon, elles n'ont plus aucune influence sur les calculs des poids relatifs ni au niveau des individus ni au niveau des ménages.

Ces personnes ne peuvent connaître qu'un seul type de transition: rejoindre l'échantillon. Leur trajectoire sera définie soit par leur absence définitive de l'étude, soit par des retours plus ou moins prolongés dans la catégorie C.

2.1.4. Les variables DIA-gnostic

Une variable "DIA-gnostic" est construite annuellement sur la base de cette matrice.

Cette variable est la "clé" du fichier.

EXEMPLE: **variable: DIA (1985 - 1986)**

codes	trajectoires	fichier: état 1986 n =
0.	memb.85/refus86	833
1.	memb.85/memb.86	5167
2.	nouv.mem. pan.86	78
3.	nouv.mem.non pan. 86	146
4.	memb.85/décédés ou émigrés 86	110
TOTAL		6334

Chaque modalité de cette variable correspond à la trajectoire d'un ensemble de personnes dans le fichier et dans le panel entre la première vague d'enquête et l'année considérée.

En 1985, la variable "DIA-gnostic" est sans objet; toutes les personnes recensées dans le fichier sont membres du panel et membres de l'échantillon.

En 1986, la variable distingue cinq vecteurs:

(0) les membres de 1985 qui refusent de répondre en 1986,
(1) les membres de 1985 qui répondent en 1986,
(2) les nouveaux membres du panel entrés dans l'échantillon en 1986,
(3) les nouveaux membres qui n'appartiennent pas au panel et qui entrent dans l'échantillon en 1986,
(4) les membres du panel présents dans l'échantillon en 1985 qui ont émigré entre les deux vagues d'enquête ou qui sont décédés au cours de cette période.

Chaque trajectoire est définie par la succession des statuts que ces personnes ont occupé au cours de la période considérée.

ex. membre panel 85/ émigré 86 / membre panel 87 / décédé 88
---- / nouv.mem.86/ membre pan.87 / mem.pan.88

Des modalités supplémentaires apparaissent chaque fois qu'un nouvel échantillon annuel est inséré dans le fichier: l'ensemble des trajectoires se diversifie.

Cette variable risque de devenir rapidement complexe et d'un maniement difficile. Une seconde variable propose chaque année une version simplifiée (à partir de la troisième vague d'enquête): cette variable simplifiée reprend uniquement les éléments définissant le statut des personnes dans le système de pondération.

En 1986, la version simplifiée et la version élargie coïncident.

A partir de 1987, la version simplifiée ne prend plus en compte que les termes extrêmes des trajectoires:

- à quel titre les personnes sont-elles entrées dans le fichier?

- quel est leur statut dans l'échantillon choisi?

Ces variables "DIA-gnostic" permettent de gérer aisément le fichier, de sélectionner les échantillons et les sous-échantillons que l'on souhaite analyser voire pondérer.

Dans l'exemple proposé:

* la sélection des groupes #0, #1 et #4 réunit l'ensemble des membres de l'échantillon interrogés en 1985, soit 6110 personnes;
* la sélection des groupes #1, #2 et #3 réunit l'ensemble des personnes interrogées en 1986;

* la sélection des vecteurs #0 et #1 permet de calculer les taux de réponse des membres du panel en 1986 et d'élaborer le dispositif de pondération de l'échantillon observé en 1986;

* la sélection du groupe #1 permet de saisir l'ensemble des personnes interrogées à deux reprises: ce type d'échantillon se prêtera aux analyses de données individuelles temporelles;

* la sélection des groupes #0, #1 et #4 permet de reconstituer l'ensemble des personnes interrogées en 1985 et de changer de niveau d'analyse. Chaque personne est désignée, dans le fichier, par un numéro d'identification. Ce numéro est strictement individuel et reste identique d'années en années. Chaque personne reçoit également un numéro correspondant au numéro d'identification du ménage auquel elle appartient. Ce numéro d'identification du ménage est identique pour toutes les personnes appartenant au même ménage. Il est donc possible de procéder à l'agrégation des valeurs individuelles au niveau du ménage et, inversement, de ventiler des valeurs observées directement au niveau du ménage sur chacun des membres de ce ménage.

2.2. LES VARIABLES DE PONDERATION

Chaque *chapitre* présente l'ensemble du système de pondération appliqué à *une vague d'enquête*.

Tous les chapitres sont structurés de la même manière. L'exposé suit la logique du calcul des pondérations: chaque *étape* de ce processus correspond à la création d'une *variable intermédiaire*. Toutes les variables ou étapes du processus de calcul sont insérées dans le fichier chaque année.

Les règles élémentaires du calcul des pondérations d'un échantillon probabiliste ne seront pas rappelées chaque fois qu'elles seront mises en oeuvre.

Par contre, le caractère longitudinal de l'étude exige que certaines règles d'ajustement de l'échantillon annuel soient précisées. L'échantillon est soumis à des déformations qui n'apparaissent pas dans une étude synchronique. Ces règles seront évoquées régulièrement. Cette précaution s'impose parce que toutes les études longitudinales n'appliquent pas nécessairement les mêmes règles.

* Les procédures permettant de gérer les *sources de biais spécifiques* sont donc décrites dans cette introduction. Ces règles sont, bien entendu, rappelées régulièrement lorsqu'elles sont mises en oeuvre (**2.2.1.**).

* Un schéma, placé en tête de chaque chapitre, récapitule l'ensemble de la procédure et permet à l'utilisateur de trouver rapidement la variable qui l'intéresse en consultant uniquement ce schéma (**2.2.2.**).

* Ces variables sont dites "*instrumentales*" ou "*opérationnelles*" en fonction de leur rôle dans le système de pondération (**2.2.3.**).

* Enfin, les *poids calculés pour les ménages* ne peuvent pas être utilisés sans quelques manipulations préalables. Ces poids ont été reventilés sur les membres du ménage afin de rassembler toutes les valeurs de pondération dans un seul fichier individuel (**2.2.4.**).

2.2.1. Processus de calcul: étapes et aspects spécifiques

2.2.1.1. *Première étape*

La première étape de cette démarche consiste à calculer les probabilités de sélection des unités d'analyse. La première année ces probabilités dépendent essentiellement de la position des unités d'analyse dans la *procédure d'échantillonnage*. Les vagues d'observation des années suivantes sont soumises à d'autres aléas. Les probabilités de sélection des unités d'analyse sont modifiées par la *répartition inégale des taux de réponse* dans l'échantillon.

Particularité de l'échantillon initial

L'échantillon initial de cette étude, est un échantillon probabiliste: chaque unité d'analyse a une chance connue et différente de zéro d'apparaître dans l'échantillon. La théorie statistique permet d'établir les propriétés des estimations effectuées dans ce type d'échantillon. Il est donc possible de procéder à des inférences statistiques.

Les unités de tirage sont sélectionnées par la méthode de l'échantillonage aléatoire simple: chaque élément de la population a la même probabilité d'être tiré; chaque combinaison d'éléments a la même probabilité d'être tiré.

Les estimations statistiques n'en sont pas pour autant à l'abri de certains biais.

Le mode de tirage de l'échantillon initial pose un problème à cet égard: l'unité de *tirage*, l'unité d'*observation* (le ménage) et les différents types d'unités d'*analyse* utilisables (individus, ménages, groupes de revenus) ne sont pas identiques.

L'échantillon des unités de tirage ne présente pas de biais systématique mais les ménages et les membres des ménages observés ne forment pas des échantillons aléatoires simples.

Ce problème peut être résolu.

En effet, il est possible d'identifier les liens qui existent entre l'unité de tirage, l'unité d'observation (le ménage) et les unités d'analyse (ménages, individus).

L'identification de ces liens permet de dériver les probabilités de sélection des unités d'analyse.

Ces probabilités étant connues, l'équiprobabilité des unités d'analyse peut être restaurée.

La pondération des unités d'analyse produit exactement cet effet: elle consiste à accorder à chaque unité d'analyse un poids relatif. Toutes les unités ne sont plus équivalentes à une unité. Chaque unité vaut une fraction ou un multiple de l'unité. Cette valeur s'obtient en divisant chaque unité par sa probabilité de sélection.

Ces différents aspects sont traités de manière détaillée dans le chapitre 3 (**3.1.**).

Généralisation aux vagues suivantes

A partir de la deuxième vague d'enquête, les déformations de l'échantillon sont liées à deux facteurs. Le premier est propre à la vie de l'échantillon: les refus de répondre. Le second reproduit dans l'échantillon les conditions d'évolution de la population de référence (naissances, émigration, décès).

Dans un premier temps, toutes les personnes, entrées à un moment quelconque dans l'étude, doivent être classées en fonction de leur position par rapport à ces deux facteurs.

* Les facteurs démographiques ajustent automatiquement le profil de l'échantillon au profil de la population de référence: les personnes émigrées et les personnes décédées peuvent donc être tenues à l'écart du calcul des nouvelles probabilités de sélection.

* La probabilité de sélection des non-membres du panel est inconnue: ces personnes ne sont donc pas prises en compte dans l'analyse de l'évolution de l'échantillon.

* Les enfants nés de parents membres du panel sont définis comme étant des membres du panel. Ils forment une catégorie particulière. Ils recevront, au terme de la procédure, un poids égal à la moyenne des poids qui auront été attribués à leurs parents pour l'année considérée.

* Un seul facteur reste à mesurer: l'importance relative des refus de répondre.

Le taux de réponse correspond au rapport entre

- le nombre de membres du panel présents dans l'échantillon

- le nombre total des membres du panel.

Les nouveaux membres du panel entrés dans l'échantillon au cours de la vague d'enquête considérée n'entrent pas dans le calcul des taux de réponses.

Tous les membres du panel entrés au cours des vagues précédentes sont pris en compte, excepté: les personnes émigrées et les personnes décédées.

Tous ces membres sont déjà affectés d'un poids: le poids relatif qui leur a été attribué au moment de leur entrée dans le panel (**2.1.3. Case A**). Cette précaution s'impose étant donné le biais de sélection lié au mode de tirage de l'échantillon initial.

Si le taux de réponse était uniforme, quelle que soit la manière de partitionner l'échantillon, il faudrait conclure que les refus de répondre n'introduisent aucun biais dans l'échantillon et ne modifient en rien les probabilités de sélection des membres du panel. Il n'y aurait pas lieu de calculer de nouvelles pondérations.

C'est précisément ce qu'il faut vérifier.

Dans un second temps, il convient donc de rechercher des variables, caractéristiques des membres du panel, dont les modalités présentent des taux de réponses très contrastés. Lorsque le contraste entre les taux est significatif, l'échantillon est affecté d'un biais qu'il y a lieu de corriger.

L'identification des variables les plus discriminantes peut être effectuée par différentes procédures (analyse discriminante, "Automatic Interaction Detector", Probit ou toute autre procédure adaptée au fait que la variable dépendante est dichotomique: présence/absence des personnes dans l'échantillon).

Suivant les conseils et l'expérience du P.S.I.D., le panel luxembourgeois a opté pour une procédure qui permet d'éviter d'une part, les contraintes de la linéarité et d'autre part, l'effet des "petits groupes". En évitant d'isoler des trop petits groupes de personnes, cette procédure évite que des taux de réponse extrêmes et propres à des groupes très marginaux influencent le calcul des pondérations.

La procédure adoptée s'apparente à l'A.I.D. mais elle présente deux caractéristiques particulières.

Elle réintroduit à chaque étape de l'analyse TOUTES les variables mises en concurrence. Une variable peut donc intervenir à plusieurs reprises dans le processus d'analyse.

Elle permet d'opérer un choix entre deux variables dont les effets sont approximativement équivalents. L'examen des résultats des discriminations qui suivraient chacune des options possibles fournit des résultats anticipés. Ces résultats peuvent s'avérer très utiles à la décision. La procédure en sens inverse permet de modifier une décision antérieure et d'obtenir une solution plus satisfaisante.

Cette procédure sera illustrée dans chaque chapitre à partir de la deuxième année. Elle peut être décrite de la manière suivante:

* Les membres du panel présents dans l'échantillon reçoivent un code ' 1 '. Les membres du panel absents de l'échantillon reçoivent un code ' 0 ' (les personnes émigrées et les personnes décédées ne sont pas prises en compte).

* Le taux de réponses correspond à la proportion de codes 1 observée dans l'ensemble de l'échantillon ou dans une catégorie déterminée de population.

* L'échantillon est pondéré en fonction de la probabilité de sélection initiale des membres. Cette variable de pondération doit être actualisée chaque année.

  En 1986, année de la seconde vague d'enquête, cette variable correspond exactement à la variable de pondération de l'échantillon initial. En effet, les premiers nouveaux membres du panel apparaissent au cours de cette seconde vague mais ils ne sont pas encore pris en compte dans le calcul des taux de pondération.

  A partir de la troisième vague, la variable de pondération est actualisée chaque année par l'adjonction des poids des nouveaux membres entrés dans l'échantillon au cours de la vague d'enquête précédente.

  Cette première pondération corrige a priori l'inégalité des probabilités de sélection initiales des membres. Dès lors, l'analyse des taux de réponse permet d'identifier *les biais d'échantillonnage liés spécifiquement à la vague d'enquête considérée.*

* Les membres du premier échantillon sont décrits par des caractéristiques qu'ils présentaient en 1985. Les nouveaux membres sont décrits par les caractéristiques qu'ils présentaient au moment ou ils sont entrés dans l'échantillon.

* Ces variables doivent donc être actualisées chaque année par l'adjonction des caractéristiques des nouveaux membres entrés dans l'échantillon au cours de la vague d'enquête précédente.

* Les variables choisies décrivent des caractéristiques du ménage, du chef de ménage et de la personne elle-même.

CARACTERISTIQUES DU MENAGE	DU CHEF DE MENAGE	DE LA PERSONNE
Canton de résidence	Age	Age
Type d'habitat	Sexe	Sexe
Taille du ménage	Etat civil	Etat civil
Nombre d'adultes	Emploi (oui/non)	Formation
Nombre d'enfants		Emploi (O/N)
Nombre d'emplois		Adulte/Enfant
		Lien avec le CM

Toutes ces caractéristiques sont attribuées aux personnes. Les caractéristiques du ménage et du chef de ménage décrivent l'environnement de la personne.

* Ces variables permettent de comparer les taux de réponses de différentes catégories de personnes et d'identifier les catégories de population sous-représentées ou sur-représentées.

Exemple:

Codes	AGE du Chef de ménage	Taux de réponse (86)
1	< 24 ans	.81
2	25 à 34 ans	.84
3	35 à 44 ans	.77
4	45 à 54 ans	.77
5	55 à 64 ans	.73
6	> 64 ans	.69
TOTAL		.77

Dans cet exemple, les personnes appartenant à un ménage dont le chef de ménage est âgé de plus de 64 ans semblent sous-représentées. Leur taux de réponse est particulièrement faible. Leur probabilité de sélection est inférieure à celle de l'échantillon. Leur comportement vis-à-vis de l'enquête a pour effet de sous-représenter dans l'échantillon la catégorie de population dont elles devraient être, théoriquement, les témoins fidèles.

D'autres distortions peuvent être plus centrales. Cette catégorie de personnes peut être absorbée entièrement ou partiellement par d'autres partitions de l'échantillon. Quelle est la partition la plus efficace? Quelle est la partition la plus discriminante?

* Avant de comparer le pouvoir discriminant des variables disponibles, celles-ci doivent être dichotomisées. Cette précaution annule les effets statistiques liés au nombre de modalités.

Les modalités d'une variable peuvent être combinées selon différents regroupements en vue d'obtenir la dichotomie la plus discriminante.

Dans l'exemple, trois variables peuvent être construites: le groupe #6 peut être opposé à l'ensemble des autres modalités, les catégories #5 et #6 peuvent former une modalité et les groupes #1 et #2 peuvent être opposés aux autres.

* Le test du ETA permet de sélectionner la ou les variables dichotomiques qui identifient et opposent les groupes les plus homogènes (voir encadré).

ENCADRE **RAPPEL**

$$E^2_{yx} = 1 - (\text{var. intra de Y} \;/\; \text{var. totale de Y})$$

$$= 1 - \frac{(Sy^2 - Sn_kY^2_k)}{(Sy^2 - NY^2)}$$

où: Sy^2 est simplement la somme générale des carrés des scores (y)

Y est la moyenne générale des y.

$Sn_kY^2_k$ est la somme sur l'ensemble des catégories k, des produits dans chaque catégorie (k) du nombre de cas (n) par le carré de la moyenne de la catégorie (Y^2)

* Généralement, ce test ne suffit pas: plusieurs variables ou plusieurs dichotomies d'une variable restent en concurrence parce que les valeurs de ETA restent très proches.

Exemple: En 1987, la meilleure dichotomie des "cantons de résidence" et la meilleure dichotomie des "âges des chefs de ménage selon leur sexe" présentaient une valeur de ETA égale à .12.

La comparaison des moyennes des groupes (t) permet de choisir la variable qui crée le contraste le plus significatif (P).

Lorsque les valeurs de t restent très proches, le nombre de cas observés dans chaque modalité peut être pris en compte. Le plus souvent, l'équilibre entre les tailles des

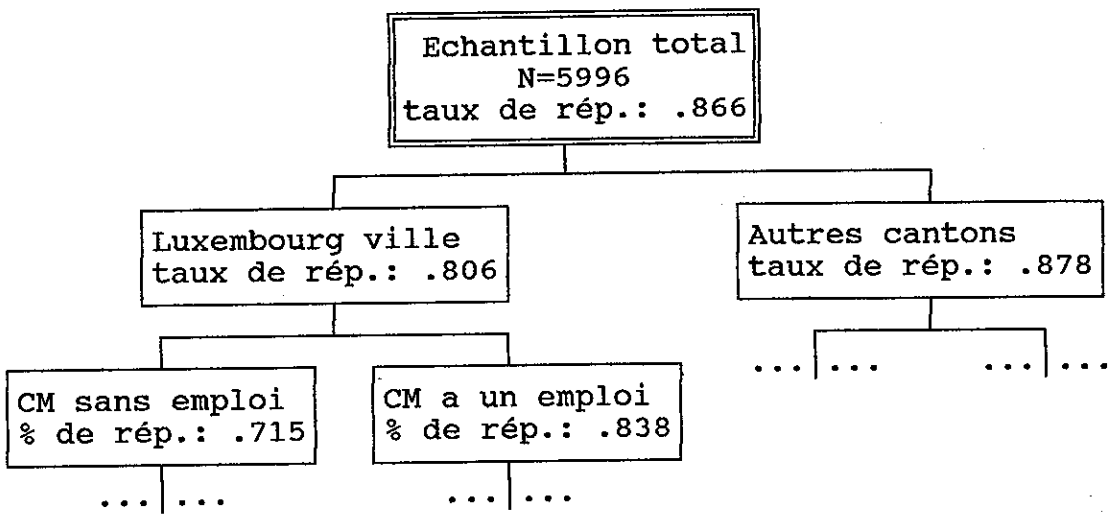
groupes sera privilégié. Dans certains cas, l'isolement d'un groupe particulier peut s'avérer plus satisfaisant pour la suite de l'analyse.

Lorsque des hésitations peuvent encore se justifier, il est utile de pouvoir examiner les conséquences de chaque option sur les résultats ultérieurs. (Voir plus haut, les avantages de la procédure adoptée par le C.E.P.S.).

* La première étape de l'analyse permet donc d'identifier la caractéristique dichotomique qui oppose les deux groupes de personnes dont les taux de réponses sont les plus homogènes (Variance intra-groupe) et les plus contrastés (Comparaison des moyennes). Cette partition de l'échantillon est adoptée à deux conditions: (1) le contraste est significatif sur le plan statistique (le biais est significatif), (2) chaque modalité compte au minimum 200 personnes.

L'analyse est interrompue lorsque l'une de ces deux conditions ne peut être remplie. Il ne faut pas en conclure pour autant que l'échantillon est exempt de tout biais lié aux refus de répondre. Il se peut que la ou les variables pertinentes n'aient pas été prises en compte.

* La première partition scinde l'échantillon en deux catégories.



La seconde partition part de l'une des deux modalités. Elle permet éventuellement de dégager une nouvelle dichotomie significative.

Une variable peut intervenir à plusieurs reprises dans la définition d'une filière.

Une variable peut intervenir dans la définition de plusieurs filières différentes et à des stades différents de la construction de ces filières.

L'analyse est interrompue

* lorsque les variables disponibles ne permettent plus de distinguer des catégories de personnes dont les taux de réponses sont significativement différents,

* ou, lorsque les variables disponibles ne permettent plus de distinguer des groupes d'au moins 200 personnes.

L'ensemble du processus sera présenté chaque année sous la forme d'un schéma arborescent retraçant les divisions successives de l'échantillon.

Au terme de l'analyse, l'échantillon est décomposé en **k** catégories de personnes. Chaque catégorie est définie par l'ensemble des caractéristiques qui ont permis de l'isoler progressivement.

L'ensemble de ces catégories ou modalités forme une variable nominale.

Cette variable portera chaque année un nom composé de deux éléments. Le premier élément est constant: TREE. Le second élément précise l'année d'observation de l'échantillon analysé: TREE87 désigne la variable applicable à l'échantillon observé en 1987.

La variable correspondante de la première année échappe à cette règle parce que la procédure utilisée est fondée sur d'autres principes.

2.2.1.2. *Deuxième étape*

Cette typologie n'est qu'un instrument. Elle permet de classer *tous les membres du panel*, qu'ils soient membres ou non de l'échantillon considéré.

Ces catégories se distinguent par le taux de participation de leurs membres à une vague d'enquête donnée. La population n'est pas représentée de manière constante dans ces différentes zones de l'échantillon et les pondérations initiales ne compensent pas ces déséquilibres.

Les membres du panel présents dans l'échantillon n'ont plus des poids égaux: ils *appartiennent* à des groupes dont les taux de participation varient.

La seconde étape de la démarche consiste donc à attribuer à chaque membre du panel *présent dans l'échantillon* le taux de réponse de la catégorie typologique à laquelle il appartient. Ce taux de réponse définit sa nouvelle probabilité de sélection.

Cette nouvelle variable n'est plus une variable nominale: elle forme une échelle. Elle comporte le même nombre de modalités que la variable typologique. Ces modalités se présentent désormais dans un ordre donné: l'échelle est tendue entre deux valeurs extrêmes qui correspondent, d'une part, au taux de réponse le plus faible et, d'autre part, au taux de réponse le plus élevé. Entre ces deux extrémités, les modalités sont ordonnées en suivant l'ordre croissant de leur valeur.

Cette variable définit donc la probabilité de sélection des membres du panel présents dans un échantillon donné.

2.2.1.3. *Troisième étape*

La troisième étape consiste à calculer le poids de chaque membre.

Lorsque le groupe typologique enregistre un taux de réponse très faible, chaque membre "survivant" doit représenter un plus grand nombre de personnes présentant les mêmes caractéristiques. Lorsque le taux de réponses s'élève, l'importance de cet ajustement se réduit.

D'une manière générale, ce poids est inversement proportionnel au taux de réponses enregistré dans le groupe.

Soit: $W_{i(t)} = 1 / R_{ki(t)}$ (1)

où: $W_{i(t)}$ désigne les poids individuels, dans une vague d'enquête réalisée au moment *t*

$k_{i(t)}$ désigne les groupes typologiques d'appartenance des individus, dans l'échantillon observé au moment *t*

R désigne les taux de réponses observés dans les catégories typologiques de l'échantillon observé au moment *t*.

A partir de la deuxième vague d'observation, ce calcul doit prendre en compte les poids relatifs des membres au moment de leur entrée dans l'étude.

L'équation (1) doit intégrer ce facteur et devient:

$$W_{i(t0+n)} = W_{i(t0)} \times (1 / R_{ki(t0+n)}) \quad (2)$$

où: $W_{i(t0+n)}$ désigne les poids des individus membres du panel, *n* années après leur entrée dans le panel

$W_{i(t0)}$ désigne les poids des individus au moment de leur entrée dans le panel; *t0* correspond à la première année de leur présence dans l'échantillon.

$k_{i(t0+n)}$ désigne les groupes typologiques d'appartenance des individus, dans l'échantillon observé n années après leur entrée dans le panel.

R désigne les taux de réponses observés dans les catégories typologiques de l'échantillon observé au moment *t0+n.*

Exemple:

En 1986: **WGT85** est la variable de **pondération** des membres du panel observés en **1985**

W est la variable qui définit la **probabilité de sélection** des membres du panel, présents dans l'échantillon observé en **1986**

W86 est la variable de **pondération** des membres du panel observés en **1986**

d'ou: W86 = WGT85 x (1 / W)

Cette variable, définissant le poids des membres du panel présents dans l'échantillon, portera chaque année un nom composé de deux éléments: W (weight) suivi de l'année au cours de laquelle l'échantillon a été observé:

W86 en 1986, W87 en 1987, ...

Il est évident que ces variables ne s'appliquent pas aux nouveaux membres du panel, entrés dans l'échantillon au cours de la vague d'enquête considérée. Ils n'entrent pas dans le calcul des taux de réponses (**2.2.1.1.**) et leur probabilité de sélection sera établie selon d'autres règles.

Ces variables ne s'appliquent qu'aux membres du panel ayant participé au moins deux fois à l'étude:

Soit: $i_{(t0+n)}$, $n > 0$

2.2.1.4. *Quatrième étape*

L'utilisation de cette variable de pondération ne se justifie que dans le cadre d'une étude qui ne prendrait pas en compte les nouveaux membres du panel entrés dans l'échantillon
au cours de la vague d'enquête considérée.

Une telle étude permet d'analyser l'effet de la régénération annuelle du panel (par la comparaison entre deux années successives) ou d'analyser l'effet cumulatif de la régénération après plusieurs années.

Toutefois il faut noter que cette variable ne peut pas toujours être utilisée telle quelle (2.2.3.).

Cette variable utilisée à l'état "brut" aura généralement pour effet de modifier la taille de l'échantillon.

Dans un échantillon non-biaisé, tous les membres représentent une unité d'analyse. Toutes les unités d'analyse ont le même poids. Toutes les unités d'analyse ont un poids unitaire et la somme des poids individuels est égale au nombre d'unités observées.

Dans un échantillon biaisé, une catégorie de personnes est sur-représentée. La probabilité de sélection de ses membres est supérieure à la probabilité de sélection des autres personnes. Cette sur-représentation d'une fraction de l'échantillon doit être corrigée. Par exemple, si les membres de cette fraction ont deux fois plus de chance d'être sélectionnés que les autres, ces membres ne doivent représenter qu'une demi-unité d'analyse (l'inverse de leur probabilité de sélection).

Lorsque l'échantillon biaisé est pondéré par ce facteur de correction, toutes les unités d'analyse n'ont plus le même poids. Certaines unités d'analyse ne sont plus prises en compte que pour une fraction d'unité.

Par conséquent, *la somme totale des poids individuels n'est plus égale au nombre d'individus.*

La moyenne des poids individuels est inférieure à l'unité.

En produisant cet effet, la variable de pondération perturbe les propriétés statistiques de l'échantillon. Certains tests paraîtront significatifs, simplement parce que la taille de l'échantillon a été modifiée.

Cet effet peut être annulé en divisant les poids par une constante qui restituera la taille initiale de l'échantillon.

L'équation (2) était:

$$W_{i(t0+n)} = W_{i(t0)} \times (1 / R_{ki(t0+n)})$$

où: $R_{ki(t0+n)}$ désigne les probabilités de sélection des individus membres du panel et présents dans l'échantillon *n* années après leur entrée dans l'étude

$1 / R_{ki(t0+n)}$ désigne l'inverse de ces probabilités

$W_{i(t0)}$ désigne le poids initial des membres du panel

$W_{i(t0+n)}$ désigne le poids des membres du panel présents dans l'échantillon au moment t0+n, compte tenu de leur poids initial.

Les poids ajustés, destinés à sauvegarder la taille réelle de l'échantillon sont:

$$W'_{i(t0+n)} = W_{i(t0+n)} / (SW_{i(t0+n)} / N_{(t0+n)}) \quad (3)$$

$$= W_{i(t0+n)} \times (N_{(t0+n)} / SW_{i(t0+n)})$$

où: $SW_{i(t0+n)}$ désigne la **somme** des poids des membres du panel présents dans l'échantillon au moment *t0+n*

$N_{(t0+n)}$ désigne le nombre de personnes présentes dans l'échantillon indépendamment de toute pondération.

La constante d'ajustement est la valeur moyenne des poids.

Exemple:

En 1987: **N***(t0+n)* = 4583. Soit: l'ensemble des membres du panel observés en 1987, hormis les nouveaux membres entrés dans l'échantillon en 1987.

SW*i(t0+n)*=

4329. Soit: la somme des poids attribués par la variable **W87** aux 4583 membres.

WGT87 est la variable de pondération **W87 ajustée:**

WGT87 = W87 x (4583 / 4329)

Les poids sont ajustés par une constante. Leur somme est égale au nombre de membres présents dans l'échantillon et leur moyenne égale 1.0.

Cette modification se répercute sur les distributions des fréquences ABSOLUES des variables observées.

Elle ne modifie pas les distributions de fréquence RELATIVES: les rapports entre les poids ne sont pas modifiés puisque tous les poids sont multipliés par une constante. Les fréquences relatives restent identiques avant et après l'ajustement de la variable de pondération (**ANNEXE CHAP. 3**: exemple des modifications des distributions de fréquence - avant pondération - après pondération - après ajustement de la variable de pondération).

Cette variable portera chaque année un nom composé de deux éléments: WGT (weight), afin de la distinguer de la variable originale (W), suivi de l'année au cours de laquelle l'échantillon a été observé:

WGT86 en 1986, WGT87 en 1987, ...

Cette variable ne permet pas encore de pondérer TOUT l'échantillon observé en t0+n. L'ensemble des nouveaux membres, appartenant ou non au panel, n'a pas encore été pris en compte.

Les nouveaux membres de l'échantillon qui n'appartiennent pas au panel doivent être pondérés avant les nouveaux membres appartenant au panel, pour des raisons qui seront expliquées plus loin.

2.2.1.5. *Cinquième étape*

Cette étape intermédiaire ne donne pas lieu à la création d'une nouvelle variable.

La procédure d'enquête permet d'observer chaque année de nouvelles personnes: l'observation se déroule au sein des ménages. Les ménages sont des unités très instables: des personnes entrent dans le ménage, quittent le ménage, se séparent pour former de nouveaux ménages, reviennent dans le ménage initial seules, accompagnées d'un conjoint, d'un ami, d'une amie d'un ou plusieurs enfants.

L'observation des personnes qui forment le ménage conduit inévitablement à l'observation de nouvelles personnes chaque année.

Dans la majorité des cas, ces personnes appartenaient à la population avant d'entrer dans l'échantillon. Elles auraient pu entrer dans l'échantillon initial (1985). Elles n'ont pas été sélectionnées au moment du tirage de l'échantillon. Leur probabilité de sélection au moment où elles entrent dans l'échantillon est donc inconnue (à moins que des hypothèses particulières soient émises en dépit du fait que ces hypothèses sont invérifiables).

Ces personnes entrent dans l'échantillon. Elles seront suivies lors des vagues d'enquêtes suivantes, aussi longtemps qu'elles accepteront de répondre. Elles seront prises en compte dans des analyses longitudinales et dans l'analyse des données collectées au niveau des ménages. Mais elles ne seront pas prises en compte dans les analyses synchroniques: elles ne sont pas membres du panel (**2.1.3. Case C**).

Ces personnes reçoivent un poids de '0'. Cet artifice leur confère des propriétés illustrées plus loin (**2.2.1.6.**).

2.2.1.6. *Sixième étape*

Les personnes qui entrent dans l'échantillon et dans la population au cours de la même année sont prises en compte au titre de membres du panel: ce sont pour la plupart des nouveaux-nés.

Ce statut leur est attribué parce qu'ils régénèrent l'échantillon conformément à l'évolution de la population. En outre, leur probabilité de sélection initiale dépend directement de la probabilité de sélection de leurs parents *au moment ou ils naissent* et entrent dans l'échantillon (**2.1.3. Case A**).

Le calcul de leur poids initial suit une règle générale: *leur poids initial est égal à la moyenne des poids de leurs parents*. Leur poids annuel est ensuite calculé selon les règles générales applicables à tous les membres du panel.

Le résultat du calcul de leur poids initial dépend de deux facteurs:

* le statut de leurs parents dans le panel,

* le statut de leurs parents dans l'échantillon.

Il faut donc envisager quatre cas de figure.

1. Le père et la mère sont membres du panel et membres de l'échantillon au moment ou l'enfant entre dans le panel.

   Le poids de l'enfant est égal à la *moyenne des poids* de ses parents.

Il est égal au poids de chaque ascendant lorsque ceux-ci ont des poids identiques.

2. L'un des deux parents n'est pas membre du panel mais les deux parents sont présents dans l'échantillon.

   Le poids de l'enfant est égal à la moyenne des poids des parents: l'un des deux parents ayant reçu un poids de 0, le poids de l'enfant est égal à la *moitié du poids* de l'ascendant membre du panel.

3. L'un des deux parents n'est pas membre de l'échantillon, pour une raison quelconque, provisoire ou définitive.

   Le poids de l'enfant est *égal au poids* de l'unique ascendant direct présent dans l'étude. Sa probabilité d'entrer dans l'échantillon dépend uniquement de la présence de l'ascendant dont il prend le poids.

4. Les deux parents sont non-membres du panel ou absents de l'échantillon.

   Les deux parents étant non-membres du panel ont un poids de 0: l'enfant reçoit la moyenne de leur poids soit 0. Ce nouveau membre de l'échantillon n'est pas membre du panel. Cet enfant aura été préalablement classé dans le groupe des nouvelles personnes *non-membres* du panel (**2.1.3. Case C**). Son poids lui sera donc attribué dans le cadre de la cinquième étape (2.2.1.5.).

   L'absence des parents donne lieu aux mêmes dispositions: le nouveau membre est classé a priori dans la catégorie des non-membres du panel. Son entrée dans l'échantillon n'est pas liée à la présence de ses ascendants directs.

Le statut des parents détermine le poids des nouveaux membres du panel. Les nouvelles personnes non-membres du panel doivent donc recevoir leur poids avant les nouveaux membres du panel (2.2.1.5.).

La sixième étape enrichit la variable de pondération annuelle.

Aux poids des membres du panel viennent s'ajouter:

* les poids des nouveaux membres du panel entrés dans l'échantillon au cours de la vague d'enquête de l'année $t0+n$,
* les poids des nouveaux membres de l'échantillon nonmembres du panel

Cette extension de la variable $W_{i(t0+n)}$ (équation (2)) (2.2.1.3.) permet de pondérer l'ensemble de l'échantillon observé au cours de cette vague d'enquête en prenant en compte les individus tels que:

$$i_{(t0+n)} = i_{(t0)}$$

Cette nouvelle variable sera désignée chaque année par le nom de la variable originale (W86, W87, ...) spécifiée par la lettre B afin de rappeler qu'il s'agit simplement d'une extension de la variable de départ. Soit:

W86B en 1986, W87B en 1987, ...

Cette variable présente les mêmes propriétés et les mêmes limites que la variable $W_{i(t0+n)}$: lorsqu'elle est utilisée pour pondérer l'échantillon, elle modifie la taille de l'échantillon.
La somme des poids n'est pas égale au nombre de personnes présentes dans l'échantillon.

En outre, elle présente une particularité: les poids '0'.

La dernière étape de la procédure consiste à ajuster cette variable de telle manière que les propriétés statistiques de l'échantillon soient sauvegardées. Le rôle des poids nuls sera examiné dans le cadre de cette dernière étape.

2.2.1.7. *Septième étape*

La procédure d'ajustement de cette variable et les raisons de cet ajustement ont été commentées plus haut (équation (3)) (**2.2.1.4.**).

L'introduction de poids nuls exige quelques commentaires parce qu'elle suscite une question: à quel titre les non-membres du panel (les poids nuls) doivent-ils être pris en compte au moment de calculer la constante d'ajustement?

Le calcul de la constante d'ajustement ne prend pas en compte les non-membres du panel: l'attribution de poids nuls vise précisément à exclure ces membres de l'analyse au moment où l'échantillon doit être pondéré. (Les raisons de leur exclusion ont déjà été développées (**2.2.1.5.**)).

L'usage de la variable de pondération aura donc un triple effet:

* il corrige l'inégalité des probabilités de sélection des unités d'analyse,

* il sauvegarde les propriétés statistiques de l'échantillon,

* il sélectionne automatiquement les unités statistiques pertinentes.

Exemple:

En 1987: **W87B** est la variable de pondération (non ajustée) applicable à l'ensemble de l'échantillon observé.

Cette variable s'applique à 4914 personnes:

1. membres du panel en 85 et 87		4583
2. nouveaux membres du panel en 87		65
3. nouveaux non-membres en 1987		266
	N =	4914

Les 266 nouveaux non-membres ont reçu un poids égal à 0.

La distribution de W87B peut être présentée de deux manières différentes: soit, les 266 unités ayant reçu un poids nul sont prises en compte, soit ces unités sont tenues à l'écart.

Dans le premier cas, N est égal à 4914. Dans le second cas N est égal à 4648:

	Avec poids 0	Sans poids 0
N =	4914	4648
moyenne =	.89	.94
sigma =	.459	.417
somme =	4385.32	4385.32

La somme des poids est identique dans les deux cas, puisque les 266 personnes mises à l'écart avaient un poids nul.

Dans les deux cas, la somme des poids est inférieure à N.
La procédure d'ajustement doit être appliquée.

Première hypothèse

Les poids sont ajustés sur une base de 4914 personnes. La somme des poids est égale à 4914 unités. Elle reste identique avant et après la mise à l'écart des personnes ayant un poids nul. L'échantillon des membres du panel est donc porté artificiellement à 4914 unités statistiques au moment où l'échantillon est pondéré.

Seconde hypothèse

Les poids sont ajustés sur la base de 4648 unités. La somme des poids est égale à 4648 unités avant et après la mise à l'écart des unités à poids nul.

	Avec poids 0	Sans poids 0
N =	4914	4648
moyenne =	.94	1.0
sigma =	.486	.442
somme =	4648	4648

Utiliser la variable de pondération ajustée sur 4648 unités, sans écarter les personnes ayant reçu un poids de 0, signifie que les analyses saisissent 4914 personnes mais ne prennent en compte que 4648 unités d'analyse. Ceci est logique puisque 266 personnes ont un poids nul.

La mise à l'écart des personnes ayant reçu un poids nul conduit exactement aux mêmes résultats.

Avant ou après la mise à l'écart des poids nuls:

* la taille de l'échantillon est identique puisqu'elle est définie par la somme des poids,

* la taille de l'échantillon correspond au nombre de membres du panel,

* les distributions de fréquence des variables observées dans l'échantillon sont identiques.

Application: Distribution des fréquences des catégories de membres de l'échantillon, **après pondération ajustée**

	avec les poids 0		sans les poids 0	
	n=	%	n=	%
1. memb.85/87	4589	99.0	4589	99.0
2. nouv.memb. panel 87	59	1.0	59	1.0
3. nouv.memb. non-pan. 87	266	0.0	-	-
TOTAL	4648	100.0	4648	100.0

Cette variable de pondération ajustée sera désignée chaque année par un nom composé de trois termes: WGT (par référence à la première variable ajustée WGT86, WGT87) suivi de l'année (WGT86, WGT87, ...) et spécifié par le suffixe TOT (signifiant que cette variable s'applique à l'ensemble de l'échantillon observé).

Soit: WGT86TOT, WGT87TOT, ...

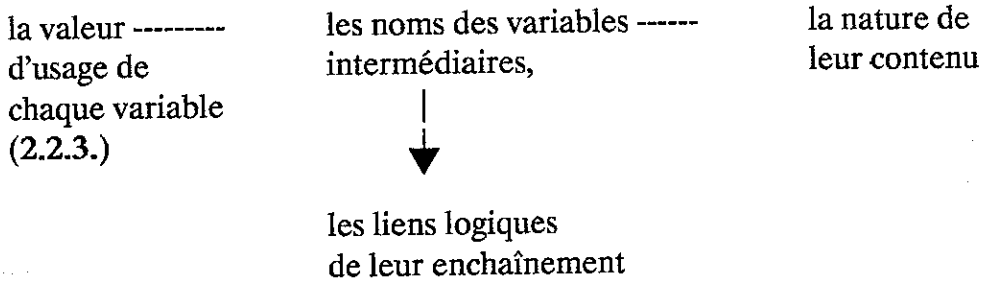
Cette variable n'a pas d'équivalent en 1985: tous les membres de l'échantillon sont membres du panel.

2.2.2. Schéma récapitulatif

Chaque chapitre décrit les variables de pondération calculées pour une année.

Un schéma récapitule au début de chaque chapitre les variables calculées au cours de la procédure qui a été décrite.

Ce schéma reprend:



Un utilisateur habitué aux techniques de pondération pourra se limiter à la consultation de ces schémas pour trouver rapidement la variable dont il souhaite faire usage.

2.2.3. La valeur d'usage des variables

Chaque variable est désignée par une 'valeur d'usage'.

Certaines variables sont dites "opérationnelles" et d'autres sont dites "instrumentales" en fonction de leur rôle dans le système de pondération.

Les variables dites "opérationnelles" peuvent être utilisées sans transformation préalable. Elles permettent de limiter les effets des biais qui affectent les échantillons.

Leur usage présuppose que le logiciel d'analyse utilisé n'effectue pas automatiquement l'ajustement des poids à la taille de l'échantillon. (Les utilisateurs du SPSSX prendront garde au fait que l'ajustement des variables de pondération n'est pas automatique. La somme des poids n'est pas égale au nombre de personnes appartenant à l'échantillon et la moyenne de ces poids n'est pas égale à 1.0).

Les variables dites "instrumentales" ont permis de calculer les variables opérationnelles. Elles ponctuent les différentes étapes de la procédure de calcul.

Lorsqu'un logiciel d'analyse des données effectue automatiquement l'ajustement des poids à la taille de l'échantillon, la variable de pondération est la variable instrumentale qui précède directement la variable ajustée. Le schéma présenté chaque année, en tête de chapitre, permet de repérer rapidement cette variable.

2.2.4. Les poids des ménages

2.2.4.1. *Les ménages dans un fichier individuel*

Les variables calculées jusqu'ici sont destinées à corriger les échantillons dont l'unité d'analyse est l'individu membre d'un ménage.

L'ensemble du fichier est fondé sur cette unité d'information et d'analyse. Toutes les variables répertoriées dans le fichier attribuent des valeurs aux *membres* des ménages.

Chaque personne est désignée par un numéro. Ce numéro est strictement individuel et reste identique d'année en année.

Chaque personne est également désignée par le numéro du ménage auquel elle appartient. Ce numéro est identique pour toutes les personnes appartenant au même ménage. Il n'est pas nécessairement identique chaque année.

Le numéro du ménage permet donc de procéder à l'agrégation de valeurs individuelles au niveau du ménage ou inversement, de ventiler des valeurs observées ou calculées au niveau du ménage sur les membres de ce ménage.

2.2.4.2. *La pondération des ménages*

Tous les ménages n'ont pas eu la même probabilité de sélection au moment du tirage de l'échantillon initial. Le ménage n'était pas l'unité de tirage.

A priori, plusieurs unités de tirage pouvaient conduire au même ménage. Par conséquent, l'échantillon des ménages est affecté d'un biais. La probabilité de sélection du ménage est proportionnelle au nombre d'unités de tirage qui peuvent conduire à identifier ce ménage comme membre de l'échantillon. Plus le nombre d'unités de tirage conduisant au même ménage s'élève, plus le coefficient de pondération de ce ménage doit être faible.

Par contre, toutes les personnes membres du ménage participaient à l'observation. Les membres du ménage ont donc, tous, la même probabilité de sélection. Cette probabilité est égale à la probabilité de sélection du ménage. Chaque membre du ménage doit donc être affecté du même poids que le ménage.

A l'issue de la *première vague d'enquête*, le poids des individus est donc dérivé du poids du ménage.

Mais la structure interne des ménages peut se modifier d'une année à l'autre. Chaque personne peut quitter le ménage, fonder un nouveau ménage, ou rejoindre un autre ménage. Au sein d'un ménage certains membres peuvent refuser de répondre, être momentanément ou définitivement absents. Chaque personne peut suivre un itinéraire personnel, indépendant du ménage d'origine.

A partir de la seconde vague d'enquête, le calcul des taux de réponses prend en compte cette mobilité individuelle: il s'effectue au niveau des individus parce que leur carrière au sein du panel n'est pas nécessairement liée à celle de leur ménage d'origine.

Le poids des individus est donc calculé avant le poids des ménages. Le poids du ménage est désormais calculé en fonction du poids des individus qui le composent.

Le poids du ménage est égal à la valeur moyenne des poids des membres qui le composent (Cf. P.S.I.D., Ann Arbor, Michigan).
(Rem. Cette propriété du poids du ménage est observée dès la première année, puisque les poids des membres du ménage sont à la fois égaux entre eux et égaux au poids du ménage).

A partir de 1987, certains ménages reçoivent un poids de 0 parce qu'ils sont composés exclusivement de personnes non-membres du panel (2.2.1.7.).

2.2.4.3. *Les poids des ménages dans le fichier individuel*

Le poids des ménages peut être calculé en agrégeant les valeurs des membres au niveau du ménage.

Cette nouvelle unité d'analyse n'est plus compatible avec le dispositif du fichier individuel.

Les valeurs du ménage sont donc reventilées sur chaque membre du ménage. Elles ne sont pas utilisables sous cette présentation mais il est aisé de recomposer les unités en agrégeant au niveau du ménage les valeurs ventilées sur ses membres. (Au moment de sauver le fichier, l'utilisateur n'omettra pas d'emporter les numéros d'identification des ménages, à toutes fins utiles ...).

Chaque année, le poids des ménages sera désigné par un nom composé de quatre éléments: M (ménages), WGT (weight), 86 (année), A (ajustée à la taille de l'échantillon).

Soit: MWGT85A MWGT86A ...

La variable correspondante, avant ajustement, est désignée par un suffixe différent.

Soit: MWGT86T ...

Ces variables sont reprises dans le schéma récapitulatif placé en tête de chapitre.

2.3. ECHANTILLONS ET VARIABLES DE PONDERATION

Les variables ont été définies essentiellement par le rôle qu'elles sont appelées à jouer dans le système de pondération.

Elles doivent encore être définies par leur rôle dans l'analyse des données.

Le fichier est évolutif et cumulatif. Par conséquent, les différentes variables de pondération ne s'appliquent pas systématiquement à toutes les personnes répertoriées dans le fichier.

Chaque année, la portée de chaque variable sera précisée: un tableau indiquera la nature et la taille de l'échantillon à partir duquel chaque variable a été calculée. Cet échantillon sera ventilé selon les différentes catégories de la variable DIA(gnostic) de l'année.

Les variables de pondération sélectionnent automatiquement l'ensemble des personnes auquel elles s'appliquent.

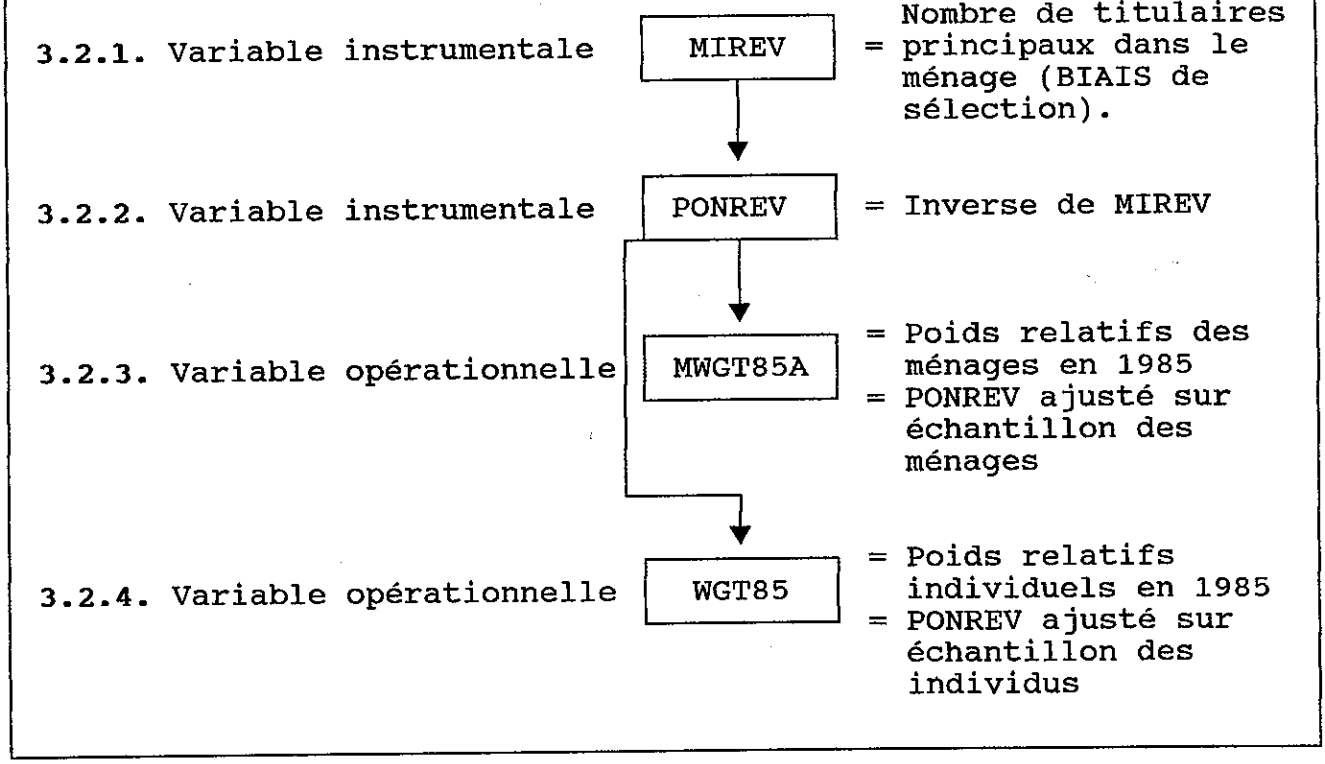
Les codes de la variable DIA(--) permettent d'effectuer des sélections de sous-échantillons.

Un tableau général, placé à la fin de ce document, récapitule les variables principales et les sous-échantillons auxquels elles s'appliquent.

ÉCHANTILLON 1985

ET VARIABLES DE PONDÉRATION

Encadré 3.1. **VARIABLES DE PONDERATION 1985**



Chapitre 3

ÉCHANTILLON 1985

ET VARIABLES DE PONDÉRATION

3.1. L'ECHANTILLON INITIAL

3.1.1. Unité de sélection et population de référence

La population de référence est définie par l'ensemble des titulaires principaux au titre de la Sécurité Sociale et de la Protection Sociale qui résident au Luxembourg.

* Un **titulaire principal** est la personne qui, dans un ménage, est un assuré social ou un bénéficiaire de la Protection sociale, à titre principal.

* Il doit résider au pays.

* Il n'est pas attaché à un système de sécurité sociale ou de protection sociale tel que celui de la Communauté Européenne.

* Il doit vivre dans un ménage "privé" et non dans un ménage "collectif".

* Il peut avoir une ou plusieurs personnes à charge (époux, enfant, autres).

* Plusieurs titulaires principaux peuvent appartenir à un même ménage. Ils correspondent à autant de "groupes de revenus" selon une définition propre à la Sécurité Sociale.

* Tous les titulaires principaux sont pris en compte, à une exception près: lorsque deux époux sont titulaires principaux à un titre quelconque, ils sont regroupés et un seul titulaire est pris en compte (par convention, il s'agit le plus souvent du mari).

3.1.2. Constitution du fichier de référence

Le schéma **3.A.** présente de manière détaillée les différentes étapes de la constitution du fichier de référence qui permettra de sélectionner l'échantillon.

Les opérations peuvent être regroupées en trois grandes étapes.

3.1.2.1. *Première étape*

* Regroupement des listes de personnes affiliées aux caisses de la Sécurité Sociale du secteur privé, à la caisse des fonctionnaires et au Fonds National de Solidarité.

* Elimination des "non-résidants" et des personnes affiliées à un système de Sécurité Sociale étranger ou international.

* Toutes les personnes retenues sont liées à la Sécurité Sociale luxembourgeoise en raison de leur emploi, de droits acquis par leur travail passé ou celui de leur conjoint.

3.1.2.2. *Deuxième étape*

* Création d'un fichier unique et complet de la population résidante "bénéficiaire" de la Sécurité Sociale Luxembourgeoise.

* Ces bénéficiaires sont les titulaires et les personnes qui sont à leur charge.

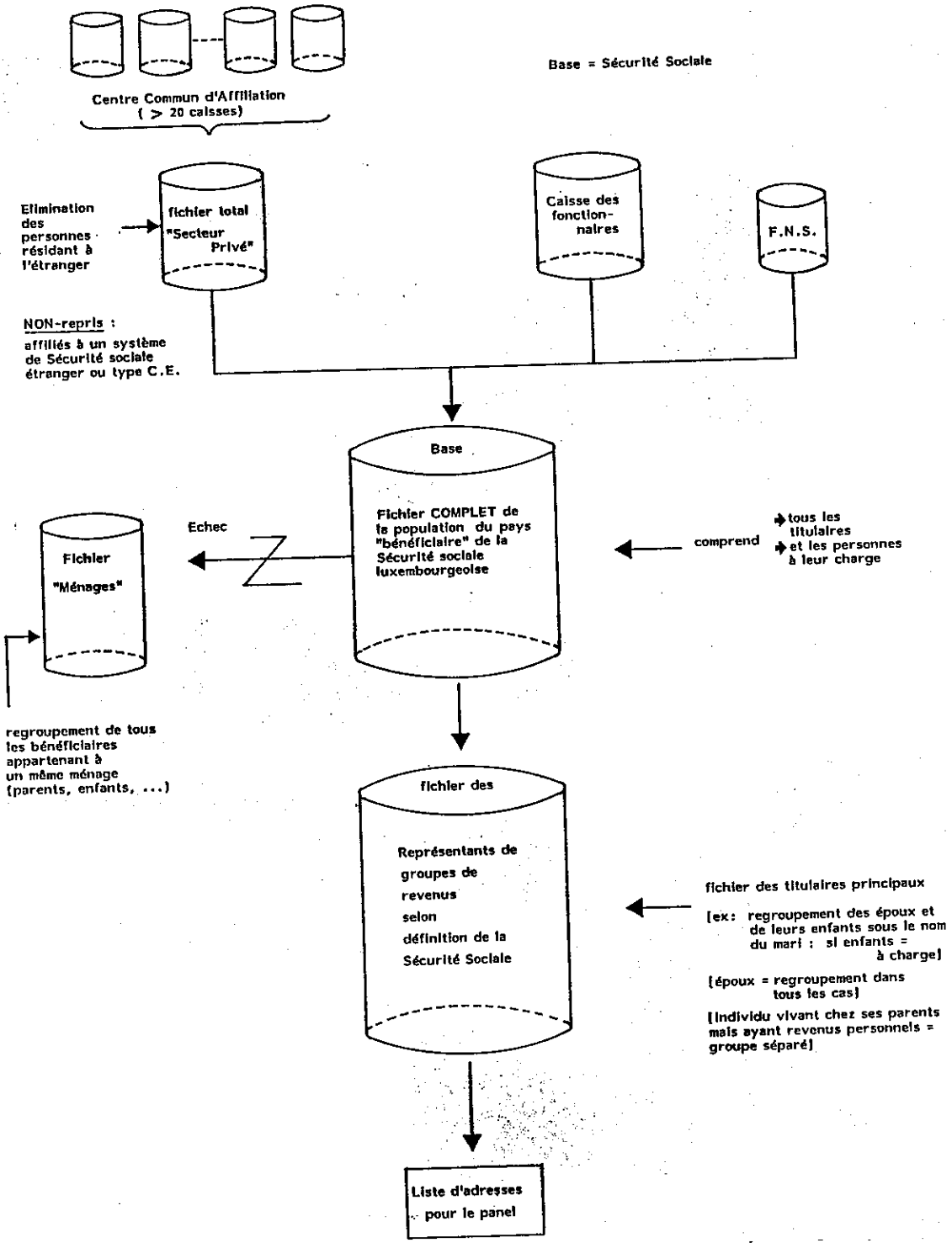
3.1.2.3. *Troisième étape*

* Tentative de regroupement de tous les bénéficiaires appartenant à un même ménage: cette opération est impossible à réaliser. Les ménages devront être identifiés par une autre procédure. Ils seront reconstitués au moment de l'enquête sur le terrain.

* Regroupement des bénéficiaires autour des titulaires principaux. Cette opération consiste à reconstituer des groupes de revenus, conformément à la définition de la Sécurité Sociale.

SCHEMA 3.A.

Historique du fichier de référence



Exemples: - regroupement des époux

- regroupement des époux et des enfants à charge

- un enfant domicilié chez ses parents mais disposant de revenus personnels forme un groupe de revenus distinct.

* A chaque titulaire principal et à l'ensemble des personnes qui lui sont attachées correspond un et un seul enregistrement, une et une seule adresse.

* Plusieurs titulaires principaux peuvent conduire à la même adresse. Cette adresse permet d'identifier un ménage. Le ménage peut correspondre à une ou plusieurs personnes, à un ou plusieurs titulaires principaux, à un ou plusieurs groupes de revenus.

Exemples: **1 ménage - 1 titulaire - 1 groupe - 4 personnes:**

Un ménage composé des 2 parents et de 2 enfants à charge.

1 ménage - 3 titulaires - 3 groupes - 5 personnes:

Un ménage composé des 2 parents, d'un enfant à charge, d'un enfant au travail, d'une belle-mère pensionnée.

* Ce fichier des adresses des titulaires principaux constitue la base de tirage de l'échantillon du panel.

3.1.3. Sélection de l'échantillon

Le fichier de référence contient les 168234 adresses des titulaires principaux.

L'échantillon est sélectionné par un tirage aléatoire simple.

La taille minimale de l'échantillon est fixée à 2000 unités.

Les aléas de l'enquête imposent le tirage de quatre listes d'adresses.

La première liste fournit 2018 adresses.
La deuxième liste fournit 434 adresses de remplacement.
La troisième liste fournit 416 adresses.
La quatrième liste fournit les 422 adresses qui permettront d'atteindre le quota de réponses voulu, soit 2012 adresses.

Au total, 3290 adresses ont été tirées aléatoirement mais elles n'ont pas nécessairement donné lieu à une enquête. Le schéma **3.B.** retraçe brièvement les histoires imbriquées de l'échantillon et de l'enquête.

Il est utile de s'arrêter un moment sur ces quelques chiffres.

Toutes les adresses disponibles n'ont pas été utilisées.

* Certains titulaires principaux n'entraient pas dans le cadre de la population cible, en dépit des précautions qui ont été prises au moment où le fichier de référence a été constitué.

Un fichier peut difficilement enregistrer les mouvements quotidiens de la population. Certains titulaires ont été écartés de l'enquête parce qu'ils n'auraient pas du se trouver dans le fichier de référence: des personnes décédées, des personnes qui résidaient dans des ménages collectifs.

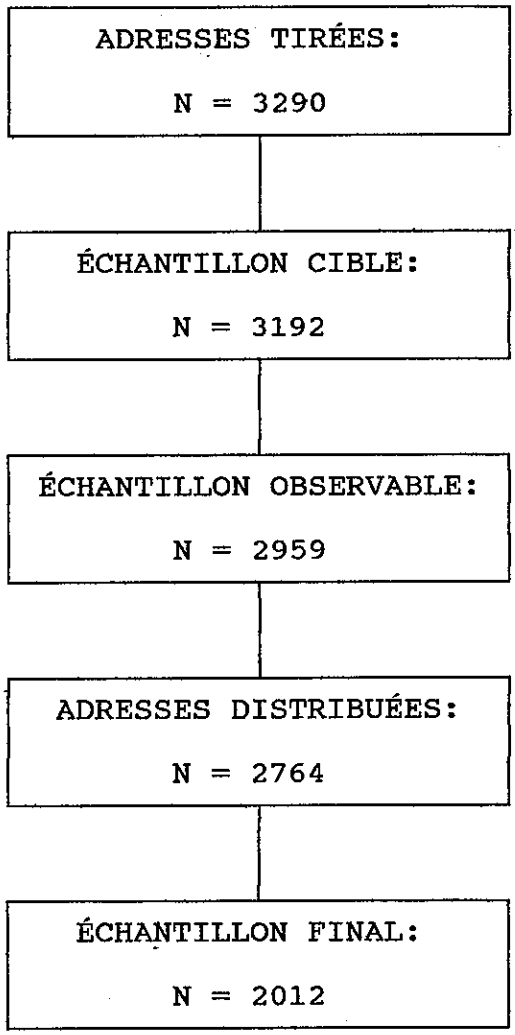
D'autre part, le tirage des adresses est un tirage dit "avec remise"; 16 personnes ont été sélectionnées à deux reprises. Ce phénomène peut être considéré comme marginal sur le plan statistique.

Suite à l'élimination de ces adresses, l'échantillon cible ne contient plus que 3192 adresses.

* Certaines personnes appartenant à la population cible n'étaient pas nécessairement "observables". Elles ne pouvaient participer à l'enquête pour des raisons très concrètes:

SCHEMA 3.B.

Historique de l'échantillon



- les personnes qui ont quitté le pays pour une durée indéterminée et qui se trouvaient toujours recensées dans le fichier de référence;

- les personnes "introuvables" soit parce qu'elles ont changé d'adresse, soit parce qu'elles ne sont jamais présentes à leur domicile;

- certaines personnes handicapées inaptes à répondre au questionnaire.

Au total, 10 % des adresses sont inutilisables.

La mise à l'écart de ces adresses, au cours de l'enquête, réduit l'échantillon observable à 2959 titulaires principaux.

* Parmi ces 2959 adresses utilisables, 2764 ont été réellement utilisées.

Chaque enquêteur dispose d'une réserve d'adresses destinée à remplacer les ménages qui refusent de répondre. Cette réserve est constituée essentiellement au moment du tirage de la quatrième liste de remplacement et dès que le seuil des 2000 ménages est atteint, l'enquête est interrompue.

Près de 6 % des adresses dites "observables" n'ont donc pas donné lieu à un contact entre l'enquêteur et le ménage. Il n'est pas possible d'estimer la proportion de ces adresses qui était réellement utilisable.

Les enquêteurs entrent en contact avec 2764 ménages correspondant à la définition de l'échantillon-cible et aptes à répondre.

A ce stade, 16 % des adresses disponibles ont été écartées.

* Ces 2764 contacts récoltent 2012 réponses, soit un taux de réponses de 72.7 %.

Cette dernière sélection porte à 38.9 % la proportion des adresses disponibles qui ont été écartées: telle est la mesure de la distance qui sépare un fichier dit "exhaustif" d'un échantillon dit "représentatif".

La plus grande partie de cette déperdition provient des refus de répondre (23 des 39 %). Ce facteur fera l'objet d'un examen particulier (3.1.5.).

3.1.4. Les unités d'analyse

La conception de l'échantillon offre la possibilité de varier les niveaux d'analyse parce que les informations ont été saisies à des niveaux différents ou parce que ces informations peuvent être agrégées (ou désagrégées) à des niveaux d'analyse différents.

Le schéma 3.C. présente ces différents niveaux d'analyse.

3.1.4.1. *Première unité d'analyse*

Les *2012 titulaires principaux* ont permis d'identifier 2012 adresses. Chaque titulaire principal est un "individu repère": il correspond à un "groupe de revenu" conformément à la définition adoptée par les services de la Sécurité Sociale. Ce groupe est composé du titulaire et de l'ensemble des personnes qui sont à sa charge.

Ces titulaires principaux forment le seul échantillon réellement probabiliste.

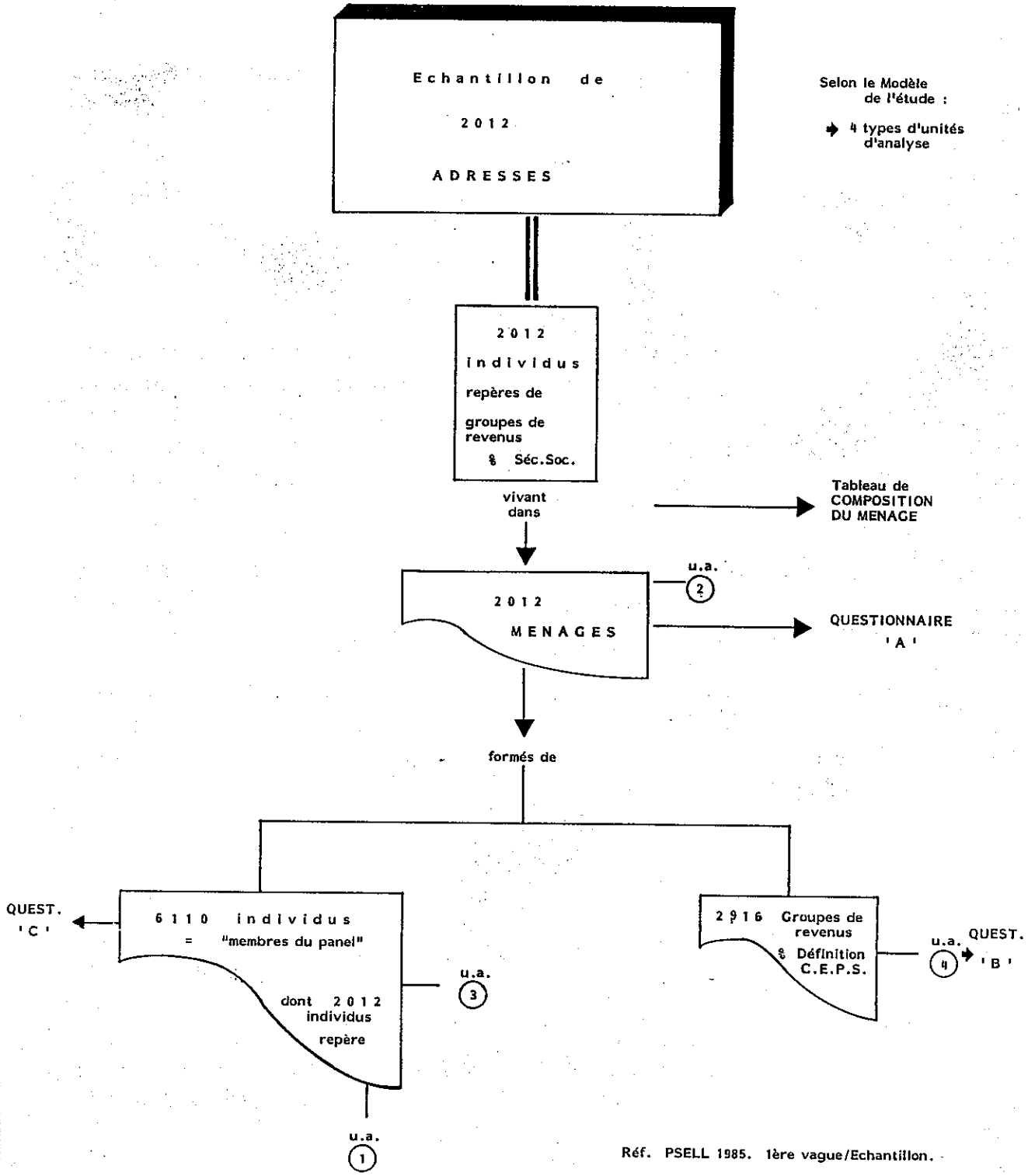
Il est utile d'anticiper quelque peu sur la suite de la présentation: un ménage peut être composé de plusieurs titulaires. Seul le titulaire principal sélectionné dans le cadre de l'échantillon peut être pris en compte au titre d'unité d'analyse appartenant à l'échantillon probabiliste. Les autres titulaires présents dans le ménage auraient pu être sélectionnés au moment du tirage, mais ils ne l'ont pas été. Ils ne peuvent pas être pris en compte au titre d'unités d'analyse appartenant à l'échantillon probabiliste.

En principe, cet échantillon est directement représentatif de l'ensemble des titulaires appartenant à la population de référence.

Les titulaires principaux peuvent être décrits par leurs caractéristiques individuelles mais aussi par d'autres variables élaborées à d'autres niveaux d'analyse. Par exemple, les variables décrivant le ménage peuvent être reventilées au niveau des titulaires principaux.

SCHEMA 3.C.

Modèle de l'étude - Unité d'analyse



Réf. PSELL 1985. 1ère vague/Echantillon.

3.1.4.2. *Deuxième unité d'analyse*

Chaque titulaire principal conduit au domicile d'un *ménage*. L'échantillon est donc formé par 2012 ménages.

Le questionnaire d'observation débute par un tableau décrivant la composition de ce ménage. Cette procédure permet d'établir les liens existant entre les membres du ménage.

La notion de ménage dépasse la notion de famille. Dans la plupart des cas, les personnes présentes dans le ménage sont liées par des liens familiaux plus ou moins étroits mais il arrive que certaines personnes dépourvues de tout lien familial avec les autres membres partagent le même toit: amis, amies, relations.

Il est possible d'identifier parmi les membres du ménage plusieurs titulaires au titre de la Sécurité sociale. Chaque titulaire aurait pu être sélectionné au moment du tirage de l'échantillon. Chacun aurait pu conduire à la même adresse. Par conséquent, tous les ménages n'avaient pas à priori la même probabilité d'être sélectionnés. La probabilité de sélection du ménage était directement proportionnelle au nombre de titulaires susceptibles de conduire à la même adresse.

L'échantillon des ménages est donc biaisé et ce biais devra être corrigé. Les informations relatives aux ménages devront être analysées sous le contrôle d'un facteur de pondération des unités d'analyse.

Les ménages sont décrits par des caractéristiques qui sont saisies directement à leur niveau: par exemple, le canton de résidence, les conditions de logement ou les équipement ménagers. Mais d'autres variables observées à d'autres niveaux d'analyse (par exemple, les membres du ménage) peuvent être agrégées au niveau des ménages.

3.1.4.3. *Troisième unité d'analyse*

Dans chaque ménage, tous les individus sont pris en compte et l'observation porte sur chacun d'eux: *6110 personnes* forment l'échantillon des membres du panel.

Tous les membres du ménage sont observés. Par conséquent, ils ont tous la même probabilité de sélection au sein du ménage. Mieux: ils ont tous la même probabilité de sélection que le ménage et le biais qui affecte l'échantillon des ménages rejaillit directement sur l'échantillon des individus.

Les données relatives aux personnes devront donc être analysées sous le contrôle d'un facteur de correction.

Le questionnaire comporte un volet de questions destinées à décrire les caractéristiques de chaque membre (dont les 2012 titulaires principaux). Chaque membre peut également être décrit par les caractéristiques du ménage auquel il appartient.

L'un des membres du ménage est désigné "chef du ménage" en fonction de certains critères. Chaque membre peut donc être décrit par les caractéristiques individuelles du chef du ménage auquel il appartient.

3.1.4.4. *Quatrième unité d'analyse*

Le dispositif de l'étude offre encore une quatrième possibilité. Il s'agit d'une construction interne au panel. Elle est citée, ici, à titre d'information parce qu'elle résulte d'un travail préalable sur les informations collectées au sein du ménage.

L'analyse des transferts monétaires entre les membres d'un ménage permet d'isoler *les groupes de revenus*. Ils ne se confondent pas nécessairement avec les "groupes de revenus" définis par les services de la Sécurité Sociale. Ils sont construits à partir de l'observation des pratiques de mise en commun des revenus entre les membres du ménage.

Parmi les 6110 personnes observées au cours de cette première vague d'enquête, il est possible d'identifier environ 2915 groupes de revenus. La probabilité de sélection de ces groupes peut être dérivée directement de la probabilité de sélection des ménages: au sein de chaque ménage les groupes de revenus sont mutuellement exclusifs et collectivement exhaustifs.

Les caractéristiques de ces groupes ne sont jamais saisies directement. Elles résultent toujours d'une reventilation des informations collectées à d'autres niveaux.

3.1.5. Des biais d'échantillonnage ?

Avant de calculer les valeurs des facteurs de pondération qui devront contrôler les analyses des données aux différents niveaux, il est important de vérifier la qualité de l'échantillonnage des titulaires principaux.

Les moyens disponibles sont très limités mais ils permettent de penser qu'il n'y a pas lieu d'introduire d'emblée des facteurs de correction au niveau de l'échantillon des titulaires principaux.

D'autre part, un biais s'est introduit dans le tirage des adresses au cours de la dernière période de l'enquête. Ce biais ne peut être corrigé mais il n'affecte qu'une très petite partie de l'échantillon.

3.1.5.1. *Les titulaires principaux*

Le procédé le plus couramment utilisé pour tester la présence de biais dans un échantillon consiste à confronter la distribution des fréquences d'une ou de plusieurs variables telles qu'elles se présentent dans l'échantillon et dans la population.

Le fichier complet des titulaires permet d'obtenir la distribution des fréquences de la population selon le sexe des personnes. La taille de ce fichier ne permet pas de reconstituer les cantons de résidence de tous les membres.

Un échantillon aléatoire de 342 adresses, sélectionné à l'aide d'une table de nombres aléatoires, permet de reconstituer les cantons de résidence des titulaires. Il permet également d'obtenir la distribution des fréquences de ces individus selon le sexe.

Ces données de référence permettent de procéder à deux types de comparaisons.

Une première comparaison permet d'estimer l'ampleur d'un biais qui serait apparu entre:

- le fichier de référence (population totale des 168234 titulaires (ou l'échantillon aléatoire de 342 adresses)),
- le tirage de l'échantillon initial des 3290 adresses,
- et la sélection des adresses réellement utilisées (2764).

Cette première comparaison ne prend pas en compte les biais liés aux refus de répondre. Elle permet essentiellement d'observer les effets de la sélection progressive des adresses pertinentes dans le cadre de l'étude.

Une seconde comparaison permet d'observer l'effet spécifique des refus de répondre, indépendamment des distortions provoquées par le nettoyage du fichier de référence. Elle confronte:

- l'échantillon des adresses qui ont réellement donné lieu à un contact entre les enquêteurs et les titulaires (2764)

- et les 2012 titulaires principaux qui ont accepté de répondre et qui forment l'échantillon final.

Ces comparaisons s'appliquent respectivement à la distribution des individus selon leur sexe et à leur répartition dans les cantons de résidence.

TABLEAU 3.1.
Répartition des titulaires selon le sexe (en %)

SEXE	Fichier complet	Echant. aléatoire	Echant. initial	Adresses utilisées	Echant. final
hommes	71.5	71.6	71.7	72.8	74.2
femmes	28.5	28.4	28.3	27.2	25.8
N =	168234	342	3290	2764	2012

Le tableau **3.1.** présente la répartition des titulaires selon le sexe. Il permet de tirer quelques conclusions.

* L'échantillon aléatoire de 342 personnes est très fidèle à la population de référence.

* L'échantillon initial composé de 3290 adresses est également très fidèle à la population de référence.

* Le nettoyage de l'échantillon initial induit une très légère sur-représentation des titulaires masculins.

* Les refus de répondre n'ont pas modifié la distribution des fréquences de manière significative, bien que les hommes soient très légèrement sur-

représentés par rapport à l'ensemble des personnes qui étaient à même de participer à l'étude.

TABLEAU 3.2.
Répartition des titulaires selon le canton de résidence (en %)

CANTON	Fichier complet	Echant. aléatoire	Echant. initial	Adresses utilisées	Echant. final
Luxembrg. ville	-	20.2	21.1	20.0	18.5
Luxembrg. campagne	-	9.1	8.8	8.9	8.0
Esch/Alz.	-	28.7	32.4	32.5	34.8
autres	-	42.0	37.7	38.6	38.7
N =	168234	342	3290	2764	2012

Le tableau 3.2. présente la répartition des titulaires selon les cantons de résidence. Les trois cantons les plus représentés sont opposés globalement à l'ensemble des autres.

Ce tableau appelle quelques remarques.

* La représentation du canton de Esch sur Alzette est plus importante dans le fichier initial que dans le fichier aléatoire composé de 342 adresses (32.4 % contre 28.7 %). La différence entre les tailles des deux échantillons peut expliquer cet écart. Ces deux échantillons ne sont que deux des innombrables échantillons qu'il est possible de tirer à partir du fichier de référence. L'écart observé n'est qu'une illustration de la marge d'erreur qui affecte tout échantillon.

* Le nettoyage du fichier initial ne modifie pas sensiblement sa structure.

* Les refus de répondre ne modifient pas d'une manière significative la structure du fichier des personnes qui étaient à même de répondre.

  Le fait le plus notable est, sans doute, une légère sous-représentation du canton de Luxembourg. Ce canton est le seul pour lequel l'écart entre le fichier des adresses utilisées et le fichier final est plus élevé que l'écart entre le fichier aléatoire (342 adresses) et le fichier de l'échantillon initial: (28.9 - 26.5) > (29.9 - 29.3).

Ce point mérite d'être évoqué parce que cette tendance s'accentuera au cours de la seconde vague d'enquête: les habitants des grandes métropoles marquent souvent plus de méfiance à l'égard des enquêteurs que les personnes demeurant dans d'autres cantons. Ce phénomène se manifeste dès la première vague d'enquête. Il n'est pas encore significatif. Par un effet de cumul, il se marquera davantage au cours de la seconde vague d'enquête. Mais il commencera à s'estomper à partir de la troisième vague et il sera supplanté par d'autres facteurs.

Dans l'ensemble, il ne semble donc pas utile d'introduire des corrections dans l'échantillon des titulaires principaux.

Les moyens disponibles n'ont pas permis de procéder à beaucoup d'autres tests. Une analyse effectuée sur la base d'une combinaison des deux variables n'a pas conduit à d'autres conclusions. Les taux de réponse ont été comparés en fonction des listes de tirage et du sexe puis en fonction des listes de tirage et des cantons. Mais ces différentes analyses n'ont apporté aucun élément nouveau justifiant l'usage d'une procédure de pondération.

3.1.5.2. *Un biais méthodologique*

La rigueur méthodologique exige que soit mentionnée ici l'existence d'un léger biais provenant de la procédure de remplacement des refus de répondre.

Ce biais n'est intervenu qu'au terme de la période d'enquête. Il n'affecte qu'une partie de la quatrième liste des adresses utilisées pour remplacer les refus de répondre. Cette liste comprend 422 adresses, soit 12.8 % des 3290 adresses mises à la disposition des enquêteurs.

Au cours de cette dernière période, le manque de coopération des habitants de certains cantons a induit une certaine tendance à remplacer les personnes qui refusaient de répondre par des habitants du même canton. Il n'est pas possible d'estimer l'ampleur exacte de ce phénomène.

Cette procédure comporte un inconvénient majeur: les unités d'analyse ne sont plus choisies de manière probabiliste. Elles forment un sous-groupe caractérisé par un critère d'homogénéité interne artificiel: la disponibilité des personnes à l'égard de l'enquête. Les personnes difficiles à contacter pour des raisons indépendantes de leur volonté, les personnes auprès desquelles il faut insister, les personnes opposées à toute participation sont sous-représentées.

Cette méthode permet de sauvegarder la distribution des fréquences attendue mais elle biaise les estimateurs statistiques. Elle exclut toute possibilité d'identifier les caractéristiques des personnes qui refusent de répondre et de corriger les biais.

Le seul outil disponible pour estimer l'effet de ce biais consiste à comparer la répartition des membres de l'échantillon, selon les cantons de résidence, à deux moments:

- au terme de la troisième liste d'adresses

- et au terme de la quatrième liste.

Cette procédure est peu précise puisque le biais n'a pas affecté toute la quatrième liste des adresses de remplacement. Elle permet néanmoins de soupçonner l'existence d'un léger glissement dans le taux de réponse des habitants du canton de Luxembourg.

TABLEAU 3.3.
Répartition des titulaires selon le canton de résidence (en %)

CANTON	Echant. aléatoire	Echant. initial	Echant. après 3 listes	Echant. final: 4 listes
Luxembrg. ville	20.2	21.1	17.5	18.5
Luxembrg. campagne	9.1	8.8	6.6	8.0
Esch/Alz.	28.7	32.4	36.2	34.8
autres	42.0	37.7	39.7	38.7
N =	342	3290	1782	2012

Le canton de Luxembourg est légèrement sous-représenté au moment où les trois premières listes d'adresses sont épuisées: il représente 24.1 % des réponses contre 29.9 % des adresses dans le fichier initial et 29.3 % des adresses dans l'échantillon aléatoire de référence.

Lorsque les adresses fournies par la quatrième liste sont prises en compte le canton de Luxembourg représente 26.5 % des réponses. Ce redressement ne suffit pas pour

rendre à la capitale, la place qu'elle devrait occuper. Mais il semble bien que des efforts particuliers ont été déployés au cours de la dernière période de l'enquête pour combler le déficit enregistré dans ce canton. Il est possible que la sous-représentation du canton de Luxembourg aurait été plus importante si cette correction n'était pas intervenue au terme de l'enquête.

L'ampleur de ce biais ne doit pas être exagérée.

En tout état de cause, il est impossible d'y remédier.

3.2. VARIABLES DE PONDERATION

REMARQUE

Les variables de pondération des échantillon observés en 1985 s'appliquent à l'ensemble de ces échantillons, soit 2012 ménages et 6110 personnes.

La variable annuelle DIA est sans objet puisque toutes ces personnes ont le même statut. Leur trajectoires ne se différencient pas encore.

La variable annuelle DIA sera utile à partir de la seconde vague d'enquête.

Néanmoins, ces personnes et ces ménages sont insérés dans un fichier global dont la taille s'accroît chaque année.

Pour isoler ces personnes, la variable DIA élaborée en 1986 peut être utile. Les variables de pondération des échantillons observés en 1985 seront donc présentées en fonction de leur ventilation sur la variable DIA.

Ceci permettra d'indiquer quels sont les vecteurs décrits par la variable DIA qu'il faut sélectionner pour reconstituer l'échantillon observé en 1985.

3.2.1. **Variable instrumentale "MIREV"**

Cette variable mesure le nombre de titulaires principaux de revenus dans le ménage. Elle permet de calculer les variables de pondération des individus et des ménages en 1985 et de corriger le biais de sélection initial.

Chaque titulaire principal conduit à une adresse (**3.1.4.2.**). Celle-ci définit l'unité du ménage.

Les renseignements collectés au sein de chaque ménage prennent en compte toutes les personnes qui résident à cette adresse.

Lorsqu'un ménage est identifié, tous ses membres appartiennent à l'échantillon. Toutes les personnes appartenant à un ménage ont donc la même probabilité de

sélection initiale. Aucun biais n'intervient au moment où s'effectue ce changement d'unité, du ménage vers ses membres.

Par contre, un biais s'est introduit lors du passage de l'unité de tirage (titulaires principaux) vers les unités d'observation (ménages et membres).

En effet, plusieurs titulaires principaux peuvent conduire à un seul ménage. Par conséquent, certains ménages sont sur-représentés dans l'échantillon initial tandis que d'autres sont sous-représentés. La probabilité de sélection initiale d'un ménage est proportionnelle au nombre de titulaires principaux qui conduisent à son adresse.

Ce biais provoque des déformations proportionnelles dans l'échantillon du ménage: les membres des ménages sur-représentés sont, eux- mêmes, sur-représentés. Le facteur de correction est donc identique aux deux niveaux d'analyse.

La variable " MIREV " dénombre les titulaires principaux dans le ménage. Elle permet de connaître la probabilité de tirage du ménage et de corriger le biais de sélection.

Ce facteur de correction correspond à l'inverse de la probabilité de sélection.

3.2.2. Variable instrumentale "PONREV"

" PONREV ", l'inverse de " MIREV ", remplit cette fonction de correction, tant au niveau des ménages qu'au niveau des individus.

TABLEAU 3.4.

Distributions des fréquences des ménages et des individus (en valeurs absolues), selon "MIREV" et "PONREV"

"MIREV"	"PONREV"	FREQUENCES	
N. de titulaires	1/MIREV	MENAGES n=	INDIVIDUS n=
1.	1.00	1263	3134
2.	.50	480	1650
3.	.33	198	899
4.	.25	50	281
5.	.20	18	121
6.	.17	3	25
	TOTAL	2012	6110

Le tableau **3.4.** se lit de la manière suivante:

Lorsqu'un ménage est caractérisé par la présence d'un seul titulaire principal, sa probabilité de sélection est égale à 1. L'inverse de cette probabilité est égal à 1.0; 1263 ménages appartenant à l'échantillon du panel présentent cette caractéristique. Ils regroupent 3134 personnes. Le poids relatif de ces 1263 ménages et de ces 3134 personnes est égal à 1.0.

Lorsqu'un ménage est caractérisé par la présence de 6 titulaires principaux, sa probabilité de sélection est égale à 6. L'inverse de cette probabilité est égal à .17; 3 ménages appartenant à l'échantillon présentent cette caractéristique. Ils regroupent 25 personnes. Le poids relatif de ces 3 ménages et de ces 25 personnes est égal à .17.

TABLEAU 3.5.
Ventilation des valeurs de "MIREV" et "PONREV" dans le fichier individuel

DIA (-gnostic)	MIREV n =	PONREV n =
0. memb. 85/refus 86	833	833
1. memb. 85/memb. 86	5167	5167
2. nouveaux memb.-panel 86		-
3. nouveaux memb. non-panel 86		-
4. memb. 85 décédés/émigrés 86	110	110
Total	6110	6110
sysmis	224	224
TOTAL	6334	6334

Les valeurs de ces deux variables sont accessibles dans le fichier individuel. Elles sont attribuées aux 3 catégories formant l'échantillon de 1985, soit les groupes #0, #1 et #4. Elles ne sont pas attribuées aux individus qui entreront dans le fichier en 1986, en raison des principes de pondération qui seront appliqués à ces deux groupes (Tableau 3.5).

La variable " PONREV " n'est pas une variable de pondération opérationnelle. Lorsque chaque ménage est multiplié par son coefficient de correction, il ne vaut plus " une unité " mais une fraction d'unité: un ménage composé de 6 titulaires principaux n'a plus la valeur d'une unité. Il vaut désormais 1/6 d'unité, soit '.17'.

Si cette variable était utilisée comme facteur de pondération, la taille de l'échantillon se trouverait réduite. Les estimateurs statistiques ne seraient plus fiables.

Les valeurs de " PONREV " doivent être divisées par une constante qui permet de rétablir la taille initiale de l'échantillon. Les deux variables de pondération opérationnelles pour 1985, MWGT85A pour les ménages et WGT85 pour les individus, résultent de cette opération.

3.2.3. Variable opérationnelle de pondération des ménages en 1985 "MWGT85A"

Cette variable permet de pondérer correctement l'échantillon des MENAGES observés en 1985.

Elle est calculée directement au niveau de chaque ménage, en en ajustant les valeurs de "PONREV" afin de ne pas modifier la taille de l'échantillon des ménages: la somme totale des valeurs de MWGT85A est égale à 2012.

Le poids du ménage est ventilé sur chacun des membres du ménage afin de rassembler toutes les variables de pondération dans un seul fichier établi au niveau des individus. La valeur de pondération de chaque ménage peut être reconstituée en agrégeant les valeurs attribuées à chacun de ses membres. Tous les membres du ménage ayant reçu la même valeur, la moyenne de ces valeurs peut être utilisée au titre de valeur de pondération du ménage.

TABLEAU 3.6.
Ventilation des valeurs de "MWGT85A" dans le fichier individuel

DIA (-gnostic)	MWGT85A n =
0. memb. 85/refus 86	833
1. memb. 85/memb. 86	5167
2. nouveaux memb.-panel 86	-
3. nouveaux memb. non-panel 86	-
4. memb. 85 décédés/émigrés 86	110
Total	6110
sysmis	224
TOTAL	6334

Le tableau **3.6.** présente les catégories de personnes qui doivent être prises en compte pour reconstituer les poids des ménages en 1985. (Les groupes #2 et #3 n'ont pas reçu de valeurs puisqu'ils n'entreront dans la formation des ménages qu'en 1986).

3.2.4. Variable opérationnelle de pondération individuelle en 1985 "WGT85"

Cette variable permet de pondérer correctement l'échantillon des INDIVIDUS interrogés en 1985. Elle a été calculée afin de corriger le biais introduit par la variation du nombre de titulaires principaux dans les ménages.

Le calcul des pondérations est effectué selon les principes déjà énoncés.

Le poids des individus correspond à l'inverse du nombre de titulaires principaux (PONREV). Ces poids sont ajustés afin de sauvegarder la taille initiale de l'échantillon. (Les tableaux **A.1**, **A.2** et **A.3** présentés en annexe donnent une illustration des transformations progressives de la distribution des individus en fonction du nombre de titulaires principaux dans le ménage).

Trois points sont à souligner.

1. Tous les membres d'un même ménage reçoivent nécessairement le même poids puisqu'ils sont affectés du même biais dont l'origine se trouve au niveau du ménage. Cette propriété ne se vérifie que dans la première vague.

Par la suite, l'évolution de la structure des ménages rendra cette propriété caduque.

2. Les personnes ont reçu le même poids que leur ménage à une CONSTANTE près, à savoir: l'écart entre les valeurs d'ajustement destinées à éviter une modification de la taille initiale des échantillons.

Cette constante mise à part, seule la variable " PONREV " intervient dans la pondération des deux échantillons. Ceci est logique puisque le biais d'origine intervient au niveau du ménage et s'applique sans autre interférence à tous les membres qui résidaient à la même adresse.

L'évolution de la structure des ménages rendra cette propriété caduque dans les vagues d'enquêtes ultérieures: les membres du ménage pourront recevoir des poids différents du poids du ménage.

3. Ces deux propriétés entraînent une conséquence évidente: le poids du ménage est égal à la valeur moyenne des poids de ses membres (à une constante près).

ménage. Le poids du ménage ne sera plus nécessairement égal au poids de chacun de ses membres, mais il restera égal à la moyenne des poids de ses membres.

TABLEAU 3.7.
Ventilation des valeurs de "WGT85" dans le fichier individuel

DIA (-gnostic)	WGT85 n =
0. memb. 85/refus 86	833
1. memb. 85/memb. 86	5167
2. nouveaux memb.-panel 86	
3. nouveaux memb. non-panel 86	-
4. memb. 85 décédés/émigrés 86	110
Total	6110
sysmis	224
TOTAL	6334

Ces pondérations ont été attribuées à toutes les personnes présentes dans l'échantillon en 1985, soit les groupes #0, #1 et #4 de la variable DIA (Tableau 3.7.).

Ces valeurs sont donc ventilées sur 6110 personnes.

Annexe

MODIFICATION DE LA DISTRIBUTION DES INDIVIDUS, EN FONCTION DU NOMBRE DE TITULAIRES PRINCIPAUX DANS LE MENAGE

- AVANT PONDERATION
- APRES PONDERATION PAR " PONREV "
- APRES AJUSTEMENT PAR " WGT85 "

Tableau **A.1.** Distribution des individus selon le nombre de titulaires principaux dans le ménage, AVANT PONDERATION

"MIREV"	FREQUENCES AVANT PONDERATION		"PONREV"
	absolues	relatives	
1.	3134	49.5	1.00
2.	1650	26.0	.50
3.	899	14.2	.33
4.	281	4.4	.25
5.	121	1.9	.20
6.	25	.4	.17
TOTAL	6110	100.0	

Le tableau **A.1.** doit être lu de la manière suivante:

- 3134 personnes sont membres d'un ménage dans lequel il n'existe qu'un seul titulaire principal (MIREV). Elles n'ont qu'une seule chance d'apparaître dans l'échantillon. Leur coefficient de correction vaudra 1.0 (PONREV). Elles représentent 49.5% du nombre total des membres de l'échantillon de 1985;

- 25 personnes sont membres d'un ménage dans lequel il existe 6 titulaires principaux. Elles auraient pu apparaître 6 fois dans l'échantillon. Leur coefficient de correction vaudra .17, soit 1/6 d'unité. Elles représentent 0.4% de l'échantillon total de 1985.

Tableau **A.2.** Distribution des individus selon le nombre de titulaires principaux dans le ménage, APRES PONDERATION par PONREV

"PONREV"	FREQUENCES absolues		FREQUENCES relatives	
	avant pondér.	après pondér.	avant pondér.	après pondér.
1.00	3134	3134	49.5	72.0
.50	1650	825	26.0	18.9
.33	899	297	14.2	6.8
.25	281	70	4.4	1.6
.20	121	24	1.9	.6
.17	25	4	.4	.1
TOTAL	6110	4354	100.0	100.0

Le tableau **A.2.** se lit de la manière suivante:

- 3134 personnes ont reçu un coefficient de correction de 1.0 puisqu'elles appartiennent chacune à un ménage où il n'existe qu'un seul titulaire principal. Le poids de ces personnes n'est pas modifié lorsque l'échantillon est pondéré par PONREV;

- par contre, 1650 personnes ont reçu un coefficient de correction de .50 parce qu'elles sur-représentent les ménages où il existe deux titulaires principaux. La pondération de l'échantillon par la variable PONREV n'accorde plus qu'un poids d'une demi-unité à ces individus. En valeur absolue, ils ne sont plus que " 825 " unités;

- au total, les 6110 personnes " pondérées " ne représentent plus que 4354 unités;

- par conséquent, les 3134 personnes du premier groupe représentent désormais 72.0% de l'échantillon total et non plus 49.5%; les 1650 personnes du deuxième groupe ne représentent plus que 18.9% de l'échantillon total pondéré; ...

Tableau **A.3.** Distribution des individus selon le nombre de titulaires principaux dans le ménage, APRES PONDERATION par WGT85

"PONREV" (poids)	FREQUENCES APRES PONDERATION			"WGT85" (poids)
	absolues avant ajustement	relatives	absolues après ajustement	
1.00	3134	72.0	4398	1.40
.50	825	18.9	1158	.70
.33	297	6.8	416	.46
.25	70	1.6	99	.35
.20	24	.6	34	.28
.17	4	.1	6	.23
TOTAL	4354	100.0	6110	TOTAL

Le tableau **A.3.** illustre l'effet de l'ajustement des poids individuels.

L'échantillon doit retrouver sa taille initiale pour sauvegarder ses propriétés statistiques. Les poids des 4354 unités disponibles doivent être ajustés à cette fin. Cet ajustement conduit à l'élaboration de la variable WGT85.

Le tableau **A.3.** se lit de la manière suivante:

1. par application de la variable "PONREV", le taux moyen de réduction de la taille de l'échantillon est de .712 (soit 4354 / 6110).

2. Pour corriger cet effet de réduction du nombre d'unités, il suffit de multiplier chaque poids (PONREV) par l'inverse du taux de réduction (1 x (1 / .712) = 1.40; 0.5 x 1.4 = .70; ...).

   Soit: WGT85 = PONREV / (4354 / 6110)
   = PONREV x 1.40

3. Tous les poids étant ajustés en fonction d'une constante, les fréquences relatives ne sont plus modifiées: la fréquence absolue de chaque groupe s'élève de manière proportionnelle.

4. Lorsque l'échantillon est pondéré par la variable WGT85 (tab. **A.3.**):

- les membres d'un ménage dans lequel il existe un seul titulaire principal, représentent 4398 unités, soit 72.0% d'un échantillon qui a retrouvé sa taille initiale

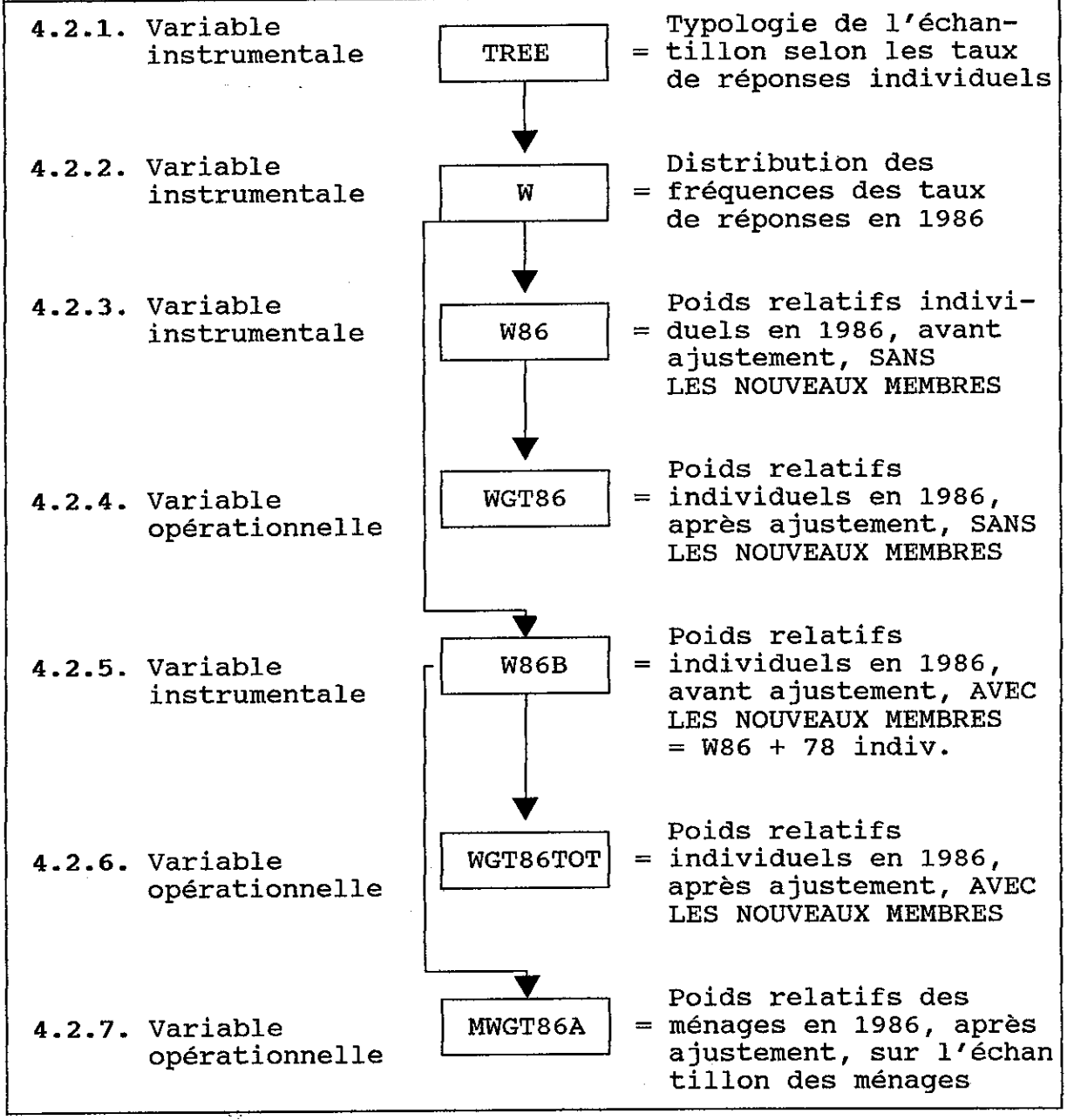
- les membres d'un ménage dans lequel il existe deux titulaires principaux, représentent 1158 unités, soit 18.9% d'un échantillon qui a retrouvé sa taille initiale

- ...

ÉCHANTILLON 1986

ET VARIABLES DE PONDÉRATION

Encadré 4.1. **VARIABLES DE PONDERATION 1986**



Chapitre 4

ÉCHANTILLON 1986

ET VARIABLES DE PONDÉRATION

L'échantillon observé en 1986 n'est plus conforme à l'échantillon observé en 1985. Plusieurs phénomènes sont à l'origine de cette évolution.

Certains évènements d'ordre démographique sont observables dans la population de référence. Ils ne doivent pas être pris en compte dans la procédure de pondération. Ils expliquent pourquoi le nouvel échantillon ne pourra jamais présenter le même profil que l'échantillon précédent.

D'autres évènements sont liés à la vie de l'échantillon et ne reflètent pas l'évolution de la population de référence: ils sont parfois à l'origine de biais systématiques dans l'échantillon.

La procédure de pondération s'applique à sauvegarder les propriétés statistiques de l'échantillon. Elle doit donc prendre en compte l'ensemble de ces phénomènes et les traiter en fonction de leur impact sur l'évolution de l'échantillon.

La variable DIA classe les personnes en fonction de leur trajectoire dans l'étude. Elle a été utilisée pour présenter les personnes observées en 1985 parce qu'elle permet de les isoler dans l'ensemble du fichier mais elle ne prend tout son sens que dans le cadre d'une perspective longitudinale: que sont devenues en 1986, les personnes interrogées en 1985? Comment l'échantillon a-t-il évolué?

4.1. L'ÉCHANTILLON 1986

TABLEAU 4.1.
Trajectoires individuelles 1985-1986

DIA	n =
0. memb. 85/refus 86	833
1. memb. 85/memb. 86	5167
2. nouvx.memb.-panel 86	78
3. nouvx.memb.-non panel 86	146
4. memb. 85/émigrés, décédés 86	110
TOTAL	6334

Le nouveau fichier contient désormais 6334 personnes classées en cinq catégories:

- les membres du panel ayant participé à l'étude en 1985 mais qui ont refusé de répondre en 1986;

- les membres du panel présents dans les deux échantillons;

- les nouveaux membres du panel entrés dans l'échantillon en 1986;

- les nouveaux membres de l'échantillon qui ne sont pas membres du panel;

- les membres du panel, présents dans l'échantillon en 1985 mais absents en 1986, soit parce qu'ils ont émigré (et dans ce cas, susceptibles de réapparaître ultérieurement dans l'échantillon), soit parce qu'ils sont décédés.

4.1.1. Les anciens membres du panel

Parmi les 6110 personnes observées en 1985, 5167 sont toujours présentes en 1986: 84.6 % des membres de l'échantillon initial ont donc accepté de participer à la seconde vague d'enquête.

Le taux de réponses exprimé de cette manière ne prend en compte ni les corrections des poids individuels effectuées en 1985, ni l'érosion de l'échantillon liée aux phénomènes démographiques. Il ne fournit pas une indication exacte de l'usure de l'échantillon.

4.1.2. Les refus

Certaines personnes présentes en 1985 ont refusé de poursuivre leur collaboration en 1986: 833 personnes ont donc été "perdues" d'une vague à l'autre. L'absence de ces personnes peut modifier la structure de l'échantillon. Elle peut introduire un biais systématique si ces pertes sont concentrées dans une ou plusieurs catégories homogènes de la population.

Ces refus représentent 13.6 % de l'échantillon initial.

Le tableau **4.2.** montre que certaines catégories de personnes ont été assimilées aux refus (groupes #3 à #6), parce que leur disparition peut produire les mêmes effets de distortion dans l'échantillon: elles n'ont pas pu être interrogées pour des raisons propres à la vie de l'échantillon mais elles ne sont pas sorties de la population de référence. Ces personnes doivent donc être prises en compte dans la procédure d'analyse des taux de réponses.

4.1.3. Les émigrés et les décès

TABLEAU 4.2.
Que sont devenues, en 1986,
les personnes interrogées en 1985?

TRAJECTOIRES 1985 - 1986	FREQUENCES C.A.	%
1. interrogés en 85 et 86	5167	84.6
2. refus explicite	775	12.7
3. absents de longue durée	3	.0
4. absents de courte durée	6	.1
5. "split off"*	17	.3
6. introuvables	32	.5
7. émigrés	60	1.0
8. décédés	50	.8
TOTAL	6110	100.0

* "Split off" = refus individuels explicites de personnes ayant formé un nouveau ménage.

Le tableau **4.2.** montre que 110 personnes ont échappé à l'observation en 1986, soit parce qu'elles ont émigré, soit parce qu'elles sont décédées.

Elles représentent 1.8 % des membres de l'échantillon initial.

Ces causes de disparition ne biaisent pas l'échantillon puisque ces personnes ont quitté simultanément l'échantillon et la population de référence. Ces personnes ne doivent donc pas être prises en compte dans la procédure d'estimation des nouvelles probabilités de sélection.

4.1.4. Les nouveaux membres du panel

Le tableau **4.1.** montre que 78 personnes sont entrées dans l'échantillon observé en 1986 au titre de membres du panel. Ces nouveaux membres sont généralement des nouveaux-nés, descendants de membres du panel. Quelques fils ou filles rentrés au foyer peu après la première observation ont été assimilés aux nouveaux membres, mais ceci ne sera plus appliqué dans les échantillons suivants.

Ces nouveaux membres régénèrent l'échantillon au sens propre du terme: chaque année une nouvelle génération entre dans le panel, tandis que les générations des personnes plus âgées quittent progressivement l'échantillon.

Cette régénération ne biaise pas l'échantillon puisque ces nouveaux membres sont entrés simultanément dans la population de référence. Elle contribue, au contraire, à lui conserver un profil comparable à celui d'une population en constante évolution.

4.1.5. Les nouveaux non-membres du panel

Le tableau **4.1.** montre que 146 personnes sont entrées dans l'échantillon observé en 1986 sans bénéficier du titre de membres du panel. La plupart de ces personnes appartenaient probablement à la population de référence au moment où l'échantillon initial fut sélectionné. Elles auraient pu faire partie de la sélection, mais ce ne fut pas le cas: elles sont entrées dans un ménage membre du panel pour des raisons diverses. Ces liens se sont probablement créés au hasard des circonstances et la présence de ces personnes dans l'échantillon ne présente aucun caractère de nécessité.

Il est donc difficile d'estimer la probabilité de sélection de ces personnes. Elles appartiennent à l'échantillon mais elles n'appartiennent pas au panel. Elles appartiendront à tous les échantillons aussi longtemps qu'elles accepteront de poursuivre leur collaboration. Elles seront donc prises en compte par les analyses de données individuelles temporelles. Mais elles n'interviendront jamais dans les analyses de données synchroniques, puisqu'il est impossible de leur attribuer un poids relatif (**2.2.1.5., 2.2.1.6.**).

4.2. LES VARIABLES DE PONDERATION

4.2.1. Variable instrumentale TREE

Les refus de répondre constituent le facteur principal de la déformation de l'échantillon (2.2.1.1.).

TABLEAU 4.3.
Distributions des fréquences des individus membres de l'échantillon 1985, selon leur situation en 1986, avant et après pondération

DIAGNOSTIC INDIVIDUEL	FREQUENCES		
	avant pondération	après pondération (wgt85)	sans les départs naturels
1. membres 85 86	5167	5192	5192
2. refus et assimilés	833	804	804
3. décès et émigrés	110	114	-
TOTAL	6110	6110	5996

Le tableau **4.3.** permet d'observer l'évolution de l'échantillon initial, avant et après correction des probabilités de sélection. La troisième colonne présente l'état de l'échantillon ainsi pondéré, après la déduction des départs naturels. L'analyse des taux de réponses ne prend en compte que 5996 unités pondérées (ou 6000 personnes avant pondération).

Les refus ne posent aucun problème aussi longtemps qu'ils se répartissent de manière aléatoire dans l'échantillon. Par contre, ils mettent en péril la qualité de l'échantillon dès qu'ils se concentrent dans des catégories de population homogènes.Ils ont des effets particulièrement destructifs lorsqu'ils se concentrent dans les catégories les plus pertinentes au regard des objectifs de l'étude.

La méthode utilisée pour détecter l'existence de sous-groupes homogènes dans cet ensemble de "refus" est présentée dans le second chapitre (2.2.1.1.).

En résumé, elle consiste à diviser l'échantillon en sous-groupes mutuellement exclusifs par une succession de divisions binaires. L'enchaînement de ces divisions successives crée des filières ou séries de modalités. Les variables qui produisent ces divisions sont choisies parce qu'elles créent les contrastes les plus significatifs entre les taux de réponses des deux sous-groupes. Ces divisions sont interrompues dès qu'un sous-groupe ne peut plus être divisé en deux sous-groupes d'une taille minimale de 200 personnes. (Cf. Procedures and Tape Codes, 1978 Interviewing Year, Wave XI, a Supplement. P.S.I.D., I.S.R., Ann Arbor, Michigan, 1979, p.10).

Le résultat de cette analyse est présenté dans le schéma **4.1.**. Une description plus détaillée des résultats est présentée en **Annexe 4.A.**

Dix-sept divisions ont été effectuées à l'aide de huit variables. Dix-huit sous-groupes sont identifiés.

La variable dépendante est une variable binaire (0, 1) qui rend compte du taux de réponses ou du taux de survie dans le panel. Les 18 sous-groupes "n'expliquent" que 5.6 % de la variance totale de la variable dépendante, mais l'objectif de cette procédure n'était pas de rendre compte de la plus grande partie de cette variance. Des mesures ont été prises pour éviter que l'analyse identifie des sous-groupes marginaux de très petite taille. La procédure évite que la variance de l'échantillon soit démesurément accrue, en maintenant la taille des sous-groupes à un minimum de 200 personnes, parce qu'elle ne prend pas en compte des taux de réponses d'une valeur extrême.

Le faible pouvoir explicatif de la variable ainsi constituée provient également du fait que les variables démographiques utilisées au cours de la procédure divisent l'échantillon en modalités dont les taux de réponses sont peu contrastés (tableau **4.4.**).

Ces taux de réponses sont calculés:

* après correction des biais liés au mode de tirage de l'échantillon initial (1985);

* sans prendre en compte les personnes émigrées et les personnes décédées.

Ces taux de réponses sont donc calculés après avoir restauré l'équiprobabilité des unités d'analyse et sans prendre en compte l'usure "démographique" de l'échantillon.

COMPARAISON DES TAUX DE REPONSE DES SOUS-GROUPES ENTRE 1985 ET 1986

SCHEMA 4.1.

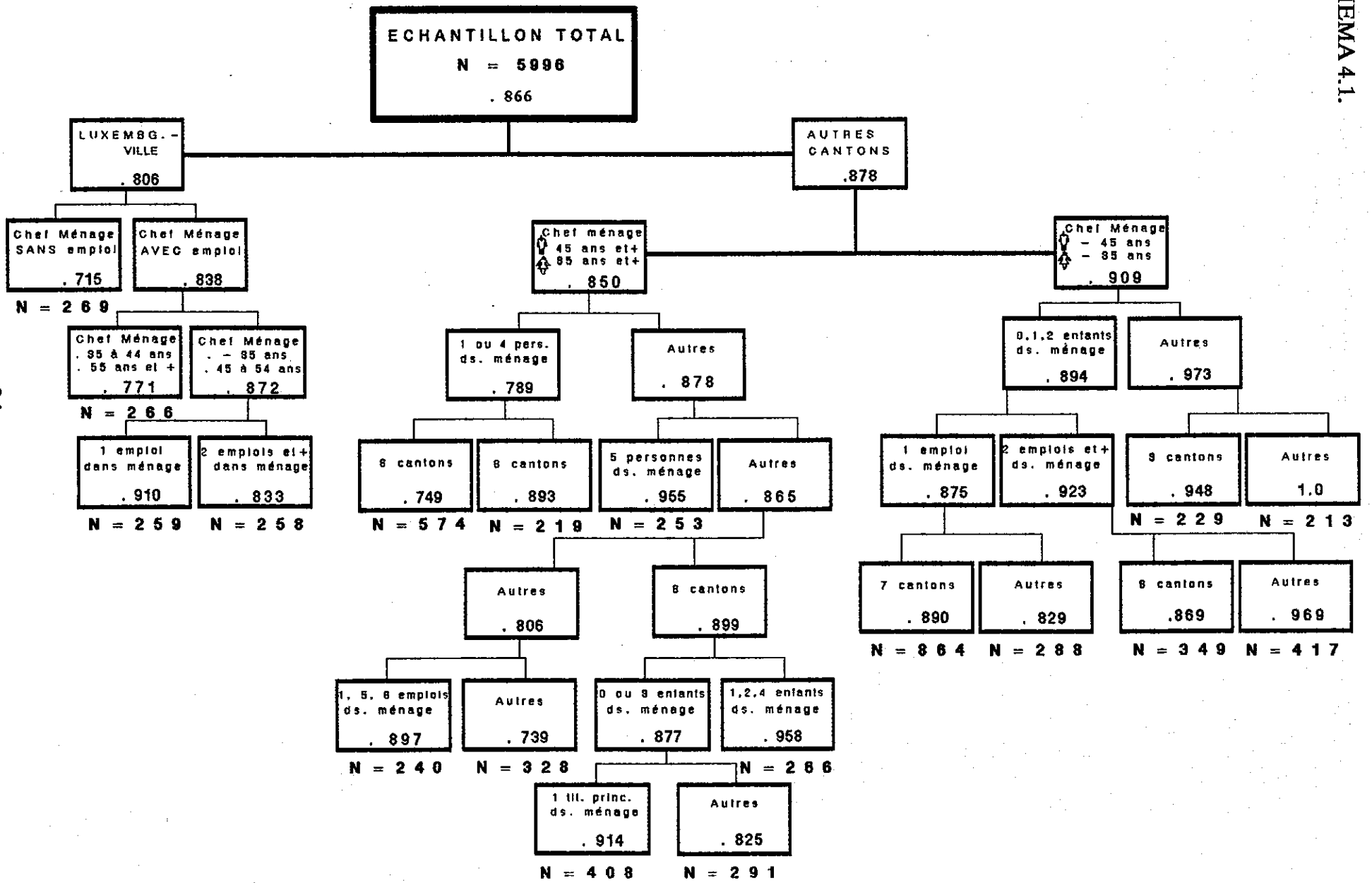


TABLEAU 4.4.

TAUX DE RÉPONSE EN 1986, DES PERSONNES INTERROGÉES EN 1985
- après exclusion des personnes décédées ou émigrées

CARACTÉRISTIQUES DES SOUS-GROUPES	TAUX DE REPONSE $\bar{x} = .86$	VARIANCE "EXPLIQUÉE" (ETA)	VARIANCE EXPLIQUÉE PAR LA MEILLEURE DICHOTOMISATION
Caractéristiques individuelles			
Sexe			
1. homme	.87	.01	.01
2. femme	.86		
Etat civil			
1. célibataire	.87	.06	.06
2. marié	.87		
3. séparé	.90		
4. divorcé	.86		
5. veuf	.79		
6. D.M.	(1.00)[2]		
Age			
1. < 17 ans	.89	.07	.06
2. 18 à 24 ans	.85		
3. 25 à 34 ans	.87		
4. 35 à 44 ans	.87		
5. 45 à 54 ans	.85		
6. 55 à 64 ans	.86		
7. 65 ans et plus	.80		
Nationalité			
1. Luxembourgeois	.85	.05	.04
2. CEE	.89		
3. hors CEE	.89		
4. D.M.	(.41)		

2. Les chiffres entre parenthèses correspondent à des catégories de moins de 20 personnes.

Situation professionnelle

1. Enfants	.88	.06	.04
2. travaille	.87		
3. congé maladie court/terme		.79	
4. congé maladie long/terme	.75		
5. Recherche (inscrit)	.79		
6. Recherche (non inscrit)	.68		
7. Pension - Invalidité	.84		
8. Retraite	.83		
9. Ménage	.85		
10. Milice	.87		

Lien avec le chef de ménage

1. Chef de ménage	.85	.06	.05
2. épouse	.87		
3. "épouse"	.95		
4. ami(e)	.91		
5. fils, fille	.87		
6. autre	.78		

Adultes

1. adulte	.85	.03	.03
2. enfant	.88		

Etat civil selon le sexe

1. Homme marié	.86	.06	.06
2. Homme non marié	.85		
3. Femme mariée	.87		
4. Femme non mariée	.81		
5. Enfant	.88		

Emploi selon le sexe

1. Homme ayant un emploi	.87	.05	.04
2. Homme sans emploi	.84		
3. Femme ayant un emploi	.85		
4. Femme sans emploi	.84		
5. Enfant	.88		

Age selon le sexe

1. Homme, de -de 18 ans	.89	.07	.05
2. Homme, de 18 à 64 ans	.86		
3. Homme, 65 ans et plus	.82		
4. Femme, de -de 18 ans	.90		
5. Femme, 18 à 64 ans	.86		
6. Femme, 65 ans et plus	.79		

Titulaire principal

1. oui	.85	.03	.03
2. non	.87		

Caractéristiques du ménage

Nombre de titulaires principaux dans le ménage			
1.	.87	.04	.04
2.	.83		
3.	.85		
4.	.86		
5.	.81		
6.	(1.00)		
Nombre d'emplois dans le ménage			
0.	.82	.05	.05
1.	.88		
2.	.86		
3.	.85		
4.	.84		
5.	.91		
6.	(1.00)		
Nombre d'adultes dans le ménage			
0.	(.00)	.08	.07
1.	.82		
2.	.88		
3.	.84		
4.	.81		
5.	.87		
6.	.78		
7.	(1.00)		
Nombre d'enfants dans le ménage			
0.	.83	.08	.07
1.	.88		
2.	.86		
3.	.92		
4.	.91		
5.	(1.00)		
6.	(.88)		
Nombre de personnes dans le ménage			
1.	.81	.07	.05
2.	.86		
3.	.86		
4.	.85		
5.	.91		
6.	.88		
7.	.85		
8.	.86		
9.	(.95)		
11.	(1.00)		

Type d'habitation

1. maison rurale	.86	.04	.03
2. maison unifamiliale individuelle	.85		
3. -"- jumelée	.87		
4. -"- en série	.86		
5. petit collectif	.88		
6. collectif moyen	.87		
7. grand collectif	.87		
8. précaire	(1.00)		
9. D.M.	.84		

Canton de résidence

1. Luxembourg-Ville	.80	.10	.09
2. Luxembourg-Campagne	.86		
3. Capellen	.83		
4. Esch-sur-Alzette	.88		
5. Mersch	.90		
6. Clervaux	.93		
7. Diekirch	.87		
8. Rédange	.85		
9. Vianden	.91		
10 Wiltz	.95		
11 Echternach	.83		
12 Grevenmacher	.86		
13 Remich	.86		

Caractéristiques du chef de ménage

Age du chef de ménage

1. < 25 ans	.89	.08	.07
2. 25 à 34 ans	.89		
3. 35 à 44 ans	.89		
4. 45 à 54 ans	.85		
5. 55 à 64 ans	.83		
6. 65 et plus	.80		

Sexe du chef de ménage

1. Homme	.87	.05	.05
2. Femme	.81		

Etat civil du chef de ménage

1. Célibataire	.87	.06	.06
2. marié	.87		
3. séparé	.92		
4. divorcé	.88		
5. veuf	.79		
6. D.M.	(1.00)		

Chef de ménage a un emploi

1. oui	.87	.06	.06
2. non	.82		

Etat civil selon le sexe du C.M.

1. Homme marié	.87	.05	.06
2. Homme non marié	.89		
3. Femme mariée	.76		
4. Femme non mariée	.81		

Emploi selon le sexe du C.M.

1. Homme ayant un emploi	.88	.07	.07
2. Homme sans emploi	.84		
3. Femme ayant un emploi	.84		
4. Femme sans emploi	.79		

Age selon le sexe du C.M.

1. Homme, < 35 ans	.89	.10	.09
2. Homme, 35 à 44 ans	.90		
3. Homme, 45 ans à 64 ans	.85		
4. Homme, 65 ans et plus	.81		
5. Femme, < 35 ans	.93		
6. Femme, 35 à 44 ans	.74		
7. Femme, 45 à 64 ans	.81		
8. Femme, 65 ans et plus	.79		

La variable TREE distingue 18 segments ou catégories de population. Ces 18 segments sont des ramifications de 3 branches maîtresses:

- les personnes qui résident à Luxembourg-ville

- les personnes qui ne résident pas à Luxembourg-ville et qui vivent dans un ménage dont le "chef de ménage" est:

    + un homme de 45 ans et plus
    + ou une femme de 35 ans et plus

- les personnes qui ne résident pas à Luxembourg-ville et qui vivent dans un ménage dont le chef de ménage est:

    + un homme de moins de 45 ans
    + ou une femme de moins de 35 ans.

Trois faits sont remarquables:

1. Le taux de réponses le plus élevé (1.0) caractérise les personnes appartenant à des ménages qui présentent le profil suivant:

    - elles habitent en-dehors de Luxembourg-ville, mais elles peuvent résider un peu partout dans le pays (9 cantons sur 12)

    - le chef de ménage est relativement jeune: si c'est un homme, il a moins de 45 ans; si c'est une femme, elle a moins de 35 ans

    - le nombre d'enfants présents dans ces ménages est beaucoup plus élevé que la moyenne nationale: 3, 4, 5 enfants contre moins d'un enfant par ménage, en moyenne, dans l'ensemble du pays.

2. Le taux de réponses le plus faible (0.71) s'observe dans des ménages résidants à Luxembourg-ville, lorsque le chef de ménage n'a pas d'emploi (chômeur, cherche un emploi, invalide, handicapé, retraité).

3. Les taux de réponses varient essentiellement en fonction des caractéristiques du ménage, du chef de ménage, ou du lieu de résidence du ménage. Des caractéristiques individuelles ont été mises en concurrence avec les caractéristiques du ménage à chaque étape de l'analyse. A aucun moment, ces caractéristiques individuelles ne se sont avérées aussi discriminantes que les caractéristiques du ménage.

TABLEAU 4.5.
Ventilation des valeurs de "TREE" dans le fichier individuel

DIA (-gnostic)	TREE n =
0. memb. 85/refus 86	833
1. memb. 85/memb. 86	5167
2. nouveaux memb.-panel 86	-
3. nouveaux memb. non-panel 86	-
4. memb. 85/décédés,émigrés 86	-
Total	6000
sysmis	334
TOTAL	6334

La variable " TREE " décrit le profil des catégories d'appartenance des individus, qu'ils aient répondu ou non en 1986. Elle n'exprime pas les probabilités de répondre. Elle ne permet donc pas de calculer le poids relatif de chaque individu dans l'ensemble de l'échantillon. Ce sera le rôle de la variable " W ".

4.2.2. Variable instrumentale "W"

La variable " W " est construite de la manière suivante: toute personne ayant effectivement répondu en 1986, reçoit la probabilité de répondre observée dans la catégorie de population à laquelle elle appartient (dans la variable TREE) (Tableau **4.6.**).

La variable " W " n'est plus une variable "nominale": elle se présente sous la forme d'une échelle. Elle comporte toujours 18 modalités, mais celles-ci ne se présentent plus dans le même ordre: l'échelle est tendue entre deux valeurs extrêmes qui correspondent, d'une part, au taux de réponses le plus faible et, d'autre part, au taux de réponses le plus élevé; entre ces deux extrémités, les modalités sont ordonnées en suivant l'ordre croissant de leur valeur.

TABLEAU 4.6.

Distribution des fréquences de l'échantillon 1986, selon les taux de réponses observés en 1986 (avant et après correction de l'échantillon initial)

" W " taux de réponses en 1986	"TREE" code	FREQUENCES	
		avant pondération n =	après pondération (WGT85) n =
.715	(1)	166	193
.739	(9)	305	243
.749	(5)	476	430
.771	(2)	191	205
.825	(11)	384	240
.829	(14)	181	239
.833	(4)	219	215
.869	(15)	287	303
.890	(13)	607	770
.893	(6)	226	196
.897	(8)	259	216
.910	(3)	175	236
.914	(10)	266	373
.948	(17)	169	217
.955	(7)	375	242
.958	(12)	286	256
.969	(16)	408	405
1.000	(18)	187	213
TOTAL		5167	5192

Il n'y a pas lieu de s'étonner si la taille de l'échantillon n'est plus identique avant et après le redressement de l'échantillon initial. Dans le cas présent, seule une fraction de l'échantillon initial (1985) a répondu en 1986: 5167 personnes sur 6110.

Or, ces 5167 personnes ont été pondérées dans le cadre de l'échantillon total, en vue de compenser les variations du nombre de titulaires principaux dans les ménages. Ces poids ont donc été ajustés en fonction de la taille de l'échantillon initial et non dans le cadre de la fraction des personnes ayant répondu (Cf. tableau **4.3.**).

Il est donc normal d'observer une différence entre les tailles de la FRACTION de l'échantillon, avant et après pondération.

La présentation de la dernière colonne du tableau **4.6.** s'imposait encore à un autre titre: l'analyse des taux de réponses a été effectuée en partant de l'échantillon initial pondéré. La comparaison des deux dernières colonnes du tableau **4.6.** permet de

constater le fait que certaines catégories de la population auraient été sur-représentées ou sous-représentées si cette précaution n'avait pas été prise.

TABLEAU 4.7.
Ventilation des valeurs de "W"
dans le fichier individuel

DIA (-gnostic)	W n =
0. memb. 85/refus 86	-
1. memb. 85/memb. 86	5167
2. nouveaux memb.-panel 86	-
3. nouveaux memb. non-panel 86	-
4. memb. 85/décédés,émigrés 86	-
Total	5167
sysmis	1167
TOTAL	6334

4.2.3. Variable instrumentale "W86"

La variable " W86 " définit le poids des personnes interrogées à deux reprises.

Ces personnes appartiennent à différentes catégories de la population et se distinguent par leur probabilité de répondre en 1986.

Lorsque le groupe enregistre un taux de réponses très faible, chaque "survivant" doit représenter un plus grand nombre de personnes présentant les mêmes caractéristiques. Lorsque le taux de réponses s'élève, l'importance de cet ajustement se réduit.

D'une manière générale, cet ajustement sera inversement proportionnel au taux de réponses enregistré dans le groupe.

Soit : 1 / W

Ce principe suffisait pour calculer les pondérations en 1985. En 1986, ce calcul doit intégrer les valeurs des poids relatifs des individus dans l'échantillon précédent. En effet, leur probabilité de sélection INITIALE a déjà dû faire l'objet d'une correction.

La pondération de 1986 ne s'applique donc pas à des individus "unités" mais à des fractions ou multiples de l'unité définis par la variable " WGT85 ".

La nouvelle variable de pondération se calcule selon la formule suivante:

$$W86 = WGT85 \times (1 / W)$$

TABLEAU 4.8.
Ventilation des valeurs de "W86" dans le fichier individuel

DIA (-gnostic)	W86 n =
0. memb. 85/refus 86	-
1. memb. 85/memb. 86	5167
2. nouveaux memb.-panel 86	-
3. nouveaux memb. non-panel 86	-
4. memb. 85/décédés,émigrés 86	-
Total	5167
sysmis	1167
TOTAL	6334

Les valeurs de cette variable sont attribuées exclusivement aux MEMBRES du PANEL qui ont répondu aux deux enquêtes [soit le groupe #1 de la variable DIA (-gnostic)(Tableau **4.8.**)].

4.2.4. Variable opérationnelle "WGT86"

La variable " W86 " n'est pas encore " opérationnelle ".

Si l'échantillon de 1986 était pondéré par ce facteur, le nombre total d'unités augmenterait de 5167 à 5997 fractions ou multiples de l'unité. Cette augmentation provient du fait que la plupart des membres de l'échantillon sont appelés à représenter un plus grand nombre de personnes qu'auparavant, suite aux refus de répondre de certains anciens membres.

Cette modification peut produire des effets indésirables: certaines statistiques sembleront significatives, simplement parce que la taille de l'échantillon est plus élevée.

Ces effets peuvent être annulés en divisant les poids par une constante qui restituera à l'échantillon sa taille initiale.

Soit: 1 / W devient (1 / W) / (5997 / 5167)

et W86 devient

WGT86 = WGT85 x ((1 / W) / (5997 / 5167))

TABLEAU 4.9.
Ventilation des valeurs de "WGT86" dans le fichier individuel

DIA (-gnostic)	WGT86 n =
0. memb. 85/refus 86	-
1. memb. 85/memb. 86	5167
2. nouveaux memb.-panel 86	-
3. nouveaux memb. non-panel 86	-
4. memb. 85/décédés,émigrés 86	-
Total	5167
sysmis	1167
TOTAL	6334

Il est important d'ajouter une remarque.

" WGT86 " ne prend pas en compte les NOUVEAUX MEMBRES du PANEL.

Les valeurs de WGT86 sont attribuées exclusivement aux membres du panel qui ont répondu aux deux enquêtes [soit le groupe #1 de la variable DIA (-gnostic)(Tableau **4.9**)].

Cette variable peut donc être utilisée afin de pondérer l'échantillon, lorsque l'analyse porte exclusivement sur les membres du panel qui appartenaient déjà au premier échantillon.

4.2.5. Variable instrumentale "W86B"

La variable " WGT86 " ne permet pas d'exploiter toutes les ressources de l'échantillon de 1986. Les nouveaux membres du panel et les nouveaux individus non membres du panel doivent être pondérés à leur tour, afin de permettre la réalisation d'analyses synchroniques dans l'ensemble de l'échantillon de 1986.

La variable " W86B " prend en compte les 78 nouveaux membres du panel; elle diffère, en cela, de la variable " W86 " et elle permet d'élargir les analyses synchroniques à l'ENSEMBLE DES INDIVIDUS MEMBRES DU PANEL EN 1986 (soit: 5245 personnes).

Les nouveaux membres de l'échantillon, NON MEMBRES DU PANEL reçoivent un poids de '0'.

TABLEAU 4.10.
Ventilation des valeurs de "W86B" dans le fichier individuel

DIA (-gnostic)	W86B n =
0. memb. 85/refus 86	-
1. memb. 85/memb. 86	5167
2. nouveaux memb.-panel 86	78
3. nouveaux memb. non-panel 86	146
4. memb. 85/décédés,émigrés 86	-
Total	5391
sysmis	943
TOTAL	6334

Les valeurs de cette variable sont attribuées à l'ensemble des personnes ayant pris part à l'enquête de 1986 [soit les groupes #1 #2 et #3 de la variable DIA (Tableau **4.10.**)].

Cette variable est purement instrumentale: les poids relatifs des individus doivent être ajustés afin de sauvegarder la taille réelle de l'échantillon.

4.2.6. Variable opérationnelle "WGT86TOT"

La variable " W86B " ne peut pas être utilisée comme telle pour pondérer l'échantillon de 1986; la taille de l'échantillon serait portée de 5245 unités à 6097 fractions ou multiples de l'unité.

Cette augmentation est due à trois facteurs:

1. chaque ancien membre du panel est appelé à représenter un plus grand nombre de personnes puisque certains membres du premier échantillon ont refusé de répondre à la deuxième enquête;

2. de nouveaux membres sont entrés dans l'échantillon au titre de membres du panel;

3. ces nouveaux membres ont reçu le poids relatif de leur ascendant direct.

Les poids relatifs doivent donc être ajustés afin que l'échantillon conserve ses propriétés statistiques.

La variable W86B devient:

$$WGT86TOT = WGT85 \times ((1 / W) / (6097 / 5245))$$

TABLEAU 4.11.
Ventilation des valeurs de "WGT86TOT" dans le fichier individuel

DIA (-gnostic)	WGT86TOT n =
0. memb. 85/refus 86	-
1. memb. 85/memb. 86	5167
2. nouveaux memb.-panel 86	78
3. nouveaux memb. non-panel 86	146
4. memb. 85/décédés,émigrés 86	-
Total	5391
sysmis	943
TOTAL	6334

Les valeurs de cette variable sont attribuées à l'ensemble des personnes ayant participé à l'enquête de 1986 [soit, les groupes #1, #2 et #3 de la variable DIA (Tableau 4.11.)].

VOIR EN ANNEXE 4C. Comparaison des caractéristiques démographiques des membres du panel en 1985 et 1986 après pondération

4.2.7. Variable opérationnelle de pondération des ménages en 1986 "MWGT86A"

La variable " MWGT86A " permet de pondérer correctement l'échantillon des MENAGES observés en 1986.

En 1985, le biais d'échantillonnage apparaît au niveau des ménages. La correction de ce biais s'effectue donc, en premier lieu, à ce niveau. Elle est répercutée sur les membres du ménage et ceux-ci reçoivent exactement le même poids que le ménage auquel ils appartiennent.

Il existe donc un rapport constant entre le poids du ménage et le poids des individus. Cette constante peut être définie de la manière suivante: le poids du ménage est égal à la valeur moyenne des poids des individus qui le composent (Cf. WGT85).

En 1986, l'échantillon peut être biaisé aux deux niveaux. Tantôt, un ménage complet disparaît et la raison du refus est commune à l'ensemble des membres du ménage. Tantôt, un seul membre du ménage refuse de répondre et disparaît pour des raisons strictement individuelles.

Il est possible de trouver un dénominateur commun à l'ensemble de ces refus: lorsqu'un ménage disparaît dans son ensemble, la raison de sa disparition est attribuée à chacune des personnes qui le composent. L'analyse des biais et le calcul des poids relatifs peuvent être effectués systématiquement au niveau individuel.

Néanmoins, le rapport entre la valeur des poids relatifs individuels et la valeur des poids relatifs des ménages garde sa propriété essentielle, il reste égal à la valeur moyenne des poids relatifs des individus.

Cette manière de procéder présente un avantage: elle permet de prendre en compte l'HÉTÉROGÉNÉITÉ des POIDS INDIVIDUELS au sein du ménage, hétérogénéité qui résulte inexorablement des modifications de la structure du ménage au cours du temps (voir **Annexe 4.B.: exemples de l'évolution des individus et des ménages**).

La procédure de calcul des pondérations des ménages est simple:

- Les valeurs des pondérations individuelles sont agrégées au niveau du ménage.

- La valeur moyenne des pondérations individuelles (W86B) détermine le poids du ménage.

- Cette valeur peut être ajustée en fonction de la taille de l'échantillon des ménages. Si ces poids étaient utilisés sans ajustement, la taille de l'échantillon s'élèverait à 1926 fractions ou multiples de l'unité. En 1986, l'échantillon des ménages compte 1793 unités. La variable de pondération " brute " doit donc être divisée par une constante de correction (1926 / 1793).

Les valeurs de cette variable ajustée (" MWGT86A ") ont été affectées à tous les membres de l'échantillon individuel (soit 5391 personnes) de telle manière que toutes les valeurs de toutes les variables de pondération se trouvent répertoriées dans un seul fichier individuel.

Il est aisé de reconstituer la variable opérationnelle: par exemple, on peut agréger au niveau du ménage les valeurs ventilées sur les individus et prendre la moyenne comme valeur de pondération.

Il faut rappeler que le poids relatif individuel (" WGT86TOT ") et le poids relatif du ménage (" MWGT86A ") vont désormais se différencier de plus en plus, d'année en année:

- ces poids sont ajustés en fonction de la taille des échantillons

- les poids individuels peuvent, désormais, varier au sein d'un même ménage.

TABLEAU 4.12.
Ventilation des valeurs de "MWGT86A" dans le fichier individuel

DIA (-gnostic)	MWGT86A n =
0. memb. 85/refus 86	-
1. memb. 85/memb. 86	5167
2. nouveaux memb.-panel 86	78
3. nouveaux memb. non-panel 86	146
4. memb. 85/décédés,émigrés 86	-
Total	5391
sysmis	943
TOTAL	6334

Le tableau **4.12.** présente les sous-échantillons individuels qu'il convient de sélectionner au moment de reconstituer les poids des ménages dans l'échantillon de 1986. Les personnes appartenant aux groupes #0 et #4 n'ont pas reçu de valeurs (code=SYSMIS). Elles n'entrent plus dans la formation des ménages en 1986.

ANNEXE 4.A.

Taux de survie, dans le panel, entre 1985 et 1986: Contrastes entre les taux de réponses de 18 segments de l'échantillon

SEGMENTS		"TREE" Code	N = avant pondération	N = après pondération	TAUX DE REPONSE
	1. RESIDENT A LUXEMBOURG-VILLE				
1.1.	chef de ménage sans emploi	1.00	242	269	.715
1.2.	chef de ménage a un emploi . est âgé de 35 à 44 ans ou 55 ans et plus	2.00	245	266	.771
1.3.	chef de ménage a un emploi . 1 seul emploi dans le ménage . chef de ménage est âgé de moins de 35 ans ou 45 à 54 ans	3.00	195	259	.910
1.4.	chef de ménage a un emploi .2 emplois et plus dans le ménage . chef de ménage est âgé de moins de 35 ans ou 45 à 54 ans	4.00	264	258	.833

Taux de survie, dans le panel, entre 1985 et 1986: Contrastes entre les taux de réponses de 18 segments de l'échantillon (suite)

SEGMENTS		"TREE" Code	N = avant pondération	N = après pondération	TAUX DE REPONSE
	2. NE RESIDENT PAS A LUXEMBOURG-VILLE Chef de ménage est . un homme de 45 ans et plus . une femme de 35 ans et plus				
2.1.	résident ds. 6 cantons (1) . 1 ou 4 personnes dans le ménage	5.00	625	574	.749
2.2.	résident ds. 6 cantons (2) . 1 ou 4 personnes dans le ménage	6.00	252	219	.893
2.3.	. 5 pers. dans le ménage	7.00	403	253	.955
2.4.	résident ds. 6 cantons (3) . 2, 3, 6 personnes et plus dans le ménage . 1, 5, 6 emplois dans le ménage	8.00	288	240	.897
2.5.	résident ds. 6 cantons (3) . 2, 3, 4 emplois dans le ménage	9.00	413	328	.739
2.6.	résident ds. 6 cantons (4) . 2, 3, 6 personnes et plus dans le ménage dont 0 ou 3 enfants . un seul titulaire principal	10.00	291	408	.914
2.7.	= 2.6. mais: . 2 titulaires principaux et plus dans le ménage	11.00	465	291	.825
2.8.	résident ds. 6 cantons (4) . 2, 3, 6 personnes et plus dans le ménage dont 1, 2, 4 enfants	12.00	301	266	.958

Taux de survie, dans le panel, entre 1985 et 1986: Contrastes entre les taux de réponses de 18 segments de l'échantillon (suite)

SEGMENTS		"TREE" Code	N = avant pondération	N = après pondération	TAUX DE REPONSE
	3. NE RESIDENT PAS A LUXEMBOURG-VILLE Chef de ménage est . un homme de moins de 45 ans . une femme de moins de 35 ans				
3.1.	résident ds. 7 cantons(5) . 0, 1, 2 enfants dans le ménage . un emploi ds le ménage	13.00	685	864	.890
3.2.	résident ds. 5 cantons(6) . 0, 1, 2 enfants dans le ménage . un emploi ds le ménage	14.00	218	288	.829
3.3.	résident ds. 6 cantons(7) . 0, 1, 2 enfants dans le ménage . 2 emplois et plus dans le ménage	15.00	324	349	.869
3.4.	résident ds. 6 cantons(8) . 0, 1, 2 enfants dans le ménage . 2 emplois et plus dans le ménage	16.00	422	417	.969
3.5.	résident ds. 3 cantons(9) . 3, 4, 5 enfants dans le ménage	17.00	180	229	.948
3.6.	résident ds.9 cantons(10) . 3, 4, 5 enfants dans le ménage	18.00	187	213	1.000
	TOTAL		6000		.866

NOTE : Les chiffres entre parenthèses (1) à (10) indiquent qu'il s'agit de différentes modalités de regroupement des 12 cantons du pays.

ANNEXE 4.B.

EXEMPLE DE L'EVOLUTION DES INDIVIDUS ET DES MENAGES

Au cours du déroulement du panel, les poids relatifs des individus se diversifient au sein des ménages. Trois types d'événements contribuent à cette diversification:

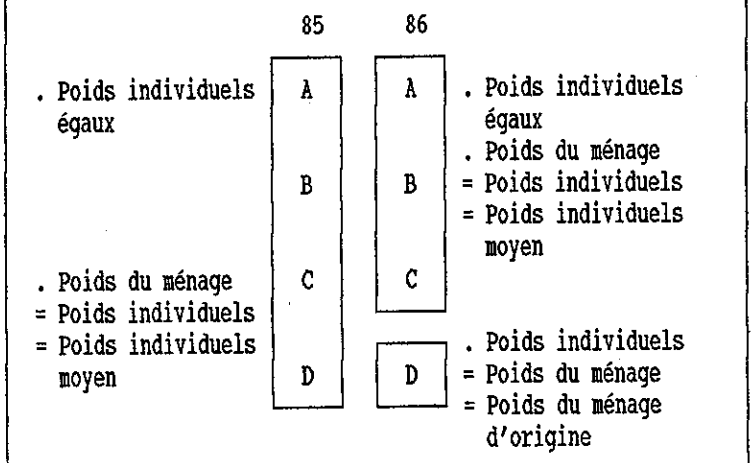
- l'insertion de nouveaux individus non-panel
- l'insertion de nouveaux individus-panel
- les modifications de la structure des ménages.

Envisageons un cas de figure.

Un ménage interrogé en 1985 est formé par 4 personnes: le père, la mère, un fils et une fille. Ces quatre personnes ont la même valeur de pondération. Le poids relatif du ménage est égal à la valeur moyenne de leur poids. En 1986, le fils quitte ses parents.

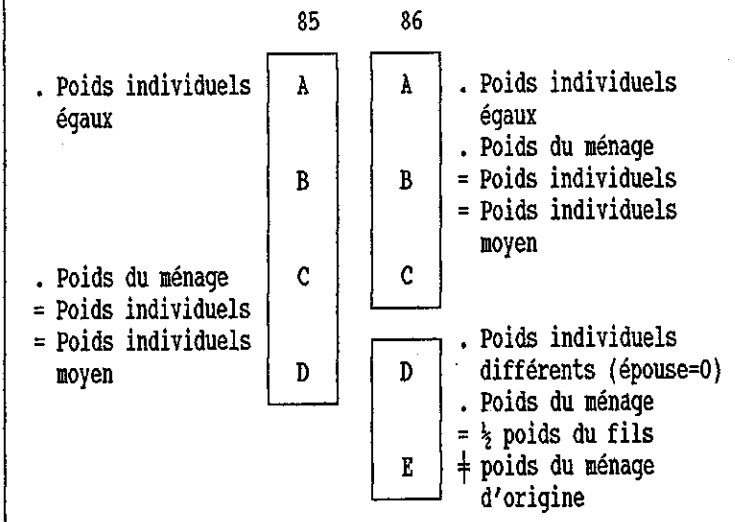
Plusieurs hypothèses peuvent être envisagées.

Première hypothèse



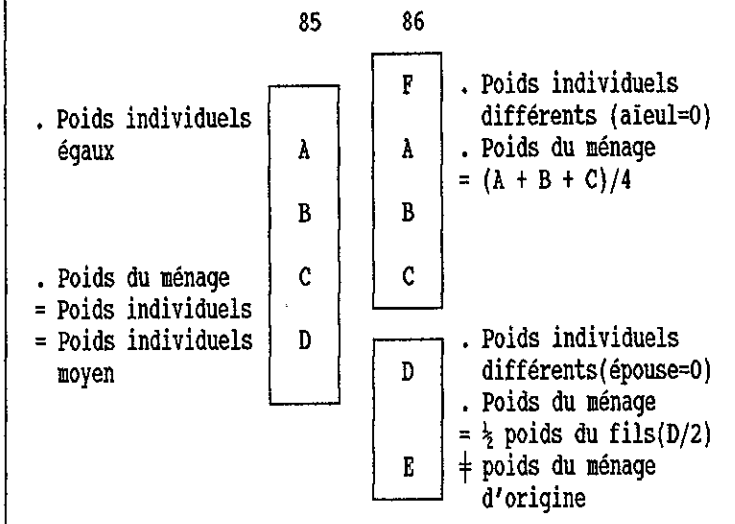
Supposons que le fils s'installe seul. D'une part,le fils forme un ménage et son propre poids individuel devient le poids du ménage qu'il forme à lui seul. D'autre part, le ménage initial ne compte plus que trois personnes. Ces trois personnes ont toujours des poids relatifs égaux et le poids du ménage reste égal à la moyenne des poids individuels (à une constante près).

Deuxième hypothèse



Dans l'hypothèse ou le fils se marie, son épouse, non-membre du panel, reçoit un poids relatif de " 0 ": leur ménage est formé de deux personnes dont les poids relatifs sont différents. Logiquement, le poids du ménage est égal à la moitié du poids relatif du mari.

Troisième hypothèse



Supposons que le grand-père maternel soit accueilli dans le ménage de sa fille, au cours de la même année. L'insertion d'une personne non-membre du panel provoque deux effets. D'une part, la structure du ménage initial est modifiée. D'autre part, le poids relatif du ménage est modifié. Le père, la mère et la fille conservent des poids identiques mais l'aïeul reçoit un poids égal à "0".
Désormais, le poids du ménage sera égal au quart de la somme des poids individuels du père, de la mère et de la fille.

Les deux ménages ainsi restructurés sont appelés à évoluer séparément. La structure des pondérations pourra encore se modifier par la suite.

C'est précisément ce qui invite à fonder le système de pondération dans les destinées individuelles. Cette approche nous paraît moins "réductrice" de l'histoire des ménages que d'autres procédures, adoptées par certains panels: prendre un seul membre du ménage comme base de référence, ou prendre des caractéristiques du ménage pour attribuer directement à tous les membres du ménage un même poids, sont des procédures d'homogénéisation qui ne permettent pas de prendre en compte l'évolution de la structure interne des ménages.

ANNEXE 4 C

TABLEAU COMPARATIF DES CARACTERISTIQUES DEMOGRAPHIQUES DES MEMBRES DU PANEL EN 1985 ET EN 1986

(après pondération)

- caractéristiques personnelles
- caractéristiques du ménage
- caractéristiques du chef de ménage

Caractéristiques Personnelles	**1985** (N = 6110)	**1986** (N = 5245)
Sexe		
1. homme	48.8	48.7
2. femme	51.2	51.3
Age		
1. < 18 ans	24.1	23.5
2. 18 - 24 ans	11.2	11.2
3. 25 - 34 ans	15.6	15.1
4. 35 - 44 ans	14.3	14.3
5. 45 - 54 ans	12.7	12.8
6. 55 - 64 ans	10.4	10.9
7. > 64 ans	11.7	12.1
Age par Sexe		
1. homme < 18 ans	12.7	12.4
2. homme 18 - 64 ans	31.7	31.8
3. homme > 64 ans	4.4	4.6
4. femme < 18 ans	11.4	11.2
5. femme 18 - 64 ans	32.6	32.6
6. femme > 64 ans	7.2	7.5
Etat civil		
1. célibataire	39.8	39.9
2. marié	49.6	49.2
3. séparé	1.0	0.8
4. divorcé	1.9	1.9
5. veuf	7.9	8.0
6. N.R.	0.0	-

Nationalité

1. Luxembourgeois	76.1	76.4
2. autres C.E.	21.3	21.3
3. hors C.E.	2.5	2.3
4. N.R.	0.1	-

Emploi

1. oui	38.1	39.8
2. non	35.4	35.3
3. enfants	26.5	24.9

Adulte/enfant

1. adulte	73.5	75.1
2. enfant	26.5	24.9

Lien avec le chef de ménage

1. chef de ménage	36.4	37.2
2. épouse	24.2	24.1
3. "épouse"	1.1	1.0
4. ami, amie	0.3	0.1
5. fils, fille	33.4	33.6
6. autres	4.6	4.0

Caractéristiques du ménage

Type d'habitation

1. maison rurale	11.4	10.7
2. unifamiliale individuelle	33.7	34.7
3. unifamiliale jumelée	13.5	8.8
4. unifamiliale en série	19.5	21.4
5. petit collectif	12.6	13.1
6. collectif moyen	7.1	7.3
7. grand collectif	1.1	1.4
8. précaire	0.3	0.0
9. N.R.	0.8	2.6

Canton		
1. Luxembourg-ville	17.6	17.6
2. Luxembourg-campagne	8.9	8.8
3. Capellen	6.9	7.0
4. Esch sur Alzette	33.5	33.6
5. Mersch	4.6	4.3
6. Clervaux	2.3	2.3
7. Diekirch	6.2	6.2
8. Redange	3.0	3.0
9. Vianden	0.8	0.7
10. Wiltz	3.0	3.1
11. Echternach	3.5	3.5
12. Grevenmacher	5.8	5.8
13. Remich	4.0	4.0
Nombre de personnes dans le ménage		
1.	7.6	8.0
2.	21.0	20.6
3.	22.5	23.1
4.	27.6	27.6
5.	12.3	12.7
6.	5.8	5.3
7.	2.1	1.8
8.	0.7	0.6
9.	0.3	0.3
11.	0.0	-
12.	-	0.1
Nombre d'enfants dans le ménage		
0.	39.2	41.9
1.	24.3	23.2
2.	22.9	23.2
3.	9.9	9.5
4.	3.5	2.2
5.	0.1	0.1
6.	0.1	-
Nombre d'adultes dans le ménage		
0.	0.1	-
1.	9.9	10.2
2.	63.0	59.5
3.	16.2	18.9
4.	7.7	6.9
5.	2.1	3.0
6.	0.8	1.2
7.	0.2	0.2
9.	-	0.1

Nombre d'emplois dans le ménage		
0.	17.1	17.5
1.	46.9	43.1
2.	29.2	29.8
3.	4.9	6.8
4.	1.4	2.1
5.	0.5	0.7
6.	0.0	0.0

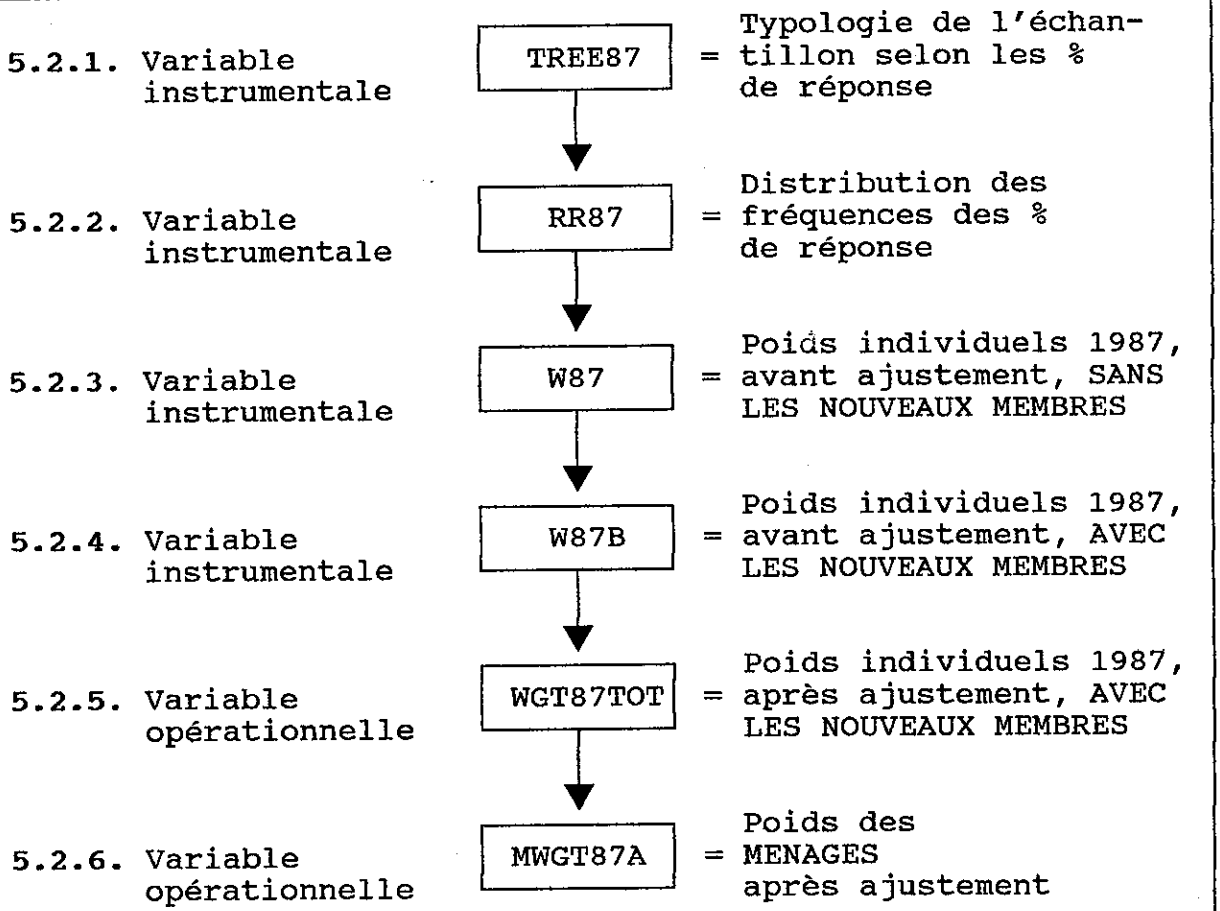
Caractéristiques du chef de ménage

Sexe		
1. homme	87.2	86.0
2. femme	12.8	14.0
Age		
1. < 24 ans	2.6	2.3
2. 25 - 34 ans	18.3	18.1
3. 35 - 44 ans	27.6	27.1
4. 45 - 54 ans	24.4	23.7
5. 55 - 64 ans	14.8	15.9
6. > 64 ans	12.2	12.9
Etat civil		
1. célibataire	6.0	6.1
2. marié	80.6	80.0
3. séparé	1.4	1.7
4. divorcé	2.5	2.7
5. veuf	9.5	9.6
6. N.R.	0.0	-
Emploi		
1. oui	74.1	73.3
2. non	25.9	26.7
Etat civil par Sexe		
1. homme non-marié	7.0	6.7
2. homme marié	80.2	79.3
3. femme non-mariée	12.4	13.4
4. femme mariée	0.4	0.6
Emploi par Sexe		
1. homme sans emploi	17.5	17.4
2. homme avec emploi	69.8	68.6
3. femme sans emploi	8.5	9.3
4. femme avec emploi	4.3	4.7

ÉCHANTILLON 1987

ET VARIABLES DE PONDÉRATION

Encadré 5.1. **VARIABLES DE PONDERATION 1987**



Chapitre 5

ÉCHANTILLON 1987

ET VARIABLES DE PONDÉRATION

L'échantillon observé en 1987 n'est plus conforme à l'échantillon observé en 1985. Phénomènes démographiques et phénomènes propres à la vie de l'échantillon conjuguent leurs effets et provoquent une dérive progressive de l'échantillon.

Les trajectoires des personnes dans l'étude se diversifient de plus en plus. La position des personnes dans la procédure de pondération dépend essentiellement de leur trajectoire.

La première partie du chapitre sera consacrée à la description de cette évolution de l'échantillon.

La seconde partie présentera les différentes variables qui ont jalonné l'élaboration du nouveau système de pondérations. Pour chaque variable, un tableau récapitule les sous-échantillons pertinents et la taille de ces sous-échantillons.

5.1. L'échantillon 1987

5.1.1. Les trajectoires individuelles: version courte

TABLEAU 5.1.
Trajectoires individuelles 1985 - 1987

DIA87	n =
0. memb. panel refus	1432
1. memb. panel réponse	4583
2. nouv. memb. panel 87	65
3. non-memb. panel réponse	266
4. non-memb. panel refus	20
5. émigrés / décédés	176
TOTAL	6542

Le nouveau fichier contient désormais 6542 personnes classées en 6 catégories principales. Le contenu de ces catégories évolue d'année en année.

Le nouvel échantillon compte 4914 personnes réparties en 3 classes: les membres du panel ayant répondu (4583), les nouveaux membres du panel entrés dans l'échantillon en 1987 (65) et les membres de l'échantillon qui ne sont pas membres du panel (266).

Le taux de réponses est calculé en rapportant le nombre des réponses des *membres du panel* (4583) au nombre total des membres du panel (6015), sans prendre en compte les nouveaux membres du panel entrés dans l'échantillon en 1987 (65) ni les personnes émigrées ou décédées.

Il convient de rappeler que ces 6 classes sont définies en fonction du point initial et du point final des trajectoires des individus dans l'ensemble de l'étude. Le point initial détermine le statut de la personne au moment ou elle est entrée dans l'étude: membre ou non-membre du panel. Le point final correspond à la présence ou à l'absence de la personne dans l'échantillon considéré.

Ceci signifie, par exemple, que les membres du panel

* entrés dans l'échantillon initial en 1985

* ayant refusé de répondre en 1986

* peuvent être classés, en 1987

- soit parmi les refus (groupe #0)

- soit parmi les réponses (groupe #1) s'ils ont accepté de renouveler leur coopération, s'ils ont rejoint un ménage d'origine toujours présent dans l'échantillon, etc ...

L'ensemble de toutes ces trajectoires est présenté au tableau 5.2.. Chaque trajectoire est définie par la succession des états occupés par les personnes au cours des années écoulées.

Le tableau 5.1. rassemble ces trajectoires en 6 classes principales parce que ces classes

* sont nécessaires et suffisantes pour permettre le calcul des nouvelles pondérations

* permettent de gérer facilement les sous-échantillons principaux qui forment l'ensemble du fichier individuel.

Le tableau 5.2. permet de recomposer d'autres échantillons en fonction des besoins de l'analyste.

La variable DIA87 comporte six grandes modalités qui peuvent être décrites de la manière suivante.

5.1.1.1. *Les refus des membres du panel*

La première catégorie est formée par les membres du panel ayant refusé de prendre part à l'enquête. Cet ensemble ne correspond plus à la première catégorie de la variable DIA utilisée en 1986. Elle réunit des membres du panel présents en 1985 ET des nouveaux membres entrés en 1986. Tous ces membres ont un point commun: ils n'ont pas répondu en 1987.

5.1.1.2. *Les membres du panel*

La deuxième catégorie (groupe #1) réunit des membres du panel présents en 1985 ET des nouveaux membres entrés en 1986.

Ils ont un point commun: ils ont répondu en 1987.

5.1.1.3. *Les nouveaux membres du panel*

Le groupe #2 isole les nouveaux membres du panel entrés dans l'échantillon en 1987. Ce sont des nouveaux-nés descendants directs de membres du panel.

Ce groupe ne contient pas les nouveaux membres du panel entrés en 1986. Ces derniers font désormais partie de l'ensemble des membres du panel et sont traités au même titre dans la procédure de pondération. Ils sont répartis entre le groupe #0, le groupe #1 et le groupe #5 en fonction de la trajectoire qu'ils ont suivi entre 1986 et 1987.

Le groupe #2 permet d'identifier les nouveaux membres de l'échantillon annuel qui recevront un poids individuel lié à celui de leurs ascendants. Les membres de ce groupe seront donc différents chaque année.

5.1.1.4. *Les non-membres du panel*

Les groupes #3 et #4 sont formés par l'ensemble des nouvelles personnes entrées dans les échantillons successifs sans accéder au titre de membres du panel.

Ces personnes sont classées dans le groupe #3 lorsqu'elles appartiennent à l'échantillon annuel. Les nouveaux non-membres du panel entrés dans l'échantillon en 1987 sont donc classés dans ce groupe au même titre que les non-membres du panel entrés dans l'étude au cours de vagues antérieures. Ils ont un point commun: ils ont répondu en 1987.

Les non-membres du panel, entrés dans les échantillons précédents, sont classés dans le groupe #4 lorsqu'ils n'ont pas participé à l'enquête annuelle.

Cette distinction entre les groupes #3 et #4 permet de reconstituer aisément l'échantillon annuel, lorsqu'il y a lieu de prendre en compte les membres du panel ET les non-membres du panel présents dans l'échantillon: il suffit de sélectionner les groupes #1, #2 et #3.

Cette procédure est très utile au moment de reconstituer les ménages ou les groupes de revenus.

Elle est indispensable au moment de calculer les poids individuels des nouveaux membres du panel: par exemple, lorsque l'un des ascendants est non-membre du panel, le poids du nouveau membre est différent selon que cet ascendant est présent ou absent de l'échantillon.

5.1.1.5. *Les émigrés et les décédés*

Le groupe #5 réunit toutes les personnes qui ont été présentes dans un échantillon au moins une fois, et qui ne sont pas présentes dans l'échantillon de l'année considérée pour une raison d'ordre *démographique*.

Cette distinction s'impose puisque les émigrés et les personnes décédées ne sont pas prises en compte dans le système de pondération annuel.

Leur classement dans une catégorie particulière facilite également la sélection de l'échantillon annuel.

Il faut rappeler qu'une personne émigrée peut rejoindre l'échantillon ultérieurement. Elle sera reclassée dans son groupe d'origine, c'est à dire dans le groupe #1 si elle était membre du panel ou dans le groupe #3 si elle n'était pas membre du panel.

5.1.2. Les trajectoires individuelles: version complète

TABLEAU 5.2.
Liste des trajectoires individuelles: 1985-1987

VARIABLE: DIA3

CODES	TRAJECTOIRES	n =
0.0	M. 85/0/0	828
1.0	M. 85/86/0	599
2.0	M. 0/86/0	5
3.0	N.M. 0/86/0	20
10.0	M. 85/0/87	5
11.0	M. 85/86/87	4502
12.0	M. 0/86/87	72
13.0	M. 85/E/87	4
14.0	M. 0/0/87	65
20.0	N.M. 0/86/87	123
21.0	N.M. 0/0/87	143
30.0	M. 85/E/E	56
31.0	M. 85/86/E	22
40.0	M. 85/D/D	50
41.0	M. 85/86/D	44
42.0	M. 0/86/D	1
43.0	N.M. 0/86/D	3
	TOTAL 1987	6542

Le tableau 5.2. est composé de la manière suivante.

La variable est nommée **DIA3** parce qu'elle correspond à l'état du fichier après la troisième vague d'enquêtes.

Les 6542 personnes recensées dans le fichier total après trois vagues d'enquêtes ont suivi 17 trajectoires différentes entre 1985 et 1987.

Chaque trajectoire est un vecteur défini par une série d'états successifs. Chaque état correspond au statut occupé par une personne dans le panel ET dans l'échantillon suite à une vague d'observation annuelle.

Chaque vecteur est désigné par un code. Ces codes ont été attribués de manière arbitraire et peuvent être modifiés. Tels qu'ils sont organisés dans le tableau **5.2.**, ils permettent d'effectuer rapidement certains regroupements.

Par exemple, les codes inférieurs à 10.0 désignent l'ensemble des personnes qui ont refusé de participer à l'enquête effectuée en 1987. Les codes 10.0 à 14.0 désignent les membres du panel présents dans l'échantillon en 1987.

Les vecteurs sont décrits de la manière suivante:

* **M.** signifie que ce vecteur contient exclusivement des **membres du panel.**

  N.M. signifie que le vecteur contient exclusivement des personnes **non membres du panel.**

* **85** signifie que ces personnes sont **membres de l'échantillon** observé en 1985.

  86 signifie que ces personnes sont membres de l'échantillon observé en 1986.

  L'année est remplacée par le code **0** lorsqu'une personne est **absente de l'échantillon** suite à un refus ou lorsqu'il n'est pas avéré que cette absence est liée à une raison d'ordre démographique.

  Par exemple: **85/86/0** désigne le vecteur des personnes

  - entrées dans l'échantillon au cours de la première vague d'enquête effectuée en 1985

  - toujours présentes dans l'échantillon observé en 1986

  - mais absentes de l'échantillon en 1987 pour des raisons assimilées à des refus.

* **E** signifie que ces personnes ont **émigré** et sont absentes de l'échantillon correspondant à la vague d'enquête de cette année. Dans certains cas cette absence est provisoire.

  Le vecteur #13 est composé de la manière suivante:

  M. 85/E/87 soit 4 membres du panel, entrés dans l'échantillon en 1985, émigrés en 1986 et revenus dans l'échantillon en 1987.

* **D** signifie que ces personnes sont **décédées.**

  Le vecteur #40 est composé de la manière suivante:

N.M. 0/86/D soit 3 personnes entrées dans l'étude en 1986 sans accéder au titre de membres du panel et décédées au cours de l'année 1987.

Le passage de la variable DIA3 à la variable DIA87 s'opère aisément. Les classes générales de trajectoires sont composées à partir du point initial et du point final de chaque vecteur.

Le **groupe #1** de la variable DIA87 regroupe l'ensemble des membres du panel entrés dans l'échantillon en 1985 ou en 1986 et présents dans l'échantillon en 1987, quelles que soit les étapes intermédiaires ayant jalonné leur trajectoire. Cette classe est donc formée par les vecteurs #10 à #13 de la variable DIA3:

10.0	**M. 85/0/87**	soit les membres du panel entrés en 1985, absents en 1986 mais revenus dans l'échantillon en 1987.
11.0	**M. 85/86/87**	soit les membres du panel entrés en 1985 et présents dans les trois vagues d'enquêtes.
12.0	**M. 0/86/87**	soit les nouveaux membres du panel entrés en 1986 et présents dans l'échantillon en 1987.
13.0	**M. 85/E/87**	soit les membres du panel entrés en 1985, émigrés en 1986 mais revenus dans l'échantillon en 1987.

Il n'existe qu'un autre vecteur propre à des membres du panel (M) dont le point final soit /87: le vecteur #14 qui désigne les nouveaux membres entrés en 1987. La procédure de pondération exige qu'ils puissent être isolés facilement. Ils forment le groupe #2 de la variable DIA87.

Le regroupement des vecteurs #10 à #14 de la variable DIA3 équivaut au regroupement des classes #1 et #2 de la variable DIA87: il désigne l'ensemble des membres du panel présents dans l'échantillon en 1987 soit 4648 personnes.

5.2. Les variables de pondération

5.2.1. Variable instrumentale TREE87

Les refus de répondre constituent le facteur principal de la déformation de l'échantillon (2.2.1.1.).

1. La notion de "taux de réponses" prendra désormais un sens particulier. Elle correspond davantage à la notion de "taux de survie dans le panel" (5.2.1.1).

2. Les refus de répondre seraient sans effet s'ils étaient répartis aléatoirement dans l'échantillon. Il convient donc de rechercher les concentrations éventuelles des refus dans des groupes démographiques homogènes afin d'en corriger les effets (5.2.1.3.).

3. Au préalable, l'inégalité des probabilités de sélection initiales des membres du panel doit être annulée en pondérant les individus par l'inverse de cette probabilité. La recherche des groupes démographiques atteints par les taux de refus les plus contrastés ne doit pas être influencée par ce biais lié au mode de tirage de l'échantillon initial (5.2.1.2.).

5.2.1.1. *Des taux de survie*

Il est important de noter que la notion de taux de réponses sera entendue désormais dans un sens particulier.

Le terme "taux de survie" est plus correct que le terme "taux de réponses". En effet, le **taux de survie** prend en compte le caractère cumulatif des refus enregistrés en 1987 et en 1986, à condition que ces derniers n'aient pas rejoint l'échantillon en 1987.

Le taux de réponses correspond désormais au nombre des réponses enregistrées, rapporté au nombre cumulé des membres du panel.

> *Il n'est pas calculé par référence au nombre des membres du panel présents dans l'échantillon précédent. Il est calculé en rapportant le nombre de réponses des membres du panel au nombre total des membres du panel entrés dans les échantillons successifs depuis la première vague d'enquête.*

Cette procédure permet de prendre en compte la mobilité des membres du panel: ils peuvent entrer et sortir de l'échantillon. Chaque fois qu'ils rejoignent l'échantillon, leur

nouvelle probabilité de sélection peut être calculée puisque les deux éléments essentiels du calcul sont connus:

* leur probabilité de sélection initiale (**5.2.1.2.**)

* et les caractéristiques démographiques du sous-groupe auquel ils appartenaient au moment ou ils sont entrés dans l'échantillon (**5.2.1.3.**).

5.2.1.2. *Correction du biais initial*

L'analyse des taux de réponses prend en compte les 6015 personnes regroupées dans les classes #0 et #1 de la variable DIA87. Elle ne prend en compte, ni les nouveaux membres du panel entrés dans l'échantillon au cours de la dernière vague d'enquête, ni les non-membres du panel, ni les personnes émigrées ou décédées.

Parmi les 6015 membres du panel, entrés dans l'échantillon en 1985 ou en 1986, 4583 ont répondu en 1987, soit 76.2 %.

Ces valeurs ne prennent pas en compte l'inégalité des probabilités de sélection initiales des individus. Il convient donc d'annuler cette inégalité avant d'entreprendre toute autre opération.

Les membres entrés dans l'échantillon en 1985 doivent être pondérés par l'inverse de leur probabilité de sélection soit, par la variable PONREV.

Les nouveaux membres entrés en 1986 doivent recevoir le poids correspondant au poids moyen de leurs ascendants directs en 1986: la probabilité de sélection de ces nouveaux membres correspond à la valeur moyenne des probabilités de sélection de leurs parents *au moment ou ces nouveaux membres entrent dans le panel.* Ces nouveaux membres ont déjà reçu ce poids au moment ou ils sont entrés dans l'échantillon: c'est la valeur de la variable W86B qui leur a été attribuée en 1986.

La synthèse de ces valeurs donne lieu à la création de la variable **IP8587.** Cette variable est sauvegardée dans le fichier global parce qu'elle sera réutilisée l'année prochaine dans le cadre de la même procédure. (Cette variable doit être ajustée en fonction de la taille de l'échantillon).

Cette première procédure modifie la valeur du taux de réponses. Le tableau **5.3.** permet d'observer les glissements des proportions avant et après la correction des biais de sélection.

TABLEAU 5.3.
Distribution des fréquences des individus membres du panel, en 1987, selon leur position dans l'échantillon, avant et après pondération

DIA87	FREQUENCES			
	avant pondération		après pondération (IP8587)	
	n	%	n	%
0. memb.pan. refus	1432	23.8	1364	22.7
1. memb.pan. réponses	4583	76.2	4652	77.3
TOTAL	6015	100.0	6015	100.0

Le taux de réponses passe de 76.2 % à 77.3 % lorsque les biais de sélection initiaux sont contrôlés.

5.2.1.3. *Les nouvelles probabilités de sélection*

La qualité de l'échantillon n'est pas affectée par les refus de répondre aussi longtemps qu'ils se distribuent de manière aléatoire dans l'ensemble de l'échantillon.

La méthode permettant de repérer des concentrations éventuelles de ces refus dans certains sous-groupes homogènes de l'échantillon a été présentée dans le second chapitre (**2.2.1.1.**).

La démarche met en concurrence le pouvoir discriminant de 17 variables:

CARACTERISTIQUES DU MENAGE	DU CHEF DE MENAGE	DE LA PERSONNE
Canton de résidence	Age	Age
Type d'habitat	Sexe	Sexe
Taille du ménage	Etat civil	Etat civil
Nombre d'adultes	Emploi (oui/non)	Formation
Nombre d'enfants		Emploi
Nombre d'emplois		Adulte/Enfant
		Lien avec le CM

Toutes ces caractéristiques sont attribuées aux personnes. Les caractéristiques du ménage et les caractéristiques du chef de ménage sont des éléments de l'environnement de la personne.

Les membres entrés dans l'échantillon en 1985 sont caractérisés par les valeurs qu'ils présentaient sur ces variables **en 1985**, au moment où ils sont entrés dans l'échantillon.

Les nouveaux membres, entrés en 1986, sont caractérisés par les valeurs qu'ils présentaient sur ces différentes variables **en 1986**. Ces variables décrivent donc le contexte du ménage dans lequel ils sont nés et *au moment où ils sont nés*.

Il est donc fort probable qu'il existe un décalage systématique entre les caractéristiques des nouveaux membres et les caractéristiques de leurs parents.

Si la taille du ménage dans lequel ils font leur entrée s'est modifiée entre 1985 et 1986, ces nouveaux membres du panel entrés en 1986 et leurs parents entrés en 1985 ne sont pas caractérisés par la même valeur sur cette variable.

L'âge du chef de ménage fournit l'exemple le plus clair: en règle générale, l'âge du chef de ménage sera différent pour les parents et pour le nouveau membre du panel entré dans la famille un an plus tard.

La synthèse de ces valeurs donne lieu à la création de nouvelles variables. Ces variables sont sauvegardées dans le fichier global parce qu'elle seront réutilisées l'année suivante dans le cadre de la même procédure.

(La variable de correction des probabilités de sélection initiales a été actualisée selon la même logique (5.2.1.2.)).

Lorsque l'ensemble des variables choisies ont été actualisées, la recherche des catégories démographiques les plus discriminantes peut être entreprise (2.2.1.1.).

En résumé, la procédure consiste à diviser l'échantillon en sous-groupes mutuellement exclusifs par une succession de divisions binaires. L'enchaînement de ces divisions successives crée des filières ou séries de modalités. Les variables qui produisent ces divisions sont choisies parce qu'elles créent les contrastes les plus significatifs entre les taux de réponses des deux sous-groupes. Ces divisions sont interrompues dès qu'un sous-groupe ne peut plus être divisé en deux sous-groupes d'une taille minimale de 200 personnes. (Cf. Procedures and Tape Codes, 1978 Interviewing Year, Wave XI, a Supplement. P.S.I.D., I.S.R., Ann Arbor, Michigan, 1979, p.10).

Le résultat de cette analyse est présenté dans le schéma 5.1.. Il reproduit sous une forme arborescente la progression de l'analyse vers des sous-groupes de plus en plus homogènes.

Dix-neuf sous-groupes sont identifiés à l'aide de 9 variables.

La variable dépendante est une variable binaire (0,1) qui rend compte du **taux de survie** dans le panel longitudinal. Les 19 sous-groupes "n'expliquent" que 6.08 % de la variance de la variable dépendante.

L'objectif de cette procédure n'était pas de rendre compte de toute la variance des taux de survie. Des mesures ont été prises en vue d'éviter que l'analyse identifie des groupes marginaux de très petite taille. La procédure évite que la variance des poids des individus soit démesurément accrue, en maintenant la taille des sous-groupes à un minimum de 200 personnes.

Le faible pouvoir explicatif de la variable constituée par cette procédure provient également du fait que les variables démographiques de base divisent l'échantillon en modalités peu contrastées (tableau 5.4.).

COMPARAISON DES TAUX DE REPONSES DES SOUS-GROUPES ENTRE 1985 ET 1987

SCHEMA 5.1.

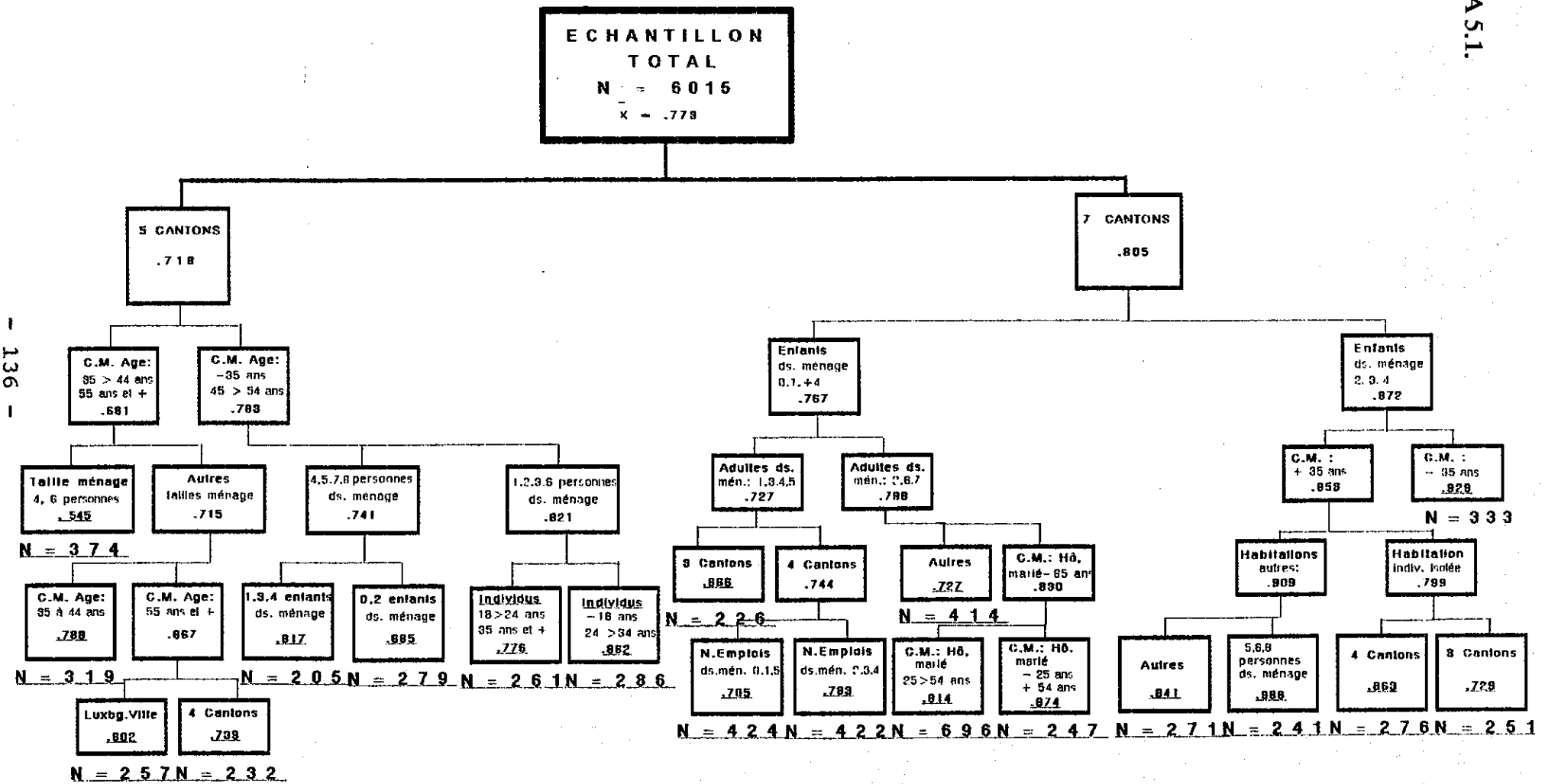


TABLEAU 5.4.

Taux de survie des membres du panel en 1987

CARACTÉRISTIQUES DES SOUS-GROUPES	TAUX DE REPONSE $\bar{x}$ = .77	VARIANCE "EXPLIQUÉE" (ETA)	VARIANCE EXPLIQUÉE PAR LA MEILLEURE DICHOTOMISATION
Caractéristiques individuelles			
Sexe			
1. masculin	.77		
2. féminin	.77	.00	.00
Etat civil			
1. célibataire	.78		
2. marié	.77		
3. séparé	.81		
4. divorcé	.70		
5. veuf	.70		
6. D.M.	(.00)*	.06	.05
Age			
1. < 17 ans	.81		
2. 18 à 24 ans	.74		
3. 25 à 34 ans	.79		
4. 35 à 44 ans	.76		
5. 45 à 54 ans	.77		
6. 55 à 64 ans	.75		
7. 65 ans et plus	.69	.08	.07
Nationalité			
1. Luxembourgeois	.76		
2. CEE	.79		
3. hors CEE	.73		
4. D.M.	(.33)	.04	.02

*Les chiffres entre parenthèses correspondent à des catégories de moins de 20 personnes.

Situation professionnelle

1. Enfants	.80		
2. Travail ou maladie	.77		
3. Demandeur d'emploi	.61		
4. Pensionné, invalide	.74		
5. Retraité	.71		
6. Ménage, milice	.76	.07	.05

Lien avec le chef de ménage

1. Chef de ménage	.76		
2. épouse	.78		
3. "épouse"	.87		
4. ami(e)	.70		
5. fils, fille	.78		
6. autre	.68	.06	.05

Adultes

1. Enfant	.80		
2. Adulte	.76	.05	.05

Etat civil selon le sexe

1. Homme			
célibataire	.77		
2. marié	.77		
3. séparé	.78		
4. divorcé	.73		
5. veuf	.73		
6. M.D.	(.00)		
7. Femme célibataire	.79		
8. mariée	.78		
9. séparée	.85		
10. divorcée	.67		
11. veuve	.69	.07	.06

Emploi selon le sexe

1. Garçon mineur	.79		
2. Homme ayant un emploi	.77		
3. Homme sans emploi	.72		
4. Fille mineure	.81		
5. Femme ayant un emploi	.76		
6. Femme sans emploi	.75	.05	.03

Age selon le sexe

1.	Homme			
	< de 18 ans	.81		
2.	de 18 à 64 ans	.76		
3.	> 64 ans	.71		
4.	Femme			
	< de 18 ans	.81		
5.	18 à 64 ans	.77		
6.	> 64 ans	.68	.07	.05

Niveau de formation terminé

1.	Enfant	.95		
2.	Pas de formation	.79		
3.	primaire ou complémentaire	.76		
4.	secondaire	.76		
5.	supérieur non-universitaire	.71		
6.	supérieur universitaire	.78		
7.	Ne sait pas	.72	.07	.05

Caractéristiques du ménage

Nombre de personnes dans le ménage

1.	.71		
2.	.77		
3.	.77		
4.	.75		
5.	.84		
6.	.79		
7.	.73		
8.	.86		
9.	.89		
11.	.90		
12.	1.00	.08	.07

Nombre d'enfants dans le ménage

0.	.73		
1.	.79		
2.	.76		
3.	.85		
4.	.83		
5.	(1.00)		
6.	(.77)	.09	.07

Nombre d'adultes dans le ménage

0.	(.00)		
1.	.72		
2.	.79		
3.	.76		
4.	.69		
5.	.73		
6.	.73		
7.	(.85)		
9.	(1.00)	.08	.07

Nombre d'emplois dans le ménage

0.	.73		
1.	.78		
2.	.78		
3.	.71		
4.	.78		
5.	.69		
6.	(1.00)	.06	.06

Type d'habitation

1. maison rurale	.76		
2. maison unifamiliale individuelle	.76		
3. maison jumelée	.76		
4. maison en série	.78		
5. petit collectif	.79		
6. collectif moyen	.75		
7. grand collectif	.83		
8. précaire	.71		
9. Non réponse	.84	.04	.02

Canton de résidence

1. Luxembourg-Ville	.72		
2. Luxembourg-Campagne	.77		
3. Capellen	.68		
4. Esch-sur-Alzette	.81		
5. Mersch	.77		
6. Clervaux	.91		
7. Diekirch	.74		
8. Rédange	.68		
9. Vianden	.91		
10 Wiltz	.75		
11 Echternach	.80		
12 Grevenmacher	.80		
13 Remich	.77	.12	.10

Caractéristiques du chef de ménage

Age du chef de ménage			
1. < 24 ans	.81		
2. 25 à 34 ans	.84		
3. 35 à 44 ans	.77		
4. 45 à 54 ans	.77		
5. 55 à 64 ans	.73		
6. 65 et plus	.69	.10	.08
Sexe du chef de ménage			
1. masculin	.78		
2. féminin	.71	.05	.05
Etat civil du chef de ménage			
1. Célibataire	.77		
2. marié	.78		
3. séparé	.84		
4. divorcé	.67		
5. veuf	.71		
6. Non réponse	(.00)	.08	.06]
Situation professionnelle du C.M.			
1. sans emploi	.72		
2. travaille	.78	.07	.07
Etat civil selon le sexe du C.M.			
1. Hommenon marié	.76		
2. marié	.78		
3. Femme non mariée	.72		
4. mariée	.47	.07	.05
Situation professionnelle du C.M. selon le sexe			
1. Hommesans emploi	.72		
2. au travail	.79		
3. Femme sans emploi	.71		
4. au travail	.71	.08	.08
Age du C.M. selon le sexe			
1. Homme < 35 ans	.83		
2. 35 à 44 ans	.78		
3. 45 ans à 64 ans	.76		
4. > 64 ans	.69		
5. Femme < 35 ans	.85		
6. 35 à 44 ans	.62		
7. 45 à 64 ans	.70		
8. > 64 ans	.69	.12	.09

Trois points peuvent être mis en évidence.

1. Le contraste très prononcé opposant la métropole et les autres cantons s'estompe après cette troisième vague d'enquête. Ceci peut être expliqué: les habitants de la capitale qui ont accepté de participer à l'étude à deux reprises restent relativement fidèles et les refus ne se concentrent plus particulièrement dans cette zone. Peu à peu, les refus se répartissent dans les cantons selon d'autres critères.

Quelques traces de ce clivage subsistent encore, mais les caractéristiques des habitants de Luxembourg-ville qui refusent de répondre se modifient progressivement.

2. Un fait nouveau doit être mentionné. Pour la première fois une caractéristique propre aux personnes émerge parmi les variables de clivage. Les variables décrivant le cadre familial ne monopolisent plus l'organisation des taux de réponses.

La description complète de cette catégorie de personnes n'apporte pas une information très claire:

"Dans certains cantons (5 sur 12)

"les personnes âgées de 18 à 24 ans et les personnes âgées de 35 ans et plus répondent moins souvent (.776) que

"les personnes âgées de moins de 18 ans et les personnes âgées de 24 à 34 ans (.862)

"lorsque toutes ces personnes appartiennent à des ménages de 1 à 3 personnes dont le chef de ménage est âgé de moins de 35 ans ou de 45 à 54 ans".

Il est malaisé de saisir le contenu exact de ce sous-groupe mais il est remarquable que ce facteur strictement individuel (l'âge des personnes) intervienne dans la différenciation des taux de réponses.

L'apparition d'un tel facteur tend à indiquer que les personnes acquièrent progressivement une certaine autonomie par rapport au contexte du ménage.

La conception de la procédure de calcul des poids individuels et des poids des ménages permet précisément de prendre en compte cette autonomie des trajectoires individuelles et des trajectoires des ménages.

Elle fonde le système de pondération dans les destinées individuelles et permet de respecter un fait relativement évident: la structure des ménages se modifie d'année en année, le ménage est une cristallisation instable et provisoire des histoires individuelles (**4.2.7., ANNEXE 4.B**).

Si les variables individuelles devaient prendre, par la suite, de plus en plus de poids dans le système de pondération, il s'avèrerait que les panels socio-économiques devraient éviter de procéder à des pondérations directes des ménages sans prendre en compte l'hétérogénéité interne de cette unité d'analyse.

3. Il faut encore souligner la croissance notable des écarts entre les taux de survie:

	1986	1987
Taux le plus élevé	1.00	.98
Taux le plus bas	.71	.54

Cette variable TREE87 décrit le profil des catégories d'appartenance des membres du panel, qu'ils aient répondu ou non en 1987. Elle permet donc de classer tous les membres des classes #0 et #1 de la variable DIA87 (tableau 5.5.).

TABLEAU 5.5.
Ventilation des valeurs de TREE87 dans le fichier individuel

DIA87	TREE87 n =
0. memb. panel refus	1432
1. memb. panel réponse	4583
2. nouv. memb. panel 87	-
3. non-memb. panel réponse	-
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	6015
sysmis	527
TOTAL	6542

TREE87 n'exprime pas des probabilités de survie. Elle décrit simplement une typologie de sous-échantillons. Elle ne permet donc pas de calculer le poids relatif des personnes dans le nouvel échantillon. Ce sera le rôle d'une autre variable: RR87.

5.2.2. Variable instrumentale RR87

La variable RR87 est construite de la manière suivante: toute personne ayant effectivement répondu en 1987 (DIA87 #1), reçoit la probabilité de sélection observée dans la catégorie de population à laquelle elle appartient (TREE87) (tableau 5.6.).

TABLEAU 5.6.
Distribution des fréquences de l'échantillon 1987, selon les taux de réponses observés en 1987 (avant et après pondération de l'échantillon initial)

RR87	TREE87	FREQUENCES	
taux de réponses en 1987	code	avant pondération n =	après pondération (IP8587) n =
.54	(1)	229	204
.60	(3)	129	155
.67	(9)	205	150
.68	(6)	173	191
.70	(10)	305	299
.73	(12)	296	301
.73	(18)	171	183
.74	(4)	206	172
.78	(7)	199	203
.78	(11)	588	331
.79	(2)	211	251
.81	(13)	428	567
.82	(5)	200	167
.84	(15)	183	228
.86	(8)	189	247
.86	(17)	249	239
.87	(14)	171	216
.93	(19)	240	309
.98	(16)	211	238
TOTAL		4583	4652

Il n'y a pas lieu de s'étonner si la taille de l'échantillon n'est plus identique avant et après le redressement de l'échantillon. Seule une fraction de l'échantillon des membres du panel a répondu en 1987: 4583 personnes sur 6015.

Ces 4583 personnes ont été pondérées dans le cadre de l'échantillon total afin de compenser les variations des probabilités de sélection initiales. Ces poids ont été ajustés: la

somme des poids est égale au nombre des individus, soit 6015. Ils n'ont pas été ajustés dans le cadre des 4583 personnes ayant participé à l'enquête (tableau **5.3.**).

La différence entre les tailles de la fraction de l'échantillon avant et après pondération est donc parfaitement normale. En outre, la comparaison des deux dernières colonnes du tableau **5.6.** permet de constater le fait que certaines catégories de la population auraient été sur-représentées ou sous-représentées si l'échantillon n'avait pas été pondéré avant d'être soumis à l'analyse des taux de réponses.

TABLEAU 5.7.
Ventilation des valeurs de RR87 dans le fichier individuel

DIA87	RR87 n =
0. memb. panel refus	-
1. memb. panel réponse	4583
2. nouv. memb. panel 87	-
3. non-memb. panel réponse	-
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4583
sysmis	1959
TOTAL	6542

Les valeurs de RR87 ne sont donc attribuées qu'aux seuls membres du panel ayant répondu en 1987 (tableau **5.7.**)

5.2.3. Variable instrumentale W87

La variable W87 définit le poids des membres du panel ayant répondu en 1987, **exception faite des nouveaux membres du panel entrés dans l'échantillon en 1987.**

Ces anciens membres appartiennent à différentes catégories de la population et se distinguent par la probabilité qu'ils répondent en 1987.

Lorsque le groupe enregistre un taux de réponses très faible, chaque "survivant" doit représenter un plus grand nombre de personnes présentant les mêmes caractéristiques démographiques. Lorsque le taux de réponses s'élève, l'importance de cet ajustement se réduit.

D'une manière générale, cet ajustement des poids des individus sera inversement proportionnel à leur probabilité de sélection c'est à dire, inversement proportionnel au taux de réponses enregistré dans leur groupe:

Soit: 1 / RR87

Dans le cas présent, ce calcul doit prendre en compte un élément supplémentaire: l'inégalité des probabilités de sélection initiales des membres du panel. Les nouveaux poids ne s'appliquent pas à des personnes "unités" mais à des fractions ou à des multiples de l'unité définis par la variable IP8587.

La nouvelle variable de pondération se calcule donc de la manière suivante:

W87 = IP8587 x (1 / RR87)

TABLEAU 5.8.
Ventilation des valeurs de W87 dans le fichier individuel

DIA87	W87 n =
0. memb. panel refus	-
1. memb. panel réponse	4583
2. nouv. memb. panel 87	-
3. non-memb. panel réponse	-
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4583
sysmis	1959
TOTAL	6542

Les valeurs de cette variable sont attribuées exclusivement aux MEMBRES DU PANEL présents dans l'échantillon en 1987, exception faite des nouveaux membres entrés dans l'échantillon en 1987 (tableau 5.8.). Le système de pondération doit donc être élargi à ces nouveaux membres.

5.2.4. Variable instrumentale W87B

La variable W87 ne permet pas encore de pondérer tout l'échantillon observé en 1987, parce qu'elle ne prend pas en compte les nouveaux membres de l'échantillon (membres ou non du panel). Il est indispensable d'étendre la portée de cette variable pour pouvoir calculer les poids des ménages.

Dans un premier temps, **les nouveaux membres de l'échantillon, non membres du panel**, reçoivent un poids nul ('0') puisqu'il est impossible de calculer leur probabilité de sélection initiale. En effet:

- ils étaient déjà présents dans la population de 1985, lorsque l'échantillon initial a été sélectionné,

- ils sont entrés dans l'échantillon par la suite et pour des raisons variées: il est donc difficile de reconstituer l'ensemble des conditions de leur entrée dans l'échantillon et d'en dériver une logique de pondération[1].

Il est indispensable de pondérer ces non-membres avant d'attribuer des poids aux **nouveaux membres de l'échantillon, membres du panel.** En effet, certains nouveaux membres de l'échantillon, **non-membres du panel** peuvent être des ascendants directs d'**enfants nouveaux membres du panel.**

Or, dans un second temps, ces nouveaux membres du panel reçoivent un **poids égal à la moyenne des poids des ascendants directs** tels qu'ils ont été calculés au moment où le nouveau membre est entré dans l'échantillon (et dans la population).

Donc:

- lorsque les ascendants directs ont des poids différents de '0', le nouveau membre reçoit un poids égal à la moyenne de ces deux poids,

- lorsque l'un des ascendants directs a un poids nul, le nouveau membre du panel reçoit logiquement un poids égal à la moitié du poids de l'ascendant direct, membre du panel,

- lorsque les deux ascendants directs ne sont pas membres du panel, le nouveau membre reçoit un poids nul: il n'est pas membre du panel.

Par contre, un nouveau membre reçoit le poids de l'ascendant direct présent dans l'échantillon observé, lorsque l'autre ascendant est ABSENT de l'échantillon au moment où le nouveau entre dans l'échantillon: la probabilité de sélection initiale de ce nouveau membre dépend entièrement de la probabilité de sélection du seul ascendant présent dans l'échantillon. (Cfr. **2.2.1.5.** et **2.2.1.6.**).

---

1. Le poids relatif de ces individus est certainement différent de '0' mais toute affectation d'un poids différent de '0' reposerait sur un ensemble d'hypothèses comportant des erreurs qu'il est impossible d'estimer.

TABLEAU 5.9.
Ventilation des valeurs de W87B
dans le fichier individuel

DIA87	W87B n =
0. memb. panel refus	-
1. memb. panel réponse	4583
2. nouv. memb. panel 87	65
3. non-memb. panel réponse	266
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4914
sysmis	1628
TOTAL	6542

Ces poids sont attribués à tous les membres de l'échantillon observé en 1987, soit 4914 personnes. Parmi celles-ci, 266 personnes ont un poids égal à 0 (groupe #3 de la variable DIA87) (tableau **5.9.**).

La somme des poids est égale à 4385.32, quel que soit le nombre de personnes prises en compte: 4914 personnes interrogées ou 4648 membres du panel.

La taille "réelle" de l'échantillon est égale à 4648 personnes membres du panel. Les 266 personnes présentes dans cet échantillon sans être membres du panel ont reçu un poids nul, précisément parce qu'il est impossible d'estimer le poids de leur présence dans cet échantillon individuel.

La somme des poids doit être ajustée à la taille de l'échantillon. La somme des poids doit être égale au nombre de personnes observées afin de sauvegarder les propriétés statistiques de l'échantillon. C'est le rôle de la variable WGT87TOT.

5.2.5. Variable opérationnelle WGT87TOT

La taille de l'échantillon pondéré par la variable W87B serait réduite à 4385 unités. Les poids doivent donc être ajustés par une constante: la moyenne des poids attribués aux individus (**2.2.1.4**).

Suite à cette opération, la valeur moyenne des poids est égale à 1.0 et la somme des poids est égale au nombre de personnes observées.

Cet ajustement est effectué sur 4648 personnes membres du panel. Il s'en suit que les analyses des données peuvent être réalisées de manière équivalente sur l'échantillon de 4914 personnes observées en 1987 ou sur l'échantillon de 4648 membres du panel observés en 1987. Dans les deux cas, la taille réelle de l'échantillon sera respectée.

Dans le premier cas le nombre total de personnes prises en compte sera apparemment égal à 4914, mais la somme des poids sera égale à 4648 unités. Seuls, les membres du panel sont pris en compte au cours des analyses, puisque 266 personnes ont un poids égal à 0: les valeurs de ces personnes sur les variables analysées sont annulées.

Dans le second cas, seuls les membres du panel sont pris en compte dans l'analyse et la mise à l'écart des non-membres du panel n'affecte en rien les résultats (**2.2.1.7.**).

La variable W87B devient:

WGT87TOT = W87B / (4385.322 / 4648)

Voir en ANNEXE 4A - Comparaison des caractéristiques des membres du panel en 1985, 1986 et 1987 après pondération

TABLEAU 5.10.
Ventilation des valeurs de WGT87TOT dans le fichier individuel

DIA87	WGT87TOT n =
0. memb. panel refus	-
1. memb. panel réponse	4583
2. nouv. memb. panel 87	65
3. non-memb. panel réponse	266
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4914
sysmis	1628
TOTAL	6542

Les valeurs de la variable WGT87TOT sont attribuées à l'ensemble des personnes ayant participé à l'enquête de 1987, soit les groupes #1, #2 et #3 de la variable DIA87 (tableau **5.10.**).

5.2.6. Variable opérationnelle de pondération des ménages en 1987 MWGT87A

La variable MWGT87A permet de pondérer correctement l'échantillon des ménages observé en 1987.

En 1985, le biais d'échantillonnage apparaît au niveau des ménages. La correction de ce biais s'effectue, en premier lieu, au niveau des ménages. Elle est répercutée sur les membres du ménage. Ceux-ci reçoivent le même poids que le ménage auquel ils appartiennent.

Il existe donc un rapport constant entre le poids du ménage et le poids des individus: le poids du ménage est égal à la valeur moyenne des poids des personnes qui le composent.

Depuis 1986, l'échantillon subit l'influence de facteurs d'évolution qui introduisent des biais aux deux niveaux d'analyse.

Tantôt un ménage complet disparaît et la raison du refus est commune à l'ensemble des membres du ménage.

Tantôt, un seul membre disparaît pour des raisons strictement individuelles.

Des nouveaux membres entrent dans les ménages. Les uns sont pris en compte au titre de membres du panel, les autres ne sont pas reconnus comme membres du panel.

De nouveaux ménages se créent et, généralement, provoquent l'inclusion d'une nouvelle personne étrangère à l'échantillon initial. Des ménages se défont et provoquent tantôt l'apparition de deux nouveaux ménages, tantôt la création d'un nouveau ménage et le retour d'une personne dans son milieu d'origine.

Cette liste n'est pas exhaustive. Elle illustre suffisamment à quel point les trajectoires des personnes et les trajectoires des ménages se différencient progressivement. Elle invite à fonder l'analyse de l'évolution de l'échantillon au niveau des trajectoires individuelles plutôt qu'au niveau des ménages.

Cependant, le rapport entre la valeur des poids relatifs individuels et la valeur des poids des ménages garde sa propriété initiale: le poids du ménage est égal à la valeur moyenne des poids des individus.

La procédure est simple:

* les valeurs des poids individuels sont agrégées au niveau du ménage;

* la valeur moyenne des poids individuels (WGT87TOT) détermine la valeur du poids du ménage (MWGT87T);

* cette valeur peut être ajustée en fonction de la taille de l'échantillon des ménages.

La taille "réelle" de l'échantillon des ménages correspond à **1641 ménages composés d'au moins un membre du panel.**

En 1987, 3 ménages sont entièrement composés par des personnes qui ne sont pas membres du panel: ils reçoivent logiquement un poids de '0'.

L'ajustement des poids des ménages s'effectue donc sur un nombre de ménages égal à 1641 unités et non sur 1644 ménages. De cette façon, les analyses seront toujours effectuées sur 1641 unités, indépendamment du fait que les ménages composés exclusivement de non-membres du panel sont pris en compte ou sont mis à l'écart (**2.2.1.7.**). A défaut d'un tel ajustement, l'échantillon pondéré par MWGT87T compterait 1692 unités. D'où:

MWGT87A = MWGT87T / (1692.372 / 1641)

TABLEAU 5.11.
Ventilation des valeurs de MWGT87A dans le fichier individuel

DIA87	MWGT87A n =
0. memb. panel refus	-
1. memb. panel réponse	4583
2. nouv. memb. panel 87	65
3. non-memb. panel réponse	266
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4914
sysmis	1628
TOTAL	6542

Les valeurs des poids des ménages (MWGT87T et MWGT87A) sont reventilées sur les individus. Chaque membre du ménage reçoit la valeur du poids du ménage auquel il appartient. Les non-membres du panel reçoivent, cette fois, un poids différent de 0, à condition qu'ils n'appartiennent pas à un ménage composé exclusivement de non-membres du panel.

Le tableau **5.11.** présente les sous-échantillons individuels qu'il convient de sélectionner dans le fichier au moment de reconstituer l'échantillon des ménages observés en 1987. Les poids des ménages s'obtiennent aisément, par exemple, en agrégeant les valeurs de MWGT87A au niveau des ménages et en prenant la moyenne comme valeur de pondération.

(Certains logiciels permettent de prendre la valeur du premier membre du ménage comme valeur de référence (SPSSX)).

Il peut être utile de rappeler que TOUS les membres appartenant à l'échantillon observé en 1987 doivent être pris en compte au moment de reconstituer l'échantillon des ménages. En l'absence des non-membres du panel, il serait impossible de calculer certaines variables propres au ménage (par exemple, la taille du ménage).

ANNEXE 5 A

TABLEAU COMPARATIF DES CARACTERISTIQUES DEMOGRAPHIQUES DES MEMBRES DU PANEL EN 1985, EN 1986 ET EN 1987 (*après pondération*)

- **caractéristiques personnelles**
- **caractéristiques du ménage**
- **caractéristiques du chef de ménage**

Caractéristiques Personnelles	**1985** (N = 6110)	**1986** (N = 5245)	**1987** (N = 4648)
Sexe			
1. homme	48.8	48.7	48.2
2. femme	51.2	51.3	51.8
Age			
1. < 18 ans	24.1	23.5	23.6
2. 18 - 24 ans	11.2	11.2	11.1
3. 25 - 34 ans	15.6	15.1	15.2
4. 35 - 44 ans	14.3	14.3	14.6
5. 45 - 54 ans	12.7	12.8	12.6
6. 55 - 64 ans	10.4	10.9	10.5
7. > 64 ans	11.7	12.1	12.4
Age par Sexe			
1. homme < 18 ans	12.7	12.4	12.0
2. homme 18 - 64 ans	31.7	31.8	31.5
3. homme > 64 ans	4.4	4.6	4.7
4. femme < 18 ans	11.4	11.2	11.6
5. femme 18 - 64 ans	32.6	32.6	32.5
6. femme > 64 ans	7.2	7.5	7.7
Etat civil			
1. célibataire	39.8	39.9	40.0
2. marié	49.6	49.2	48.8
3. séparé	1.0	0.8	0.8
4. divorcé	1.9	1.9	2.3
5. veuf	7.9	8.0	8.1
6. N.R.	0.0	-	-

Nationalité

1. Luxembourgeois	76.1	76.4	77.9
2. autres C.E.	21.3	21.3	20.3
3. hors C.E.	2.5	2.3	1.9
4. N.R.	0.1	-	-

Emploi

1. oui	38.1	39.8	37.1
2. non	35.4	35.3	34.2
3. enfants	26.5	24.9	28.7

Adulte/enfant

1. adulte	73.5	75.1	74.7
2. enfant	26.5	24.9	25.3

Lien avec le chef de ménage

1. chef de ménage	36.4	37.2	36.6
2. épouse	24.2	24.1	23.9
3. "épouse"	1.1	1.0	0.9
4. ami, amie	0.3	0.1	0.2
5. fils, fille	33.4	33.6	34.2
6. autres	4.6	4.0	4.2

Caractéristiques du ménage

Type d'habitation

1. maison rurale	11.4	10.7	9.3
2. unifamiliale individuelle	33.7	34.7	34.7
3. unifamiliale jumelée	13.5	8.8	9.1
4. unifamiliale en série	19.5	21.4	25.3
5. petit collectif	12.6	13.1	13.1
6. collectif moyen	7.1	7.3	5.8
7. grand collectif	1.1	1.4	1.1
8. précaire	0.3	0.0	0.1
9. N.R.	0.8	2.6	1.7

Canton

1. Luxembourg-ville	17.6	17.6	17.7
2. Luxembourg-campagne	8.9	8.8	8.8
3. Capellen	6.9	7.0	6.9
4. Esch sur Alzette	33.5	33.6	34.3
5. Mersch	4.6	4.3	3.7
6. Clervaux	2.3	2.3	2.5
7. Diekirch	6.2	6.2	6.7
8. Redange	3.0	3.0	2.8
9. Vianden	0.8	0.7	0.8
10. Wiltz	3.0	3.1	3.0
11. Echternach	3.5	3.5	3.5
12. Grevenmacher	5.8	5.8	5.7
13. Remich	4.0	4.0	3.6

Nombre de personnes dans le ménage

1.	7.6	8.0	8.0
2.	21.0	20.6	19.6
3.	22.5	23.1	22.8
4.	27.6	27.6	28.7
5.	12.3	12.7	12.9
6.	5.8	5.3	4.9
7.	2.1	1.8	2.1
8.	0.7	0.6	0.7
9.	0.3	0.3	0.2
11.	0.0	-	0.1
12.	-	0.1	-

Nombre d'enfants dans le ménage

0.	39.2	41.9	41.7
1.	24.3	23.2	23.6
2.	22.9	23.2	22.8
3.	9.9	9.5	9.1
4.	3.5	2.2	2.7
5.	0.1	0.1	0.1
6.	0.1	-	-

Nombre d'adultes dans le ménage

0.	0.1	-	-
1.	9.9	10.2	10.4
2.	63.0	59.5	58.4
3.	16.2	18.9	18.9
4.	7.7	6.9	7.8
5.	2.1	3.0	2.8
6.	0.8	1.2	1.4
7.	0.2	0.2	0.2
9.	-	0.1	0.1

Nombre d'emplois dans le ménage			
0.	17.1	17.5	17.4
1.	46.9	43.1	42.1
2.	29.2	29.8	30.0
3.	4.9	6.8	6.9
4.	1.4	2.1	2.5
5.	0.5	0.7	0.9
6.	0.0	0.0	0.1

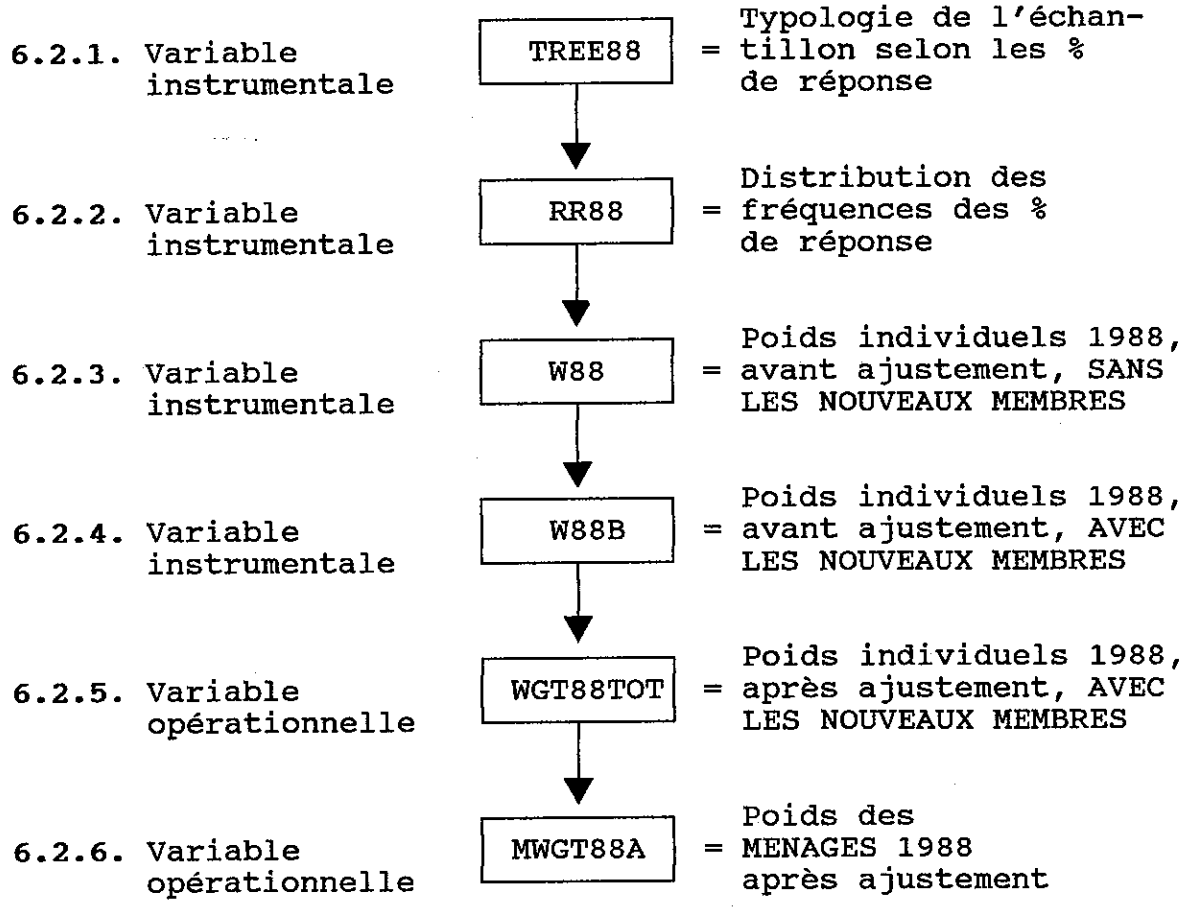
Caractéristiques du chef de ménage

Sexe			
1. homme	87.2	86.0	86.3
2. femme	12.8	14.0	13.7
Age			
1. < 24 ans	2.6	2.3	1.8
2. 25 - 34 ans	18.3	18.1	18.7
3. 35 - 44 ans	27.6	27.1	27.3
4. 45 - 54 ans	24.4	23.7	23.0
5. 55 - 64 ans	14.8	15.9	15.8
6. > 64 ans	12.2	12.9	13.4
Etat civil			
1. célibataire	6.0	6.1	6.5
2. marié	80.6	80.0	79.6
3. séparé	1.4	1.7	1.2
4. divorcé	2.5	2.7	2.7
5. veuf	9.5	9.6	10.0
6. N.R.	0.0	-	-
Emploi			
1. oui	74.1	73.3	73.3
2. non	25.9	26.7	26.7
Etat civil par Sexe			
1. homme non-marié	7.0	6.7	7.0
2. homme marié	80.2	79.3	79.3
3. femme non-mariée	12.4	13.4	13.4
4. femme mariée	0.4	0.6	0.3
Emploi par Sexe			
1. homme sans emploi	17.5	17.4	17.6
2. homme avec emploi	69.8	68.6	68.7
3. femme sans emploi	8.5	9.3	9.1
4. femme avec emploi	4.3	4.7	4.6

ÉCHANTILLON 1988

ET VARIABLES DE PONDÉRATION

Encadré 6.1. **VARIABLES DE PONDERATION 1988**



Chapitre 6

ÉCHANTILLON 1988

ET VARIABLES DE PONDÉRATION

L'échantillon observé en 1988 n'est plus conforme à l'échantillon observé en 1985. Phénomènes démographiques et phénomènes propres à la vie de l'échantillon conjuguent leurs effets et provoquent une dérive progressive de l'échantillon.

Les trajectoires des personnes dans l'étude se diversifient de plus en plus. Ces trajectoires déterminent la position des personnes dans la procédure de pondération.

La première partie du chapitre sera consacrée à la description de cette évolution de l'échantillon.

La seconde partie présentera les différentes variables qui ont jalonné l'élaboration du nouveau système de pondérations.

Pour chaque variable, un tableau récapitule les sous-échantillons pertinents et la taille de ces sous-échantillons.

6.1. L'échantillon 1988

6.1.1. Les trajectoires individuelles: version courte

TABLEAU 6.1.
Trajectoires individuelles 1985 - 1988

VARIABLE DE SÉLECTION: DIA88	n =
0. memb. panel refus	1586
1. memb. panel réponse	4413
2. nouv. memb. panel 88	64
3. non-memb. panel réponse	388
4. non-memb. panel refus	53
5. émigrés / décédés	261
TOTAL	6765

Le nouveau fichier contient désormais 6765 personnes classées en 6 catégories principales. Le contenu de ces catégories évolue d'année en année.

Le nouvel échantillon compte 4865 personnes réparties en 3 classes:

- les membres du panel ayant répondu (4413),
- les nouveaux membres du panel entrés dans l'échantillon en 1988 (64)
- et les membres de l'échantillon qui ne sont pas membres du panel (388).

Le taux de réponses est calculé en rapportant le nombre des réponses des *membres du panel* (4413) au nombre total des membres du panel (5999, soit: 1586 + 4413).

Les nouveaux membres du panel entrés dans l'échantillon en 1988 (64) et les personnes émigrées ou décédées ne sont pas prises en compte dans le calcul du taux de réponse.

Ces 6 classes sont définies en fonction du point initial et du point final des trajectoires des individus dans l'ensemble de l'étude. Le point initial détermine le statut de la personne au moment ou elle est entrée dans l'étude: membre ou non-membre du panel. Le point final correspond à la présence ou à l'absence de la personne dans l'échantillon considéré.

Ceci signifie, par exemple, que les membres du panel

* entrés dans l'échantillon initial en 1985

* ayant refusé de répondre en 1986 et en 1987

* peuvent être classés, en 1988

- soit parmi les refus (groupe #0)

- soit parmi les réponses (groupe #1) s'ils ont accepté de renouveler leur coopération, s'ils ont rejoint un ménage d'origine toujours présent dans l'échantillon, etc ...

L'ensemble de toutes ces trajectoires est présenté au tableau **6.2.**. Chaque trajectoire est définie par la succession des états occupés par les personnes au cours des années écoulées.

Le tableau **6.1.** rassemble ces trajectoires en 6 classes principales parce que ces classes

* sont nécessaires et suffisantes pour permettre le calcul des nouvelles pondérations

* permettent de gérer facilement les sous-échantillons principaux qui forment l'ensemble du fichier individuel.

Le tableau **6.2.** permet de recomposer d'autres échantillons en fonction des besoins de l'analyste.

La variable DIA88 comporte six grandes modalités qui peuvent être décrites de la manière suivante.

6.1.1.1. *Les refus des membres du panel*

La première catégorie est formée par les membres du panel ayant refusé de prendre part à l'enquête. Cet ensemble ne correspond plus à la première catégorie de la variable DIA utilisée en 1986 ni à la variable DIA87 utilisée en 1987. Elle réunit des membres du panel présents en 1985 ET des nouveaux membres entrés en 1986 ou en 1987.

Tous ces membres ont un point commun: **ils n'ont pas répondu en 1988.**

6.1.1.2. *Les membres du panel*

Le groupe #1 réunit:

- des membres du panel présents en 1985,
- des nouveaux membres entrés en 1986,
- des nouveaux membres entrés en 1987.

Ils ont un point commun: **ils ont répondu en 1988.**

Tous ces membres sont réunis dans cette catégorie, même s'ils n'ont pas répondu au cours de la seconde ou de la troisième vague d'enquêtes.

6.1.1.3. *Les nouveaux membres du panel*

Le groupe #2 isole les nouveaux membres du panel entrés dans l'échantillon en 1988. Ce sont des nouveaux-nés descendants directs de membres du panel.

Ce groupe ne contient pas les nouveaux membres du panel entrés en 1986 ou en 1987. Ces derniers font désormais partie de l'ensemble des membres du panel. Ils sont traités à ce titre dans la procédure de pondération. Ils sont répartis entre le groupe #0, le groupe #1 et le groupe #5 en fonction de la trajectoire qu'ils ont suivi entre 1986 et 1988.

Ce groupe #2 permet d'identifier les nouveaux membres de l'échantillon annuel qui recevront un poids individuel lié à celui de leurs ascendants. Les membres de ce groupe sont donc différents chaque année.

6.1.1.4. *Les non-membres du panel*

Les groupes #3 et #4 sont formés par l'ensemble des nouvelles personnes entrées dans les échantillons successifs sans accéder au titre de membres du panel.

Ces personnes sont classées dans le groupe #3 lorsqu'elles appartiennent à l'échantillon annuel. Les nouveaux non-membres du panel entrés dans l'échantillon en 1988 sont donc classés dans ce groupe au même titre que les non-membres du panel entrés dans l'étude au cours de vagues antérieures. Ils ont un point commun: **ils ont répondu en 1988.**

Les non-membres du panel, entrés dans les échantillons précédents, sont classés dans le groupe #4 **lorsqu'ils n'ont pas participé à l'enquête en 1988.**

Cette distinction entre les groupes #3 et #4 permet de reconstituer aisément l'échantillon annuel, lorsqu'il y a lieu de prendre en compte les membres du panel ET les non-membres du panel présents dans l'échantillon: il suffit de sélectionner les groupes #1, #2 et #3.

Cette distinction est également très utile au moment de regrouper les individus appartenant à un même ménage ou à un même groupe de revenu.

Elle est indispensable au moment de calculer le poids individuel des nouveaux membres du panel: par exemple, le poids du nouveau membre est différent selon que l'un des ascendants directs est absent de l'échantillon ou non-membre du panel (**6.2.4.**).

6.1.1.5. *Les émigrés et les décédés*

Le groupe #5 réunit toutes les personnes qui ont été présentes dans un échantillon au moins une fois, et qui ne sont pas présentes dans l'échantillon de l'année considérée pour une raison d'ordre *démographique*.

Cette distinction s'impose puisque les émigrés et les personnes décédées ne sont pas prises en compte dans le système de pondération annuel.

Leur classement dans une catégorie particulière facilite également la sélection de l'échantillon annuel et la reconstitution des ménages.

Il faut rappeler qu'une personne émigrée peut rejoindre l'échantillon ultérieurement. Elle sera reclassée dans son groupe d'origine, c'est à dire dans le groupe #1 si elle était membre du panel ou dans le groupe #3 si elle n'était pas membre du panel.

6.1.2. Les trajectoires individuelles: version complète

TABLEAU 6.2.
Liste des trajectoires individuelles: 1985-1988

VARIABLE: DIA4		
CODES	TRAJECTOIRES	n =
1.0	FM/85/0/0/0	826
2.0	FM/85/86/0/0	475
3.0	FM/0/86/0/0	3
4.0	FM/85/0/87/0	1
5.0	FM/85/86/87/0	269
6.0	FM/0/86/87/0	4
7.0	FM/85/E/87/0	4
8.0	FM/0/0/87/0	4
10.0	FM/85/0/0/88	2
11.0	FM/85/86/0/88	121
12.0	FM/0/86/0/88	2
13.0	FM/85/0/87/88	4
14.0	FM/85/86/87/88	4154
15.0	FM/0/86/87/88	68
16.0	FM/0/0/87/88	60
17.0	FM/85/86/E/88	2
18.0	FM/0/0/0/88	64
20.0	NF/0/86/0/0	18
21.0	NF/0/86/87/0	13
22.0	NF/0/0/87/0	22

VARIABLE: DIA4 (suite)		
CODES	TRAJECTOIRES	n =
30.0	NF/0/86/0/88	2
31.0	NF/0/86/87/88	109
32.0	NF/0/0/87/88	118
33.0	NF/0/0/0/88	159
40.0	FM/85/E/-/-	56
41.0	FM/85/86/E/-	20
42.0	FM/85/86/87/E	25
43.0	FM/0/0/87/E	1
50.0	FM/85/D/-/-	50
51.0	FM/85/86/D/-	44
52.0	FM/85/86/0/D	3
53.0	FM/85/86/87/D	54
54.0	FM/0/86/D/-	1
60.0	NF/0/86/D/-	3
61.0	NF/0/86/87/D	1
62.0	NF/0/0/87/D	3
	TOTAL	6765

Le tableau **6.2.** est composé de la manière suivante.

La variable est nommée **DIA4** parce qu'elle correspond à l'état du fichier après la quatrième vague d'enquêtes.

Les 6765 personnes recensées dans le fichier total après quatre vagues d'enquêtes ont suivi 36 trajectoires différentes entre 1985 et 1988 (contre 17, entre 1985 et 1987).

Chaque trajectoire est un vecteur défini par une série d'états successifs. Chaque état correspond au statut occupé par une personne dans le panel ET dans l'échantillon suite à une vague d'observation annuelle.

Chaque vecteur est désigné par un code. Ces codes ont été attribués de manière arbitraire et peuvent être modifiés. Tels qu'ils sont organisés dans le tableau **6.2.**, ils permettent d'effectuer rapidement certains regroupements.

Par exemple, les codes inférieurs à 10.0 désignent l'ensemble des personnes qui ont refusé de participer à l'enquête effectuée en 1988. Les codes 10.0 à 18.0 désignent les membres du panel présents dans l'échantillon en 1988.

Les vecteurs sont décrits de la manière suivante:

* **FM** signifie que ce vecteur contient exclusivement des **membres du panel** (Full Member).

  NF signifie que le vecteur contient exclusivement des personnes **non membres du panel** (Non Full member).

* **85** signifie que ces personnes sont **membres de l'échantillon** observé en 1985.

  86 signifie que ces personnes sont membres de l'échantillon observé en 1986.

  L'année est remplacée par le code **0** lorsqu'une personne est **absente de l'échantillon** suite à un refus ou lorsqu'il n'est pas avéré que cette absence est liée à une raison d'ordre démographique.

  Par exemple: **85/86/0/0** désigne le vecteur des personnes

  - entrées dans l'échantillon au cours de la première vague d'enquête effectuée en 1985
  - toujours présentes dans l'échantillon observé en 1986
  - mais absentes de l'échantillon en 1987 et en 1988 pour des raisons assimilées à des refus.

* **E** signifie que ces personnes ont **émigré** et sont absentes de l'échantillon correspondant à la vague d'enquête de cette année. Dans certains cas cette absence est provisoire.

Le vecteur #7 est composé de la manière suivante:

- **FM. 85/E/87/0** soit 4 membres du panel, entrés dans l'échantillon en 1985, émigrés en 1986 et revenus dans l'échantillon en 1987 et refusant de répondre en 1988.

* **D** signifie que ces personnes sont **décédées.**

Le vecteur #62 est composé de la manière suivante:

- **NF/0/0/87/D** soit 3 personnes entrées dans l'étude en 1987 sans accéder au titre de membres du panel et décédées au cours de l'année 1988.

Le passage de la variable DIA4 à la variable DIA88 s'opère aisément. Les classes générales de trajectoires sont composées à partir du point initial et du point final de chaque vecteur.

Le **groupe #1** de la variable DIA88 regroupe l'ensemble des membres du panel entrés dans l'échantillon en 1985, en 1986 ou en 1987 et présents dans l'échantillon en 1988, quelles que soit les étapes intermédiaires ayant jalonné leur trajectoire. Cette classe est donc formée par les vecteurs #10 à #18 de la variable DIA4:

Exemples:

10.0 **FM/85/0/0/88**
soit les membres du panel entrés en 1985, absents en 1986 et en 1987 mais revenus dans l'échantillon en 1988.

11.0 **FM/85/86/0/88**
soit les membres du panel entrés en 1985, présents dans l'échantillon en 1986, absents en 1987 parce qu'ils ont refusé de répondre, mais revenus en 1988.

14.0 **FM/85/86/87/88**
soit les membres du panel entrés en 1985 et présents dans tous les échantillons annuels.

18.0 **FM/0/0/0/88** soit les 64 nouveaux membres du panel entrés en 1988.

La procédure de pondération exige que ce dernier vecteur (#18) puissent être isolé facilement. Il correspond au groupe #2 de la variable DIA88.

Le regroupement des vecteurs #10 à #18 de la variable DIA4 équivaut au regroupement des classes #1 et #2 de la variable DIA88: il désigne l'ensemble des membres du panel présents dans l'échantillon en 1988 soit 4477 personnes (dont 64 nouveaux membres entrés dans l'échantillon en 1988).

6.2. Les variables de pondération

6.2.1. Variable instrumentale TREE88

Les refus de répondre constituent le facteur principal de la déformation de l'échantillon (**2.2.1.1.**).

1. La notion de "taux de réponses" prendra désormais un sens particulier. Elle correspond davantage à la notion de "taux de survie dans le panel" (**6.2.1.1**).

2. Les refus de répondre seraient sans effet s'ils étaient répartis aléatoirement dans l'échantillon. Il convient donc de rechercher les concentrations éventuelles des refus dans des groupes démographiques homogènes afin d'en corriger les effets (**6.2.1.3.**).

3. Au préalable, l'inégalité des probabilités de sélection initiales des membres du panel doit être annulée en pondérant les individus par l'inverse de cette probabilité. La recherche des groupes démographiques atteints par les taux de refus les plus contrastés ne doit pas être influencée par ce biais lié au mode de tirage de l'échantillon initial (**6.2.1.2.**).

6.2.1.1. *Des taux de survie*

Il est important de noter que la notion de taux de réponses sera entendue désormais dans un sens particulier.

Le terme "taux de survie" est plus correct que le terme "taux de réponses". En effet, le **taux de survie** prend en compte le **cumul** des refus enregistrés au cours des enquêtes successives à condition que ces personnes n'aient pas rejoint l'échantillon en 1988.

Le taux de réponses correspond désormais au nombre des réponses enregistrées au cours d'une vague d'enquête, rapporté au nombre cumulé des membres du panel.

> *Il n'est pas calculé par référence au nombre des membres du panel présents dans l'échantillon précédent. Il est calculé en rapportant le nombre de réponses des membres du panel au nombre total des membres du panel entrés dans les échantillons successifs depuis la première vague d'enquête.*

Cette procédure permet de prendre en compte la mobilité des membres du panel: ils peuvent entrer et sortir de l'échantillon. Chaque fois qu'ils rejoignent l'échantillon, leur

nouvelle probabilité de sélection peut être calculée puisque les deux éléments essentiels du calcul sont connus:

* leur probabilité de sélection initiale (**6.2.1.2.**)

* et les caractéristiques démographiques du sous-groupe auquel ils appartenaient au moment ou ils sont entrés dans l'échantillon (**6.2.1.3.**).

6.2.1.2. *Correction du biais initial*

L'analyse des taux de réponses prend en compte les 5999 personnes regroupées dans les classes #0 et #1 de la variable DIA88. Elle ne prend en compte, ni les nouveaux membres du panel entrés dans l'échantillon au cours de la dernière vague d'enquête, ni les non-membres du panel, ni les personnes émigrées ou décédées.

Parmi les 5999 membres du panel, entrés dans l'échantillon de 1985 à 1987, 4413 ont répondu en 1988, soit 73.6%.

Ces valeurs ne prennent pas en compte l'inégalité des probabilités de sélection initiales des individus. Il convient donc d'annuler cette inégalité avant d'entreprendre toute autre opération.

Les membres entrés dans l'échantillon en 1985 doivent être pondérés par l'inverse de leur probabilité de sélection soit, par la variable PONREV.

La probabilité de sélection des nouveaux membres correspond à la valeur moyenne des probabilités de sélection de leurs parents *au moment ou ces nouveaux membres entrent dans le panel*. Ces nouveaux membres ont déjà reçu ce poids au moment ou ils sont entrés dans l'échantillon: c'est la valeur de la variable W86B attribuée en 1986 aux nouveaux membres de l'année; c'est la valeur de W87B attribuée en 1987 aux nouveaux membres entrés en 1987.

La synthèse de ces valeurs donne lieu à la création de la variable **IP8588**. Cette variable est sauvegardée dans le fichier global parce qu'elle sera réutilisée l'année prochaine dans le cadre de la même procédure. (Cette variable doit être ajustée en fonction de la taille de l'échantillon).

Cette première procédure modifie la valeur du taux de réponses. Le tableau **6.3.** permet d'observer les glissements des proportions avant et après la correction des biais de sélection.

TABLEAU 6.3.

Distribution des fréquences des individus membres du panel, en 1988, selon leur position dans l'échantillon, avant et après pondération

DIA88	FREQUENCES			
	avant pondération		après pondération (IP8588)	
	n	%	n	%
0. memb.pan. refus	1586	26.4	1506	25.1
1. memb.pan. réponses	4413	73.6	4493	74.9
TOTAL	5999	100.0	5999	100.0

Le taux de réponses passe de 73.6% à 74.9% lorsque les biais de sélection initiaux sont contrôlés.

6.2.1.3. *Les nouvelles probabilités de sélection*

La qualité de l'échantillon n'est pas affectée par les refus de répondre aussi longtemps qu'ils se distribuent de manière aléatoire dans l'ensemble de l'échantillon.

La méthode permettant de repérer des concentrations éventuelles de ces refus dans certains sous-groupes homogènes de l'échantillon a été présentée dans le second chapitre (2.2.1.1.).

La démarche met en concurrence le pouvoir discriminant de 17 variables:

CARACTERISTIQUES DU MÉNAGE	DU CHEF DE MÉNAGE	DE LA PERSONNE
Canton de résidence	Age	Age
Type d'habitat	Sexe	Sexe
Taille du ménage	Etat civil	Etat civil
Nombre d'adultes	Emploi (oui/non)	Formation
Nombre d'enfants		Emploi
Nombre d'emplois		Adulte/Enfant
		Lien avec le CM

Toutes ces caractéristiques sont attribuées aux personnes. Les caractéristiques du ménage et les caractéristiques du chef de ménage sont des éléments de l'environnement de la personne.

Les membres entrés dans l'échantillon en 1985 sont caractérisés par les valeurs qu'ils présentaient sur ces variables **en 1985**, au moment où ils sont entrés dans l'échantillon.

Les nouveaux membres, entrés en 1986 et en 1987, sont caractérisés par les valeurs qu'ils présentaient sur ces différentes variables au moment de leur entrée dans l'échantillon. Ces variables décrivent donc le contexte du ménage dans lequel ils sont nés et *au moment où ils sont nés.*

Il est donc fort probable qu'il existe un décalage systématique entre les caractéristiques des nouveaux membres et les caractéristiques de leurs parents.

Si la taille du ménage dans lequel ils font leur entrée s'est modifiée entre 1985 et 1987, les nouveaux membres du panel entrés en 1987 ne sont pas caractérisés par la même valeur que leurs parents sur cette variable.

L'âge du chef de ménage fournit un exemple simple: en règle générale, l'âge du chef de ménage sera différent pour les parents et pour le nouveau membre du panel entré dans la famille deux ans plus tard.

La synthèse de ces valeurs donne lieu à la création de nouvelles variables. Ces variables sont sauvegardées dans le fichier global parce qu'elles seront réutilisées l'année suivante dans le cadre de la même procédure.

(La variable de correction des probabilités de sélection initiales a été actualisée selon la même logique (**6.2.1.2.**)).

Lorsque l'ensemble des variables choisies ont été actualisées, la recherche des catégories démographiques les plus discriminantes peut être entreprise (**2.2.1.1.**).

En résumé, la procédure consiste à diviser l'échantillon en sous-groupes mutuellement exclusifs par une succession de divisions binaires. L'enchaînement de ces divisions successives crée des filières ou séries de modalités. Les variables qui produisent ces divisions sont choisies parce qu'elles créent les contrastes les plus significatifs entre les taux de réponses des deux sous-groupes. Ces divisions sont interrompues dès qu'un sous-groupe ne peut plus être divisé en deux sous-groupes d'une taille minimale de 200 personnes. (Cf. Procedures and Tape Codes, 1978 Interviewing Year, Wave XI, a Supplement. P.S.I.D., I.S.R., Ann Arbor, Michigan, 1979, p.10).

Le résultat de cette analyse est présenté dans le schéma **6.1.**. Il reproduit sous une forme arborescente la progression de l'analyse vers des sous-groupes de plus en plus homogènes.

Vingt et un sous-groupes sont identifiés à l'aide de 12 variables.

La variable dépendante est une variable binaire (0,1) qui rend compte du **taux de survie** dans le panel longitudinal:

$$\text{Taux de survie} = \frac{\text{Nombre de codes 1}}{\text{Nombre de cas dans le sous-groupe}}$$

Les 21 sous-groupes "n'expliquent" que 6.98% de la variance de la variable dépendante.

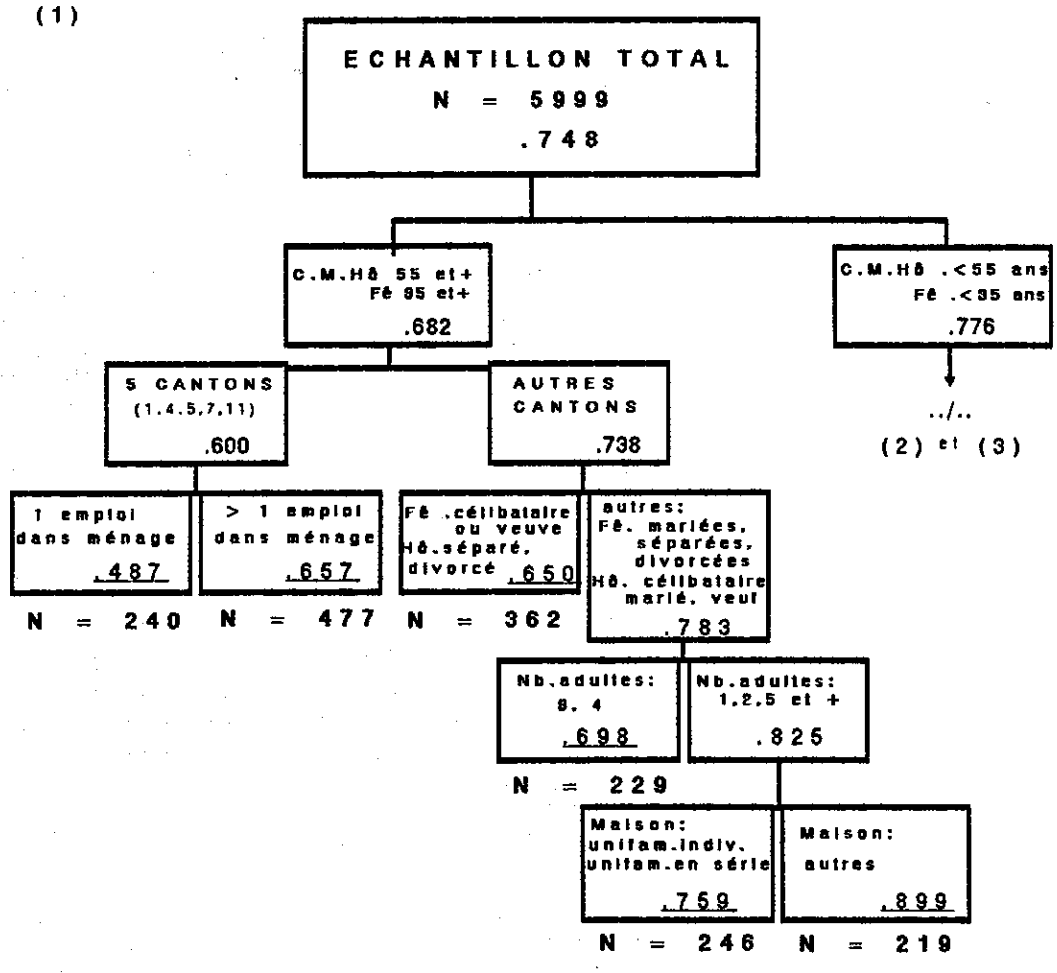
L'objectif de cette procédure n'était pas de prendre en compte toute la variance des taux de survie. Des mesures ont été prises en vue d'éviter que l'analyse identifie des groupes marginaux de très petite taille. La procédure évite que la variance des poids des individus soit démesurément accrue, en maintenant la taille des sous-groupes à un minimum de 200 personnes.

Le faible pouvoir explicatif de la typologie constituée par cette procédure provient également du fait que les variables démographiques de base divisent l'échantillon en modalités dont les moyennes sont peu contrastées (tableau **6.4.**).

SCHEMA 6.1.

Analyse des taux de réponses des sous-groupes

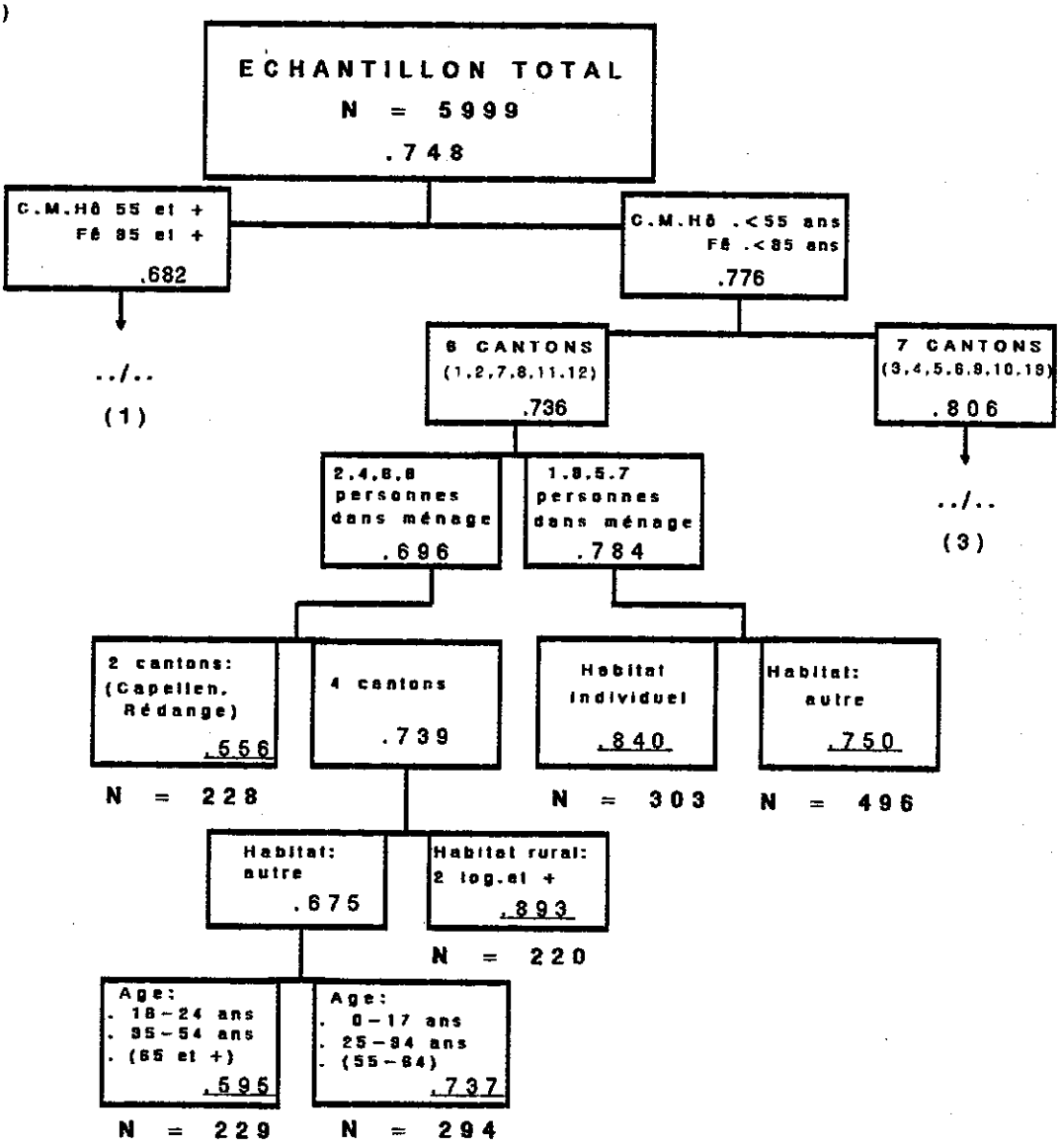
entre 1985 et 1988 (après pondération initiale)



Analyse des taux de réponses des sous-groupes

entre 1985 et 1988 (après pondération initiale)

(2)



Analyse des taux de réponses des sous-groupes

entre 1985 et 1988 (après pondération initiale)

(3)

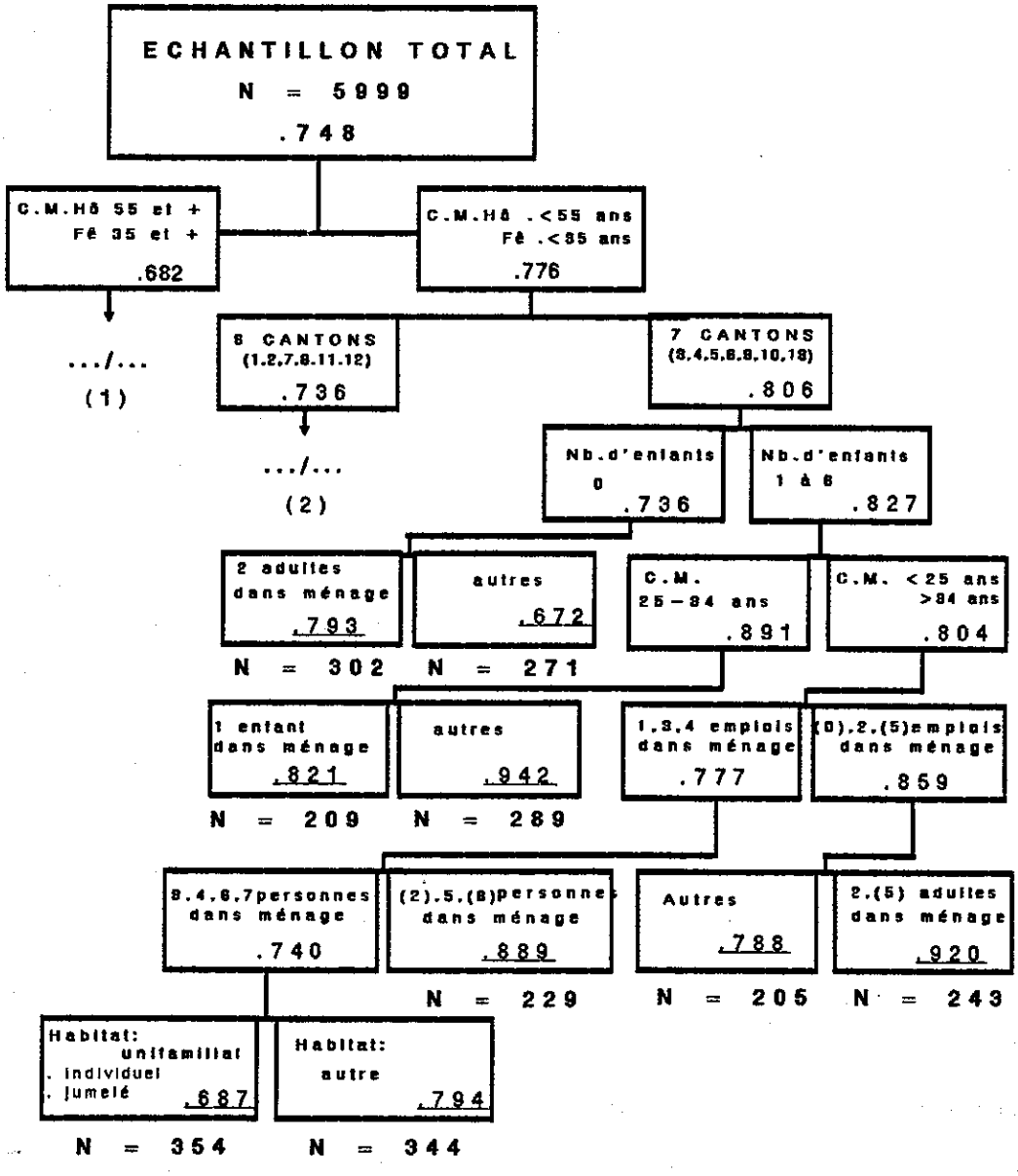


TABLEAU 6.4.

Taux de survie des membres du panel en 1988

CARACTÉRISTIQUES DES SOUS-GROUPES	TAUX DE REPONSE $\bar{x}$ = .748	VARIANCE "EXPLIQUÉE" (ETA)	VARIANCE EXPLIQUÉE PAR LA MEILLEURE DICHOTOMISATION
Caractéristiques individuelles			
Sexe			
1. masculin	.75		
2. féminin	.74	.00	.00
Etat civil			
1. célibataire	.75		
2. marié	.76		
3. séparé	.66	.07	.07
4. divorcé	.66		
5. veuf	.65		
6. D.M.	(.00)*		
Age			
1. < 17 ans	.79		
2. 18 à 24 ans	.68		
3. 26 à 34 ans	.76		
4. 35 à 44 ans	.75	.10	.09
5. 45 à 54 ans	.75		
6. 55 à 64 ans	.74		
7. 65 ans et plus	.66		
Nationalité			
1. Luxembourgeois	.74		
2. CEE	.78		
3. hors CEE	.62	.07	.04
4. D.M.	(.08)		

* Les chiffres entre parenthèses correspondent à des catégories de moins de 20 personnes.

Situation professionnelle			
1. Enfants	.78		
2. Travail ou maladie	.74		
3. Demandeur d'emploi	.61	.06	.04
4. Pensionné, invalide	.71		
5. Retraité	.70		
6. Ménage, milice	.73		
Lien avec le chef de ménage			
1. Chef de ménage	.74		
2. épouse	.76		
3. "épouse"	.73	.06	.06
4. ami(e)	(.62)		
5. fils, fille	.75		
6. autre	.62		
Adultes			
1. Enfant	.78		
2. Adulte	.73	.05	.05
Etat civil selon le sexe			
1. Homme célibataire	.74		
2. marié	.76		
3. séparé	.60		
4. divorcé	.69		
5. veuf	.69		
6. M.D.	(.00)	.07	.07
7. Femme célibataire	.75		
8. mariée	.76		
9. séparée	.72		
10. divorcée	.64		
11. veuve	.64		
Emploi selon le sexe			
1. Garçon mineur	.77		
2. Homme ayant un emploi	.74		
3. Homme sans emploi	.72	.05	.04
4. Fille mineure	.79		
5. Femme ayant un emploi	.73		
6. Femme sans emploi	.72		

Age selon le sexe

1. Homme, < de 18 ans	.79		
2. de 18 à 64 ans	.73		
3. > 64 ans	.71	.09	.07
4. Femme, < de 18 ans	.80		
5. 18 à 64 ans	.74		
6. > 64 ans	.63		

Niveau de formation terminé

1. Enfant	.92		
2. Pas de formation	.76		
3. primaire ou complémentaire	.73		
4. secondaire	.74	.08	.03
5. supérieur non-universitaire	.67		
6. supérieur universitaire	.72		
7. Ne sait pas	.74		

Caractéristiques du ménage

Nombre de personnes dans le ménage

1.	.69		
2.	.74		
3.	.75		
4.	.73		
5.	.80	.06	.05
6.	.73		
7.	.72		
8.	.83		
9.	(.89)		
11.	(.90)		
12.	(1.00)		

Nombre d'enfants dans le ménage

0.	.71		
1.	.76		
2.	.76		
3.	.78	.07	.06
4.	.85		
5.	(1.00)		
6.	(.77)		

Nombre d'adultes dans le ménage

0.	(.00)		
1.	.71		
2.	.77		
3.	.71	.09	.08
4.	.65		
5.	.74		
6.	.74		
7.	(.89)		
9.	(1.00)		

Nombre d'emplois dans le ménage

0.	.72		
1.	.75		
2.	.77		
3.	.66	.06	.04
4.	.71		
5.	.86		
6.	(.78)		

Type d'habitation

1. maison rurale	.75		
2. maison unifamiliale individuelle	.74		
3. -"- jumelée	.73		
4. -"- en série	.75		
5. petit collectif	.75		
6. collectif moyen	.72	.02	.01
7. grand collectif	.81		
8. précaire	(.71)		
9. Non réponse	.78		

Canton de résidence

1. Luxembourg-Ville	.70		
2. Luxembourg-Campagne	.68		
3. Capellen	.77		
4. Esch-sur-Alzette	.77		
5. Mersch	.74		
6. Clervaux	.85	.10	.09
7. Diekirch	.71		
8. Rédange	.70		
9. Vianden	.91		
10 Wiltz	.79		
11 Echternach	.66		
12 Grevenmacher	.75		
13 Remich	.79		

Caractéristiques du chef de ménage

Age du chef de ménage			
1. < 24 ans	.74		
2. 25 à 34 ans	.82		
3. 35 à 44 ans	.76		
4. 45 à 54 ans	.73	.10	.08
5. 55 à 64 ans	.69		
6. 65 et plus	.67		
Sexe du chef de ménage			
1. masculin	.75		
2. féminin	.67	.06	.06
Etat civil du chef de ménage			
1. Célibataire	.77		
2. marié	.76		
3. séparé	.64		
4. divorcé	.63	.08	.07
5. veuf	.67		
6. Non réponse	(.00)		
Situation professionnelle du C.M.			
1. sans emploi	.70		
2. travaille	.76	.05	.05
Etat civil selon le sexe du C.M.			
1. Homme non marié	.71		
2. marié	.76		
3. Femme non mariée	.68	.07	.06
4. mariée	.43		
Situation professionnelle du C.M. selon le sexe			
1. Homme sans emploi	.72		
2. au travail	.76		
3. Femme sans emploi	.67	.07	.06
4. au travail	.69		
Age du C.M. selon le sexe			
1. Homme, < 35 ans	.80		
2. 35 à 44 ans	.76		
3. 45 à 54 ans	.74		
4. 55 à 64 ans	.70		
5. 65 ans et +	.69	.11	.10
6. Femme, < 35 ans	.85		
7. 35 à 44 ans	.62		
8. 45 à 54 ans	.65		
9. 55 à 64 ans	.65		
10. 65 ans et +	.64		

La lecture détaillée de l'arbre (schéma **6.1**) n'apporte pas d'informations très claires.

Trois points peuvent être mis en évidence.

1. Un contraste très prononcé entre la métropole et les autres cantons du pays avait été observé au terme de la deuxième vague d'enquête. Il s'est effacé.

2. Une caractéristique propre aux individus a fait son apparition au terme de la troisième vague d'enquête (l'âge des individus). Cette caractéristique intervient à nouveau au terme de la quatrième vague d'enquête: les variables décrivant le cadre familial ne monopolisent plus l'organisation des taux de réponse.

   Si les variables individuelles devaient prendre, par la suite, de plus en plus de poids dans le système de pondération, il s'avèrerait que les panels socio-économiques devraient éviter de procéder à des pondérations directes des ménages sans prendre en compte l'hétérogénéité interne de cette unité d'analyse.

3. L'écart entre le taux de survie maximum et le taux de survie minimum a connu une croissance notable entre la deuxième et la troisième vague. Par contre, **cet écart est resté constant** en 1987 et en 1988.

	1986	1987	1988
Taux le plus élevé	1.00	.98	.94
Taux le plus bas	.71	.54	.48

TABLEAU 6.5.
Ventilation des valeurs de TREE88
dans le fichier individuel

DIA88	TREE88 n =
0. memb. panel refus	1586
1. memb. panel réponse	4413
2. nouv. memb. panel 88	-
3. non-memb. panel réponse	-
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	5999
sysmis	766
TOTAL	6765

Cette variable TREE88 décrit le profil des catégories d'appartenance des membres du panel, qu'ils aient répondu ou non en 1987. Elle permet donc de classer tous les membres des classes #0 et #1 de la variable DIA88 (tableau **6.5.**).

TREE88 n'exprime pas des probabilités de survie. Elle décrit simplement une typologie de sous-échantillons. Elle ne permet donc pas de calculer le poids relatif des personnes dans le nouvel échantillon. Ce sera le rôle d'une autre variable: RR88.

6.2.2. Variable instrumentale RR88

La variable RR88 est construite de la manière suivante: toute personne ayant effectivement répondu en 1988 (DIA88 #1), reçoit la probabilité de sélection observée dans la catégorie de population à laquelle elle appartient (TREE88) (tableau **6.6.**).

TABLEAU 6.6.

Distribution des fréquences de l'échantillon 1988, selon les taux de réponses observés en 1988 (avant et après pondération de l'échantillon initial)

RR88 taux de réponses en 1988	TREE88 code	FREQUENCES avant pondération n =	FREQUENCES après pondération (IP8588) n =
.48	(1)	121	117
.55	(7)	110	127
.59	(8)	150	137
.65	(3)	278	235
.65	(2)	339	313
.67	(14)	277	182
.68	(17)	200	244
.69	(4)	252	161
.73	(9)	184	217
.75	(12)	326	373
.75	(5)	158	187
.78	(20)	251	162
.79	(13)	182	240
.79	(18)	221	274
.82	(15)	139	172
.84	(11)	250	255
.88	(19)	198	204
.89	(10)	184	197
.89	(6)	211	198
.92	(21)	173	224
.94	(16)	209	273
TOTAL		4413	4493

Il n'y a pas lieu de s'étonner si la taille de l'échantillon n'est plus identique avant et après le redressement de l'échantillon. Seule une fraction de l'échantillon des membres du panel a répondu en 1988: 4413 personnes sur 5999.

Ces 4413 personnes ont été pondérées (par IP8588) dans le cadre de **l'échantillon total**, afin de compenser les variations des probabilités de sélection initiales. Ces poids ont été ajustés: la somme des poids est égale au nombre des individus, soit 5999. Ils n'ont pas été ajustés dans le cadre des 4413 personnes ayant effectivement participé à l'enquête (tableau **6.3.**).

La différence entre les tailles de la **fraction de l'échantillon** avant et après pondération est donc parfaitement normale. En outre, la comparaison des deux dernières colonnes du tableau **6.6.** permet de constater le fait que certaines catégories de la population

auraient été sur-représentées ou sous-représentées si l'échantillon n'avait pas été pondéré avant d'être soumis à l'analyse des taux de réponses.

TABLEAU 6.7.
Ventilation des valeurs de RR88 dans le fichier individuel

DIA88	RR88 n =
0. memb. panel refus	-
1. memb. panel réponse	4413
2. nouv. memb. panel 88	-
3. non-memb. panel réponse	-
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4413
sysmis	2352
TOTAL	6765

Les valeurs de RR88 ne sont donc attribuées qu'aux seuls membres du panel ayant répondu en 1988 (tableau **6.7.**).

6.2.3. Variable instrumentale W88

La variable W88 définit le poids des membres du panel ayant répondu en 1988, **exception faite des nouveaux membres du panel entrés dans l'échantillon en 1988.**

Ces anciens membres appartiennent à différentes catégories de la population et se distinguent par la probabilité qu'ils répondent en 1988.

Lorsque le groupe enregistre un taux de réponses très faible, chaque "survivant" doit représenter un plus grand nombre de personnes présentant les mêmes caractéristiques démographiques. Lorsque le taux de réponses s'élève, l'importance de cet ajustement se réduit.

D'une manière générale, cet ajustement des poids des individus sera inversement proportionnel à leur probabilité de sélection c'est à dire, inversement proportionnel au taux de réponses enregistré dans leur groupe:

Soit: 1 / RR88

Dans le cas présent, ce calcul doit prendre en compte un élément supplémentaire: l'inégalité des probabilités de sélection initiales des membres du panel. Les nouveaux poids ne s'appliquent pas à des personnes "unités" mais à des fractions ou à des multiples de l'unité définis par la variable IP8588.

La nouvelle variable de pondération se calcule donc de la manière suivante:

W88 = IP8588 x (1 / RR88)

TABLEAU 6.8.
Ventilation des valeurs de W88 dans le fichier individuel

DIA88	W88 n =
0. memb. panel refus	-
1. memb. panel réponse	4413
2. nouv. memb. panel 88	-
3. non-memb. panel réponse	-
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4413
sysmis	2352
TOTAL	6765

Les valeurs de cette variable sont attribuées exclusivement aux MEMBRES DU PANEL présents dans l'échantillon en 1988, exception faite des nouveaux membres entrés dans l'échantillon en 1988 (tableau **6.8.**). Le système de pondération doit donc être élargi à ces nouveaux membres.

6.2.4. Variable instrumentale W88B

La variable W88 ne permet pas encore de pondérer tout l'échantillon observé en 1988, parce qu'elle ne prend pas en compte les nouveaux membres de l'échantillon (membres ou non du panel). Il est indispensable d'étendre la portée de cette variable pour pouvoir calculer les poids des ménages.

Dans un premier temps, **les nouveaux membres de l'échantillon, non membres du panel**, reçoivent un poids nul ('0') puisqu'il est impossible de calculer leur probabilité de sélection initiale. En effet:

- ils étaient déjà présents dans la population de 1985, lorsque l'échantillon initial a été sélectionné,

- ils sont entrés dans l'échantillon par la suite et pour des raisons variées: il est donc difficile de reconstituer l'ensemble des conditions de leur entrée dans l'échantillon et d'en dériver une logique de pondération[1].

Il est indispensable de pondérer ces non-membres avant d'attribuer des poids aux **nouveaux membres de l'échantillon, membres du panel.** En effet, certains nouveaux membres de l'échantillon, **non-membres du panel** peuvent être des ascendants directs d'**enfants nouveaux membres du panel.**

Or, dans un second temps, ces nouveaux membres du panel reçoivent un **poids égal à la moyenne des poids des ascendants directs** tels qu'ils ont été calculés au moment où le nouveau membre est entré dans l'échantillon (et dans la population).

Donc:

- lorsque les ascendants directs ont des poids différents de '0', le nouveau membre reçoit un poids égal à la moyenne de ces deux poids;

- lorsque l'un des ascendants directs a un poids nul, le nouveau membre du panel reçoit logiquement un poids égal à la moitié du poids de l'ascendant direct, membre du panel;

- lorsque les deux ascendants directs ne sont pas membres du panel, le nouveau membre reçoit un poids nul: il n'est pas membre du panel.

Par contre, un nouveau membre reçoit le poids de l'ascendant direct présent dans l'échantillon observé, lorsque l'autre ascendant est ABSENT de l'échantillon au moment où le nouveau entre dans l'échantillon: la probabilité de sélection initiale de ce nouveau membre dépend entièrement de la probabilité de sélection du seul ascendant présent dans l'échantillon. (Cfr. **2.2.1.5.** et **2.2.1.6.**).

1. Le poids relatif de ces individus est certainement différent de '0' mais toute affectation d'un poids différent de '0' reposerait sur un ensemble d'hypothèses comportant des erreurs qu'il est impossible d'estimer.

TABLEAU 6.9.
Ventilation des valeurs de W88B
dans le fichier individuel

DIA88	W88B n =
0. memb. panel refus	-
1. memb. panel réponse	4413
2. nouv. memb. panel 88	64
3. non-memb. panel réponse	388
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4865
sysmis	1900
TOTAL	6765

Ces poids sont attribués à tous les membres de l'échantillon observé en 1988, soit 4865 personnes. Parmi celles-ci, 388 personnes ont un poids égal à 0 (groupe #3 de la variable DIA88) (tableau **6.9.**).

La somme des poids est égale à 4377.65, quel que soit le nombre de personnes prises en compte: 4865 **personnes interrogées** ou 4477 **membres du panel.**

La taille "réelle" de l'échantillon est égale à 4477 membres du panel. Les 388 personnes présentes dans cet échantillon sans être membres du panel ont reçu un poids nul, précisément parce qu'il est impossible d'estimer le poids de leur présence dans cet échantillon individuel.

La somme des poids doit être ajustée à la taille de l'échantillon. La somme des poids doit être égale au nombre de membres du panel afin de sauvegarder les propriétés statistiques de l'échantillon. C'est le rôle de la variable WGT88TOT.

6.2.5. Variable opérationnelle WGT88TOT

La taille de l'échantillon pondéré par la variable W88B serait réduite à 4377 unités. Les poids doivent donc être ajustés par une constante: la moyenne des poids attribués aux individus (**2.2.1.4**). Suite à cette opération, la valeur moyenne des poids sera égale à 1.0 et la somme des poids sera égale au nombre de membres du panel.

Cet ajustement est effectué sur 4477 personnes **membres du panel**. Il s'en suit que les analyses des données peuvent être réalisées de manière équivalente sur l'échantillon de 4865 **personnes observées** en 1988 ou sur l'échantillon de 4477 **membres du panel** observés en 1988. Dans les deux cas, la taille réelle de l'échantillon sera respectée (4477 unités).

Dans le premier cas le nombre total de personnes prises en compte sera **apparemment** égal à 4865, mais la somme des poids sera égale à 4477 unités. Seuls, les membres du panel sont pris en compte au cours des analyses, puisque 388 personnes ont un poids égal à 0: les valeurs de ces personnes sur les variables analysées sont annulées.

Dans le second cas, seuls les membres du panel sont pris en compte dans l'analyse et la mise à l'écart des non-membres du panel n'affecte en rien les résultats (**2.2.1.7.**).

La variable W88B devient:

WGT88TOT = W88B / (4377.65 / 4477)

TABLEAU 6.10.
Ventilation des valeurs de WGT88TOT dans le fichier individuel

DIA88	WGT88TOT n =
0. memb. panel refus	-
1. memb. panel réponse	4413
2. nouv. memb. panel 88	64
3. non-memb. panel réponse	388
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4865
sysmis	1900
TOTAL	6765

Les valeurs de la variable WGT88TOT sont attribuées à l'ensemble des personnes ayant participé à l'enquête de 1988, soit les groupes #1, #2 et #3 de la variable DIA88 (tableau **6.10.**).

6.2.6. Variable opérationnelle de pondération des ménages en 1988 MWGT88A

La variable MWGT88A permet de pondérer correctement l'échantillon des **ménages** observé en 1988.

En 1985, le biais d'échantillonnage apparaît au niveau des ménages. La correction de ce biais s'effectue, en premier lieu, au niveau des ménages. Elle est répercutée sur les membres du ménage. Ceux-ci reçoivent le même poids que le ménage auquel ils appartiennent.

Il existe donc un rapport constant entre le poids du ménage et le poids des individus: le poids du ménage est égal à la valeur moyenne des poids des personnes qui le composent.

Depuis 1986, l'échantillon subit l'influence de facteurs d'évolution qui introduisent des biais aux deux niveaux d'analyse.

Tantôt un ménage complet disparaît et la raison du refus est commune à l'ensemble des membres du ménage.

Tantôt, un seul membre disparaît pour des raisons strictement individuelles.

Des nouveaux membres entrent dans les ménages. Les uns sont pris en compte au titre de membres du panel, les autres ne sont pas reconnus comme membres du panel.

De nouveaux ménages se créent et, généralement, provoquent l'inclusion d'une nouvelle personne étrangère à l'échantillon initial. Des ménages se défont et provoquent tantôt l'apparition de deux nouveaux ménages, tantôt la création d'un nouveau ménage et le retour d'une personne dans son milieu d'origine.

Cette liste n'est pas exhaustive. Elle illustre suffisamment à quel point les trajectoires des personnes et les trajectoires des ménages se différencient progressivement. Elle invite à fonder l'analyse de l'évolution de l'échantillon au niveau des trajectoires individuelles plutôt qu'au niveau des ménages.

Cependant, le rapport entre la valeur des poids relatifs individuels et la valeur des poids des ménages garde sa propriété initiale: le poids du ménage est égal à la valeur moyenne des poids des individus.

La procédure est simple.

* Les valeurs des poids individuels sont agrégées au niveau du ménage.

* La valeur moyenne des poids individuels (WGT88TOT) détermine la valeur du poids du ménage (MWGT88T).

  Le poids du ménage est calculé en fonction du **nombre de personnes ayant répondu** et non en fonction du nombre de personnes présentes dans le ménage.

* Cette valeur peut être ajustée en fonction de la taille de l'échantillon des ménages (MWGT88A).

La taille "réelle" de l'échantillon des ménages correspond à **1673 ménages composés d'au moins un membre du panel.**

En 1988, 19 ménages sont entièrement composés par des personnes qui ne sont pas membres du panel: ils reçoivent logiquement un poids de '0'.

L'ajustement des poids des ménages s'effectue donc sur un nombre de ménages égal à 1673 unités et non sur 1673 + 19 ménages. De cette façon, les analyses seront toujours effectuées sur 1673 unités, indépendamment du fait que les ménages composés exclusivement de non-membres du panel sont pris en compte ou sont mis à l'écart (**2.2.1.7.**). A défaut d'un tel ajustement, l'échantillon pondéré par MWGT88T compterait 1668 unités. D'où:

MWGT88A = MWGT88T / (1667.625 / 1673)

TABLEAU 6.11.
Ventilation des valeurs de MWGT88A dans le fichier individuel

DIA88	MWGT88A n =
0. memb. panel refus	-
1. memb. panel réponse	4413
2. nouv. memb. panel 88	64
3. non-memb. panel réponse	388
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4865
sysmis	1900
TOTAL	6765

Les valeurs des poids des ménages (MWGT88T et MWGT88A) sont reventilées sur les individus. Chaque membre du ménage reçoit la valeur du poids du ménage auquel il appartient. Les non-membres du panel reçoivent, cette fois, un poids différent de 0, à condition qu'ils n'appartiennent pas à un ménage composé exclusivement de non-membres du panel. Cette procédure, purement technique, permet de regrouper l'ensemble des poids dans un seul fichier individuel.

Le tableau **6.11.** présente les sous-échantillons individuels qu'il convient de sélectionner dans le fichier au moment de reconstituer l'échantillon des ménages observés en 1988. Les poids des ménages s'obtiennent aisément, par exemple, en agrégeant les valeurs de MWGT88A au niveau des ménages et en prenant la moyenne comme valeur de

pondération. (Certains logiciels permettent de prendre la valeur du premier membre du ménage comme valeur de référence (SPSSX)).

Il peut être utile de rappeler que TOUS les membres appartenant à l'échantillon observé en 1988 doivent être pris en compte au moment de reconstituer l'échantillon des ménages. En l'absence des non-membres du panel, il serait impossible de calculer certaines variables propres au ménage (par exemple, la taille du ménage).

ANNEXE 6 A

TABLEAU COMPARATIF DES CARACTERISTIQUES DEMOGRAPHIQUES DES MEMBRES DU PANEL EN 1985, EN 1986, EN 1987 ET EN 1988 (*après pondération*)

- caractéristiques personnelles
- caractéristiques du ménage
- caractéristiques du chef de ménage

Caractéristiques Personnelles	1985 (N=6110)	1986 (N=5245)	1987 (N=4648)	1988 (N=4477)
Sexe				
1. homme	48.8	48.7	48.2	48.0
2. femme	51.2	51.3	51.8	52.0
Age				
1. < 18 ans	24.1	23.5	23.6	22.9
2. 18 - 24 ans	11.2	11.2	11.1	11.0
3. 25 - 34 ans	15.6	15.1	15.2	14.6
4. 35 - 44 ans	14.3	14.3	14.6	15.1
5. 45 - 54 ans	12.7	12.8	12.6	13.0
6. 55 - 64 ans	10.4	10.9	10.5	11.5
7. > 64 ans	11.7	12.1	12.4	11.9
Age par Sexe				
1. homme < 18 ans	12.7	12.4	12.0	11.6
2. homme 18 - 64 ans	31.7	31.8	31.5	32.0
3. homme > 64 ans	4.4	4.6	4.7	4.4
4. femme < 18 ans	11.4	11.2	11.6	11.2
5. femme 18 - 64 ans	32.6	32.6	32.5	33.2
6. femme > 64 ans	7.2	7.5	7.7	7.5
Etat civil				
1. célibataire	39.8	39.9	40.0	40.0
2. marié	49.6	49.2	48.8	48.9
3. séparé	1.0	0.8	0.8	0.9
4. divorcé	1.9	1.9	2.3	2.0
5. veuf	7.9	8.0	8.1	8.2
6. N.R.	0.0	-	-	-

Nationalité

1. Luxembourgeois	76.1	76.4	77.9	78.7
2. autres C.E.	21.3	21.3	20.3	19.7
3. hors C.E.	2.5	2.3	1.9	1.6
4. N.R.	0.1	-	-	-

Emploi

1. oui	38.1	39.8	37.1	40.3
2. non	35.4	35.3	34.2	34.6
3. enfants	26.5	24.9	28.7	25.1

Adulte/enfant

1. adulte	73.5	75.1	74.7	74.9
2. enfant	26.5	24.9	25.3	25.1

Lien avec le chef de ménage

1. chef de ménage	36.4	37.2	36.6	37.8
2. épouse	24.2	24.1	23.9	23.8
3. "épouse"	1.1	1.0	0.9	1.1
4. ami, amie	0.3	0.1	0.2	0.1
5. fils, fille	33.4	33.6	34.2	33.5
6. autres	4.6	4.0	4.2	3.7

Caractéristiques du ménage

Type d'habitation

1. maison rurale	11.4	10.7	9.3	7.5
2. unifamiliale individuelle	33.7	34.7	34.7	33.5
3. unifamiliale jumelée	13.5	8.8	9.1	10.3
4. unifamiliale en série	19.5	21.4	25.3	26.5
5. petit collectif	12.6	13.1	13.1	12.4
6. collectif moyen	7.1	7.3	5.8	6.2
7. grand collectif	1.1	1.4	1.1	0.8
8. précaire	0.3	0.0	0.1	0.1
9. N.R.	0.8	2.6	1.7	-
0. maison de retraite	-	-	-	0.3

Canton

1. Luxembourg-ville	17.6	17.6	17.7	16.2
2. Luxembourg-campagne	8.9	8.8	8.8	9.3
3. Capellen	6.9	7.0	6.9	7.1
4. Esch sur Alzette	33.5	33.6	34.3	33.7
5. Mersch	4.6	4.3	3.7	4.2
6. Clervaux	2.3	2.3	2.5	2.5
7. Diekirch	6.2	6.2	6.7	6.7
8. Redange	3.0	3.0	2.8	3.0
9. Vianden	0.8	0.7	0.8	0.8
10. Wiltz	3.0	3.1	3.0	3.0
11. Echternach	3.5	3.5	3.5	3.4
12. Grevenmacher	5.8	5.8	5.7	6.3
13. Remich	4.0	4.0	3.6	3.9

Nombre de personnes dans le ménage

1.	7.6	8.0	8.0	8.5
2.	21.0	20.6	19.6	21.1
3.	22.5	23.1	22.8	22.1
4.	27.6	27.6	28.7	28.6
5.	12.3	12.7	12.9	12.7
6.	5.8	5.3	4.9	5.0
7.	2.1	1.8	2.1	1.2
8.	0.7	0.6	0.7	0.7
9.	0.3	0.3	0.2	-
11.	0.0	-	0.1	-
12.	-	0.1	-	0.1

Nombre d'enfants dans le ménage

0.	39.2	41.9	41.7	43.4
1.	24.3	23.2	23.6	21.2
2.	22.9	23.2	22.8	24.1
3.	9.9	9.5	9.1	8.7
4.	3.5	2.2	2.7	2.5
5.	0.1	0.1	0.1	0.1
6.	0.1	-	-	-

Nombre d'adultes dans le ménage

0.	0.1	-	-	-
1.	9.9	10.2	10.4	11.0
2.	63.0	59.5	58.4	60.3
3.	16.2	18.9	18.9	17.4
4.	7.7	6.9	7.8	7.8
5.	2.1	3.0	2.8	2.4
6.	0.8	1.2	1.4	0.8
7.	0.2	0.2	0.2	0.3
9.	-	0.1	0.1	-

Nombre d'emplois dans le ménage

0.	17.1	17.5	17.4	17.3
1.	46.9	43.1	42.1	43.4
2.	29.2	29.8	30.0	28.9
3.	4.9	6.8	6.9	7.1
4.	1.4	2.1	2.5	2.6
5.	0.5	0.7	0.9	0.6
6.	0.0	0.0	0.1	0.1

Caractéristiques du chef de ménage

Sexe

1. homme	87.2	86.0	86.3	83.6
2. femme	12.8	14.0	13.7	16.4

Age

1. < 24 ans	2.6	2.3	1.8	2.1
2. 25 - 34 ans	18.3	18.1	18.7	17.7
3. 35 - 44 ans	27.6	27.1	27.3	26.6
4. 45 - 54 ans	24.4	23.7	23.0	23.7
5. 55 - 64 ans	14.8	15.9	15.8	16.9
6. > 64 ans	12.2	12.9	13.4	13.1

Etat civil

1. célibataire	6.0	6.1	6.5	6.2
2. marié	80.6	80.0	79.6	79.2
3. séparé	1.4	1.7	1.2	1.3
4. divorcé	2.5	2.7	2.7	3.0
5. veuf	9.5	9.6	10.0	10.3
6. N.R.	0.0	-	-	-

Emploi

1. oui	74.1	73.3	73.3	72.0
2. non	25.9	26.7	26.7	28.0

Etat civil par Sexe

1. homme non-marié	7.0	6.7	7.0	6.2
2. homme marié	80.2	79.3	79.3	77.4
3. femme non-mariée	12.4	13.4	13.4	14.6
4. femme mariée	0.4	0.6	0.3	1.8

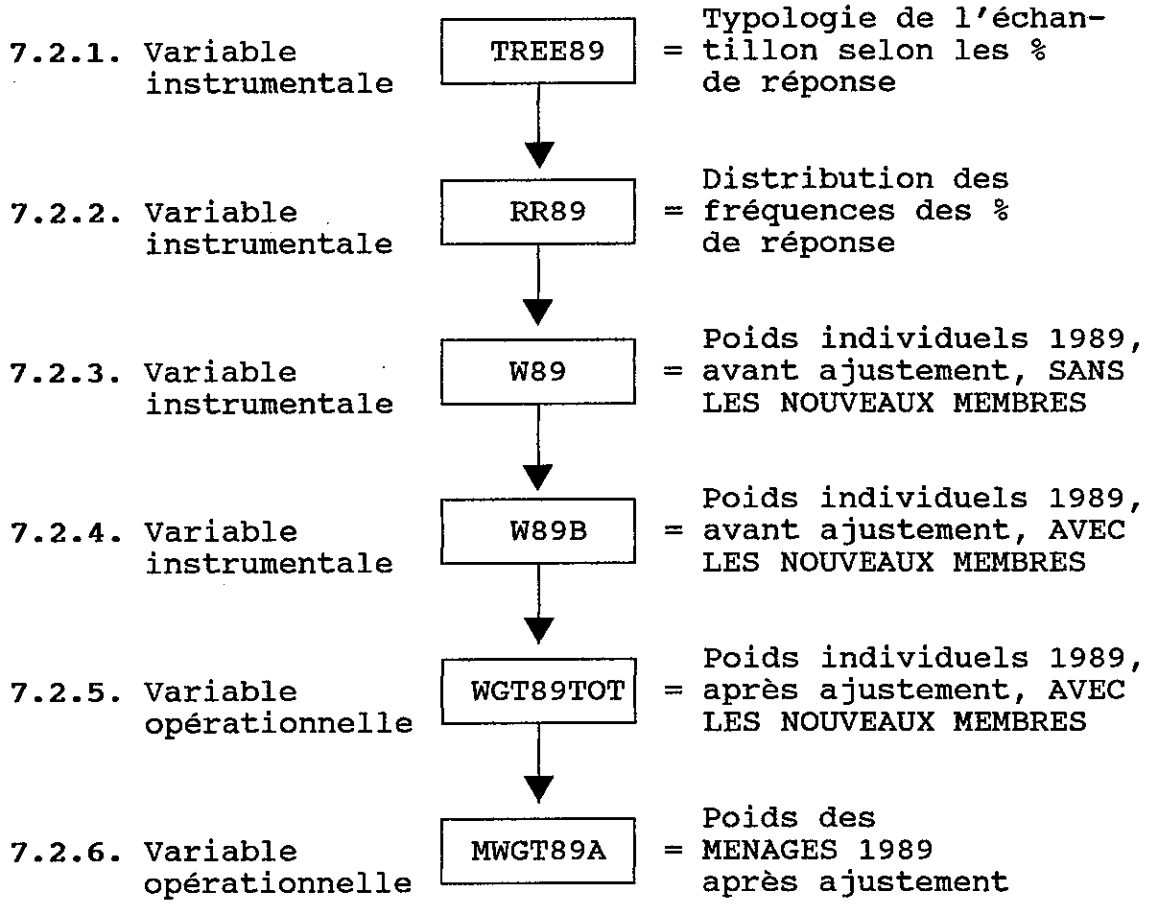
Emploi par Sexe

1. homme sans emploi	17.5	17.4	17.6	19.0
2. homme avec emploi	69.8	68.6	68.7	64.6
3. femme sans emploi	8.5	9.3	9.1	9.0
4. femme avec emploi	4.3	4.7	4.6	7.4

ÉCHANTILLON 1989

ET VARIABLES DE PONDÉRATION

Encadré 7.1. **VARIABLES DE PONDERATION 1989**



Chapitre 7

ÉCHANTILLON 1989

ET VARIABLES DE PONDÉRATION

L'échantillon observé en 1989 n'est plus conforme à l'échantillon observé en 1985. Phénomènes démographiques et phénomènes propres à la vie de l'échantillon conjuguent leurs effets et provoquent une dérive progressive de l'échantillon.

Les trajectoires des personnes dans l'étude se diversifient de plus en plus. Ces trajectoires déterminent la position des personnes dans la procédure de pondération.

La première partie du chapitre sera consacrée à la description de cette évolution de l'échantillon.

La seconde partie présentera les différentes variables qui ont jalonné l'élaboration du nouveau système de pondérations.

Pour chaque variable, un tableau récapitule les sous-échantillons pertinents et la taille de ces sous-échantillons.

7.1. L'échantillon 1989

7.1.1. Les trajectoires individuelles: version courte

TABLEAU 7.1.
Trajectoires individuelles 1985 - 1989

VARIABLE DE SÉLECTION: DIA89	n =
0. memb. panel refus	1771
1. memb. panel réponse	4220
2. nouv. memb. panel 88	71
3. non-memb. panel réponse	464
4. non-memb. panel refus	85
5. émigrés / décédés	343
FICHIER TOTAL	6954

Le nouveau fichier contient désormais 6954 personnes classées en 6 catégories principales. Le contenu de ces catégories évolue d'année en année.

Le nouvel échantillon compte 4755 personnes réparties en 3 classes:

- les membres du panel ayant répondu (4220),
- les nouveaux membres du panel entrés dans l'échantillon en 1989 (71)
- et les membres de l'échantillon qui ne sont pas membres du panel (464).

Le taux de réponses est calculé en rapportant le nombre des réponses des *membres du panel* (4220) au nombre total des membres du panel (5991, soit: 1771 + 4220).

Les nouveaux membres du panel entrés dans l'échantillon en 1989 (71) et les personnes émigrées ou décédées ne sont pas prises en compte dans le calcul du taux de réponse.

Ces 6 classes sont définies en fonction du point initial et du point final des trajectoires des individus dans l'ensemble de l'étude. Le point initial détermine le statut de la personne au moment ou elle est entrée dans l'étude: membre ou non-membre du panel. Le point final correspond à la présence ou à l'absence de la personne dans l'échantillon considéré.

Ceci signifie, par exemple, que les membres du panel

* entrés dans l'échantillon initial en 1985

* ayant refusé de répondre en 1986, en 1987 et en 1988

* peuvent être classés, en 1989

- soit parmi les refus (groupe #0)

- soit parmi les réponses (groupe #1) s'ils ont accepté de renouveler leur coopération, s'ils ont rejoint un ménage d'origine toujours présent dans l'échantillon, etc ...

L'ensemble de toutes ces trajectoires est présenté au tableau **7.2.**. Chaque trajectoire est définie par la succession des états occupés par les personnes au cours des années écoulées.

Le tableau **7.1.** rassemble ces trajectoires en 6 classes principales parce que ces classes

* sont nécessaires et suffisantes pour permettre le calcul des nouvelles pondérations

* permettent de gérer facilement les sous-échantillons principaux qui forment l'ensemble du fichier individuel.

Le tableau **7.2.** permet de recomposer d'autres échantillons en fonction des besoins de l'analyste.

La variable DIA89 comporte six grandes modalités qui peuvent être décrites de la manière suivante.

7.1.1.1. *Les refus des membres du panel*

La première catégorie est formée par les membres du panel ayant refusé de prendre part à l'enquête. Cet ensemble ne correspond plus à la première catégorie de la variable DIA utilisée en 1986 ni à la variable DIA88 utilisée en 1988. Elle réunit des membres du panel présents en 1985 ET des nouveaux membres entrés en 1986, en 1987 ou en 1988.

Tous ces membres ont un point commun: **ils n'ont pas répondu en 1989.**

7.1.1.2. *Les membres du panel*

Le groupe #1 réunit:

- des membres du panel présents en 1985,
- des nouveaux membres entrés en 1986,
- des nouveaux membres entrés en 1987,
- des nouveaux membres entrés en 1988.

Ils ont un point commun: **ils ont répondu en 1989.**

Tous ces membres sont réunis dans cette catégorie, même s'ils n'ont pas répondu au cours de la seconde, de la troisième ou de la quatrième vague d'enquêtes.

7.1.1.3. *Les nouveaux membres du panel*

Le groupe #2 isole les nouveaux membres du panel entrés dans l'échantillon en 1989. Ce sont des nouveaux-nés descendants directs de membres du panel.

Ce groupe ne contient pas les nouveaux membres du panel entrés entre 1986 et 1988. Ces derniers font désormais partie de l'ensemble des membres du panel. Ils sont traités à ce titre dans la procédure de pondération. Ils sont répartis entre le groupe #0, le groupe #1 et le groupe #5 en fonction de la trajectoire qu'ils ont suivie entre 1986 et 1989.

Ce groupe #2 permet d'identifier les nouveaux membres de l'échantillon annuel qui recevront un poids individuel lié à celui de leurs ascendants. Les membres de ce groupe sont donc différents chaque année.

7.1.1.4. *Les non-membres du panel*

Les groupes #3 et #4 sont formés par l'ensemble des nouvelles personnes entrées dans les échantillons successifs sans accéder au titre de membres du panel.

Ces personnes sont classées dans le groupe #3 lorsqu'elles appartiennent à l'échantillon annuel. Les nouveaux non-membres du panel entrés dans l'échantillon en 1989 sont donc classés dans ce groupe au même titre que les non-membres du panel entrés dans l'étude au cours de vagues antérieures. Ils ont un point commun: **ils ont répondu en 1989.**

Les non-membres du panel, entrés dans les échantillons précédents, sont classés dans le groupe #4 **lorsqu'ils n'ont pas participé à l'enquête en 1989.**

Cette distinction entre les groupes #3 et #4 permet de reconstituer aisément l'échantillon annuel, lorsqu'il y a lieu de prendre en compte les membres du panel ET les non-membres du panel présents dans l'échantillon: il suffit de sélectionner les groupes #1, #2 et #3.

Cette distinction est également très utile au moment de regrouper les individus appartenant à un même ménage ou à un même groupe de revenu.

Elle est indispensable au moment de calculer le poids individuel des nouveaux membres du panel: par exemple, le poids du nouveau membre est différent selon que l'un des ascendants directs est absent de l'échantillon ou non-membre du panel (**7.2.4.**).

7.1.1.5. *Les émigrés et les décédés*

Le groupe #5 réunit toutes les personnes qui ont été présentes dans un échantillon au moins une fois, et qui ne sont pas présentes dans l'échantillon de l'année considérée pour une raison d'ordre *démographique*.

Cette distinction s'impose puisque les émigrés et les personnes décédées ne sont pas prises en compte dans le système de pondération annuel.

Leur classement dans une catégorie particulière facilite également la sélection de l'échantillon annuel et la reconstitution des ménages.

Il faut rappeler qu'une personne émigrée peut rejoindre l'échantillon ultérieurement. Elle sera reclassée dans son groupe d'origine, c'est à dire dans le groupe #1 si elle était membre du panel ou dans le groupe #3 si elle n'était pas membre du panel.

7.1.2. Les trajectoires individuelles: version complète

TABLEAU 7.2.
Liste des trajectoires individuelles: 1985-1989

VARIABLE: DIA4		
CODES	TRAJECTOIRES	n =
10.00	FM/85/0/0/0/0	825
11.00	FM/85/0/0/0/89	1
20.00	FM/85/86/0/0/0	470
21.00	FM/85/86/0/0/89	5
30.00	FM/0/86/0/0/0	3
40.00	FM/85/0/87/0/0	1
50.00	FM/85/86/87/0/0	238
51.00	FM/85/86/87/0/89	31
60.00	FM/0/86/87/0/0	4
70.00	FM/85/E/87/0/0	4
80.00	FM/0/0/87/0/0	2
81.00	FM/0/0/87/0/0/89	2
101.00	FM/85/0/0/88/89	2
110.00	FM/85/86/0/88/0	21
111.00	FM/85/86/0/88/89	99
113.00	FM/85/86/0/88/D	1
121.00	FM/0/86/0/88/89	2
131.00	FM/85/0/87/88/89	4
140.00	FM/85/86/87/88/0	193
141.00	FM/85/86/87/88/89	3891
142.00	FM/85/86/87/88/E	33
143.00	FM/85/86/87/88/D	37
150.00	FM/0/86/87/88/0	3
151.00	FM/0/86/87/88/89	64
152.00	FM/0/86/87/88/E	1
160.00	FM/0/0/87/88/0	3
161.00	FM/0/0/87/88/89	57
171.00	FM/85/86/E/88/89	2
180.00	FM/0/0/0/88/0	4
181.00	FM/0/0/0/88/89	59
182.00	FM/0/0/0/88/E	1
183.00	FM/0/0/0/0/89	71

VARIABLE: DIA4 (suite)		
CODES	TRAJECTOIRES	n =
200.00	NF/0/86/0/0/0	17
201.00	NF/0/86/0/0/89	1
210.00	NF/0/86/87/0/0	13
220.00	NF/0/0/87/0/0	22
300.00	NF/0/86/0/88/0	1
301.00	NF/0/86/0/88/89	1
310.00	NF/0/86/87/88/0	4
311.00	NF/0/86/87/88/89	103
313.00	NF/0/86/87/88/D	2
320.00	NF/0/0/87/88/0	8
321.00	NF/0/0/87/88/89	106
322.00	NF/0/0/87/88/E	3
323.00	NF/0/0/87/88/D	1
330.00	NF/0/0/0/88/0	20
331.00	NF/0/0/0/88/89	135
332.00	NF/0/0/0/88/E	3
333.00	NF/0/0/0/88/D	1
334.00	NF/0/0/0/0/89	118
400.00	FM/85/E/-/-/-	56
410.00	FM/85/86/E/-/-	19
411.00	FM/85/86/E/-/89	1
420.00	FM/85/86/87/E/-	25
430.00	FM/0/0/87/E/-	1
500.00	FM/85/D/-/-/-	50
510.00	FM/85/86/D/-/-	44
520.00	FM/85/86/0/D/-	3
530.00	FM/85/86/87/D/-	54
540.00	FM/0/86/D/-/-	1
600.00	NF/0/86/D/-/-	3
610.00	NF/0/86/87/D/-	1
620.00	NF/0/0/87/D/-	3
FICHIER TOTAL		6954

Le tableau **7.2.** est composé de la manière suivante.

La variable est nommée **DIA5** parce qu'elle correspond à l'état du fichier après la cinquième vague d'enquêtes.

Les 6954 personnes recensées dans le fichier total après cinq vagues d'enquêtes ont suivi 63 trajectoires différentes entre 1985 et 1989 (contre 36, entre 1985 et 1988).

Chaque trajectoire est un vecteur défini par une série d'états successifs. Chaque état correspond au statut occupé par une personne dans le panel ET dans l'échantillon suite à une vague d'observation annuelle.

Chaque vecteur est désigné par un code. Ces codes ont été attribués de manière arbitraire et peuvent être modifiés.

Les vecteurs 11, 21, 51 et 81 désignent des membres du panel absents pendant au moins une année et revenus dans l'échantillon en 1989 (soit, 39 personnes).

Les vecteurs sont décrits de la manière suivante:

* **FM** signifie que ce vecteur contient exclusivement des **membres du panel** (Full Member).

  NF signifie que le vecteur contient exclusivement des personnes **non membres du panel** (Non Full member).

* **85** signifie que ces personnes sont **membres de l'échantillon** observé en 1985.

  86 signifie que ces personnes sont membres de l'échantillon observé en 1986.

  L'année est remplacée par le code **0** lorsqu'une personne est **absente de l'échantillon** suite à un refus ou lorsqu'il n'est pas avéré que cette absence est liée à une raison d'ordre démographique.

  Par exemple: **85/86/0/0/0** désigne le vecteur des personnes

  - entrées dans l'échantillon au cours de la première vague d'enquête effectuée en 1985
  - toujours présentes dans l'échantillon observé en 1986
  - mais absentes de l'échantillon en 1987, en 1988 et en 1989 pour des raisons assimilées à des refus.

* **E** signifie que ces personnes ont **émigré** et sont absentes de l'échantillon correspondant à la vague d'enquête de cette année. Dans certains cas cette absence est provisoire.

Le vecteur #70 est composé de la manière suivante:

- **FM. 85/E/87/0/0** soit 4 membres du panel, entrés dans l'échantillon en 1985, émigrés en 1986, revenus dans l'échantillon en 1987 et refusant de répondre en 1988 et en 1989.

* **D** signifie que ces personnes sont **décédées.**

Le vecteur #620 est composé de la manière suivante:

- **NF/0/0/87/D/-** soit 3 personnes entrées dans l'étude en 1987 sans accéder au titre de membres du panel et décédées au cours de l'année 1988.

Le passage de la variable DIA5 à la variable DIA89 s'opère aisément. Les classes générales de trajectoires sont composées à partir du point initial et du point final de chaque vecteur.

Le **groupe #1** de la variable DIA89 regroupe l'ensemble des membres du panel entrés dans l'échantillon en 1985, en 1986, en 1987 ou en 1988 et toujours présents dans l'échantillon en 1989, quelles que soient les étapes intermédiaires ayant jalonné leur trajectoire. Cette classe est donc formée par les vecteurs #11, #21, #51, #81, #101, #111, #121, #131, #141, #151, #161, #171, #181 de la variable DIA5.

Le vecteur #141 regroupe l'ensemble des membres du panel présents tous les ans dans l'échantillon. Ce groupe est important lorsque l'analyste souhaite analyser uniquement les personnes qui ont participé à l'enquête chaque année.

Le regroupement du vecteur #183 et de tous les vecteurs concernant des membres du panel dont le code se termine par "1" équivaut au regroupement des classes #1 et #2 de la variable DIA89: il désigne l'ensemble des membres du panel présents dans l'échantillon en 1989 soit 4291 personnes (dont 71 nouveaux membres entrés dans l'échantillon en 1989).

7.2. Les variables de pondération

7.2.1. Variable instrumentale TREE89

Les refus de répondre constituent le facteur principal de la déformation de l'échantillon (**2.2.1.1.**).

1. La notion de "taux de réponses" a pris désormais un sens particulier. Elle correspond davantage à la notion de "taux de survie dans le panel" (**7.2.1.1**).

2. Les refus de répondre seraient sans effet s'ils étaient répartis aléatoirement dans l'échantillon. Il convient donc de rechercher les concentrations éventuelles des refus dans des groupes démographiques homogènes afin d'en corriger les effets (**7.2.1.3.**).

3. Au préalable, l'inégalité des probabilités de sélection initiales des membres du panel doit être annulée en pondérant les individus par l'inverse de cette probabilité. La recherche des groupes démographiques atteints par les taux de refus les plus contrastés ne doit pas être influencée par ce biais lié au mode de tirage de l'échantillon initial (**7.2.1.2.**).

7.2.1.1. *Des taux de survie*

Il est important de rappeler que la notion de taux de réponses est entendue désormais dans un sens particulier.

Le terme "taux de survie" est plus correct que le terme "taux de réponses". En effet, le **taux de survie** prend en compte le **cumul** des refus enregistrés au cours des enquêtes successives à condition que ces personnes n'aient pas rejoint l'échantillon en 1989.

Le taux de réponses correspond, en fait, au nombre des réponses enregistrées au cours d'une vague d'enquête, rapporté au nombre cumulé des membres du panel.

> *Il n'est pas calculé par référence au nombre des membres du panel présents dans l'échantillon précédent. Il est calculé en rapportant le nombre de réponses des membres du panel au nombre total des membres du panel entrés dans les échantillons successifs depuis la première vague d'enquête.*

Cette procédure permet de prendre en compte la mobilité des membres du panel: ils peuvent entrer et sortir de l'échantillon. Chaque fois qu'ils rejoignent l'échantillon, leur

nouvelle probabilité de sélection peut être calculée puisque les deux éléments essentiels du calcul sont connus:

* leur probabilité de sélection initiale (7.2.1.2.)

* et les caractéristiques démographiques du sous-groupe auquel ils appartenaient au moment ou ils sont entrés dans l'échantillon (7.2.1.3.).

7.2.1.2. *Correction du biais initial*

L'analyse des taux de réponses prend en compte les 5991 personnes regroupées dans les classes #0 et #1 de la variable DIA89. Elle ne prend en compte, ni les nouveaux membres du panel entrés dans l'échantillon au cours de la dernière vague d'enquête, ni les non-membres du panel, ni les personnes émigrées ou décédées.

Parmi les 5991 membres du panel, entrés dans l'échantillon de 1985 à 1988, 4220 ont répondu en 1989, soit 70.4%.

Ces valeurs ne prennent pas en compte l'inégalité des probabilités de sélection initiales des individus. Il convient donc d'annuler cette inégalité avant d'entreprendre toute autre opération.

Les membres entrés dans l'échantillon en 1985 doivent être pondérés par l'inverse de leur probabilité de sélection soit, par la variable PONREV.

La probabilité de sélection des nouveaux membres correspond à la valeur moyenne des probabilités de sélection de leurs parents *au moment ou ces nouveaux membres entrent dans le panel*. Ces nouveaux membres ont déjà reçu ce poids au moment ou ils sont entrés dans l'échantillon: c'est la valeur de la variable W86B attribuée en 1986 aux nouveaux membres de l'année; c'est la valeur de W87B attribuée en 1987 aux nouveaux membres entrés en 1987; c'est, encore, la valeur de W88B attribuée en 1988 aux nouveaux membres entrés en 1988.

La synthèse de ces valeurs donne lieu à la création de la variable **IP8589**. Cette variable est sauvegardée dans le fichier global parce qu'elle sera réutilisée l'année prochaine dans le cadre de la même procédure. (Cette variable doit être ajustée en fonction de la taille de l'échantillon).

Cette première procédure modifie la valeur du taux de survie. Le tableau **7.3.** permet d'observer les glissements des proportions avant et après la correction des biais de sélection.

TABLEAU 7.3.
Distribution des fréquences des individus membres du panel, en 1989, selon leur position dans l'échantillon, avant et après pondération

DIA89	FREQUENCES			
	avant pondération		après pondération (IP8589A)	
	n	%	n	%
0. memb.pan. refus	1771	29.6	1676.	28.0
1. memb.pan. réponses	4220	70.4	4314.	72.0
TOTAL	5991	100.0	5991	100.0

Le taux de survie passe de 70.4% à 72.0% lorsque les biais de sélection initiaux sont contrôlés.

7.2.1.3. *Les nouvelles probabilités de sélection*

La qualité de l'échantillon n'est pas affectée par les refus de répondre aussi longtemps qu'ils se distribuent de manière aléatoire dans l'ensemble de l'échantillon.

La méthode permettant de repérer des concentrations éventuelles de ces refus dans certains sous-groupes homogènes de l'échantillon a été présentée dans le second chapitre (2.2.1.1.).

La démarche met en concurrence le pouvoir discriminant de 16 variables:

CARACTERISTIQUES DU MÉNAGE	DU CHEF DE MÉNAGE	DE LA PERSONNE
Canton de résidence	Age	Age
Type d'habitat	Sexe	Sexe
Taille du ménage	Etat civil	Etat civil
Nombre d'adultes	Emploi (oui/non)	Emploi
Nombre d'enfants		Adulte/Enfant
Nombre d'emplois		Lien avec le CM

Toutes ces caractéristiques sont attribuées aux personnes. Les caractéristiques du ménage et les caractéristiques du chef de ménage sont des éléments de l'environnement de la personne.

Les membres entrés dans l'échantillon en 1985 sont caractérisés par les valeurs qu'ils présentaient sur ces variables **en 1985**, au moment où ils sont entrés dans l'échantillon.

Les nouveaux membres, entrés de 1986 à 1988, sont caractérisés par les valeurs qu'ils présentaient sur ces différentes variables au moment de leur entrée dans l'échantillon. Ces variables décrivent donc le contexte du ménage dans lequel ils sont nés et *au moment où ils sont nés.*

Il est donc fort probable qu'il existe un décalage systématique entre les caractéristiques des nouveaux membres et les caractéristiques de leurs parents.

Si la taille du ménage dans lequel ils font leur entrée s'est modifiée entre 1985 et 1988, les nouveaux membres du panel entrés en 1988 ne sont pas caractérisés par la même valeur que leurs parents sur cette variable.

L'âge du chef de ménage fournit un exemple simple: en règle générale, l'âge du chef de ménage sera différent pour les parents et pour le nouveau membre du panel entré dans la famille deux ans plus tard.

La synthèse de ces valeurs donne lieu à la création de nouvelles variables. Ces variables sont sauvegardées dans le fichier global parce qu'elles seront réutilisées l'année suivante dans le cadre de la même procédure.

(La variable de correction des probabilités de sélection initiales a été actualisée selon la même logique (**7.2.1.2.**)).

Lorsque l'ensemble des variables choisies ont été actualisées, la recherche des catégories démographiques les plus discriminantes peut être entreprise (**2.2.1.1.**).

En résumé, la procédure consiste à diviser l'échantillon en sous-groupes mutuellement exclusifs par une succession de divisions binaires. L'enchaînement de ces divisions successives crée des filières ou séries de modalités. Les variables qui produisent ces divisions sont choisies parce qu'elles créent les contrastes les plus significatifs entre les taux de réponses des deux sous-groupes. Ces divisions sont interrompues dès qu'un sous-groupe ne peut plus être divisé en deux sous-groupes d'une taille supérieure ou égale à 5% de l'échantillon. (Cf. Procedures and Tape Codes, 1978 Interviewing Year, Wave XI, a Supplement. P.S.I.D., I.S.R., Ann Arbor, Michigan, 1979, p.10).

Le résultat de cette analyse est présenté dans le schéma **7.1.**. Il reproduit sous une forme arborescente la progression de l'analyse vers des sous-groupes de plus en plus homogènes.

Dix-sept sous-groupes sont identifiés à l'aide de 9 variables.

La variable dépendante est une variable binaire (0,1) qui rend compte du **taux de survie** dans le panel longitudinal:

$$\text{Taux de survie} = \frac{\text{Nombre de codes 1}}{\text{Nombre de cas dans le sous-groupe}}$$

Les 17 sous-groupes "n'expliquent" que 5.6% de la variance de la variable dépendante.

L'objectif de cette procédure n'était pas de prendre en compte toute la variance des taux de survie. Des mesures ont été prises en vue d'éviter que l'analyse identifie des groupes marginaux de très petite taille. La variance des poids des individus est maintenue dans des limites acceptables[1]

Le faible pouvoir explicatif de la typologie constituée par cette procédure provient également du fait que les modalités initiales des variables démographiques divisent l'échantillon en segments dont les moyennes sont peu contrastées (tableau **7.4.**).

1. La variance des poids des individus (ou variance de l'échantillon) a évolué de la manière suivante: 1985:'.18'; 1986:'.18'; 1987:'.19'; 1988: '.20'; 1989: '.18'.

Analyse des taux de réponses des sous-groupes entre 1985 et 1989

(après pondération initiale)

SCHEMA 7.1.

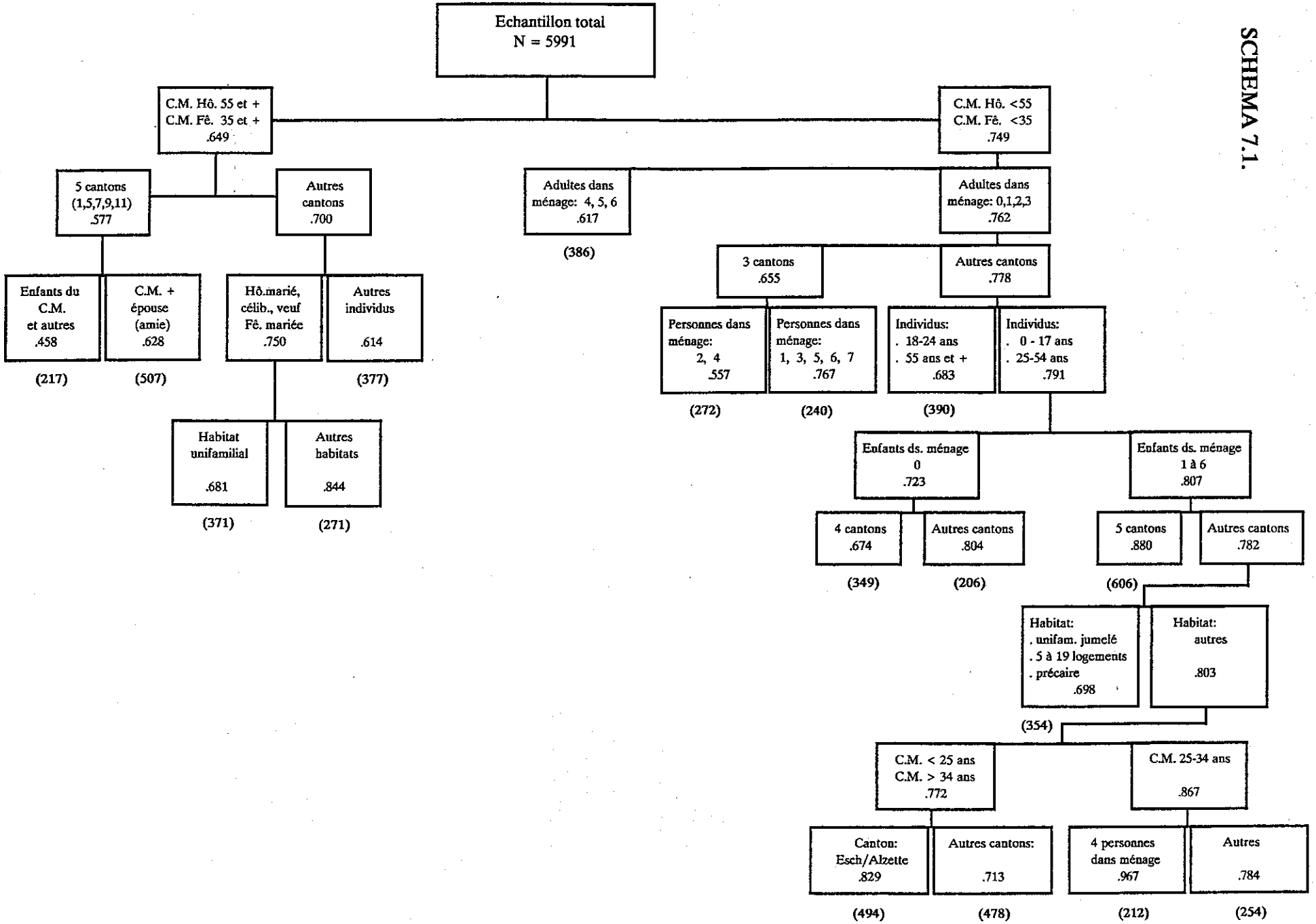


TABLEAU 7.4.

Taux de survie des membres du panel en 1989

CARACTÉRISTIQUES DES SOUS-GROUPES	TAUX DE REPONSE $\bar{x}$ = .720	VARIANCE "EXPLIQUÉE" (ETA)	VARIANCE EXPLIQUÉE PAR LA MEILLEURE DICHOTOMISATION
Caractéristiques individuelles			
Sexe			
1. masculin	.725		
2. féminin	.715	.01	.01
Etat civil			
1. célibataire	.723		
2. marié	.736		
3. séparé	.660	.07	.06
4. divorcé	.625		
5. veuf	.622		
6. D.M.	(.000)*		
Age			
1. < 17 ans	.776		
2. 18 à 24 ans	.640		
3. 25 à 34 ans	.734		
4. 35 à 44 ans	.724	.11	.10
5. 45 à 54 ans	.722		
6. 55 à 64 ans	.718		
7. 65 ans et plus	.620		
Nationalité			
1. Luxembourgeois	.717		
2. CEE	.741		
3. hors CEE	.643	.05	.02
4. D.M.	(.082)		

* Les chiffres entre parenthèses correspondent à des catégories de moins de 20 personnes.

Situation professionnelle

1. Enfants	.762		
2. Travail ou maladie	.709		
3. Demandeur d'emploi	.590	.07	.04
4. Pensionné, invalide	.695		
5. Retraité	.669		
6. Ménage, milice	.707		

Lien avec le chef de ménage

1. Chef de ménage	.714		
2. épouse	.740		
3. "épouse"	.636	.06	.06
4. ami(e)	(.608)		
5. fils, fille	.730		
6. autre	.594		

Adultes

1. Enfant	.762		
2. Adulte	.701	.06	.06

Etat civil selon le sexe

1. Homme célibataire	.719		
2. marié	.737		
3. séparé	.596		
4. divorcé	.673		
5. veuf	.686		
6. M.D.	(.000)*	.07	.07
7. Femme célibataire	.729		
8. mariée	.735		
9. séparée	.723		
10. divorcée	.582		
11. veuve	.611		

Caractéristiques du ménage

Nombre de personnes dans le ménage

1.	.658		
2.	.714		
3.	.735		
4.	.705		
5.	.771	.07	.05
6.	.706		
7.	.674		
8.	.835		
9.	(.894)		
11.	(.836)		
12.	(.000)		

Nombre d'enfants dans le ménage

0.	.677		
1.	.734		
2.	.741		
3.	.750	.08	.07
4.	.819		
5.	(1.000)		
6.	(.777)		

Nombre d'adultes dans le ménage

0.	(.000)		
1.	.675		
2.	.749		
3.	.683	.10	.08
4.	.608		
5.	.710		
6.	.625		
7.	(.921)		
9.	(.000)		

Nombre d'emplois dans le ménage

0.	.695		
1.	.726		
2.	.743		
3.	.607	.07	.05
4.	.656		
5.	.805		
6.	(.000)		

Type d'habitation

1. maison rurale	.720		
2. maison unifamiliale individuelle	.719		
3. -"- jumelée	.702		
4. -"- en série	.722		
5. petit collectif	.730		
6. collectif moyen	.716	.02	.01
7. grand collectif	.812		
8. précaire	(.714)		
9. Non réponse	(.746)		

Canton de résidence

1. Luxembourg-Ville	.673		
2. Luxembourg-Campagne	.741		
3. Capellen	.649		
4. Esch-sur-Alzette	.741		
5. Mersch	.745		
6. Clervaux	.839	.09	.08
7. Diekirch	.721		
8. Rédange	.649		
9. Vianden	.812		
10 Wiltz	.750		
11 Echternach	.659		
12 Grevenmacher	.725		
13 Remich	.788		

Caractéristiques du chef de ménage

Age du chef de ménage

1. < 24 ans	.712		
2. 25 à 34 ans	.793		
3. 35 à 44 ans	.731		
4. 45 à 54 ans	.715	.10	.08
5. 55 à 64 ans	.673		
6. 65 et plus	.626		

Sexe du chef de ménage

1. masculin	.729		
2. féminin	.652	.05	.05

Etat civil du chef de ménage

1. Célibataire	.711		
2. marié	.733		
3. séparé	.632		
4. divorcé	.622	.07	.06
5. veuf	.653		
6. Non réponse	(.000)		

Situation professionnelle du C.M.

1. sans emploi	.672		
2. travaille	.735	.06	.06

Etat civil selon le sexe du C.M.

1. Homme	non marié	.674	
2.	marié	.734	
3. Femme	non mariée	.657	.06	.06
4.	mariée	.515	

Situation professionnelle du C.M. selon le sexe

1.	Homme	sans emploi	.686		
2.		au travail	.739		
3.	Femme	sans emploi	.643	.07	.06
4.		au travail	.668		

Age du C.M. selon le sexe

1.	Homme	< 35 ans	.777		
2.		35 à 44 ans	.742		
3.		45 à 54 ans	.721		
4.		55 à 64 ans	.686		
5.		65 ans et +	.631	.12	.10
6.	Femme	< 35 ans	.842		
7.		35 à 44 ans	.579		
8.		45 à 54 ans	.656		
9.		55 à 64 ans	.602		
10.		65 ans et +	.619		

La lecture détaillée de l'arbre (schéma 7.1) n'apporte pas d'informations très claires.

Trois points peuvent être mis en évidence.

1. Un contraste très prononcé entre la métropole et les autres cantons du pays avait été observé au terme de la deuxième vague d'enquête. Il s'est effacé.

2. Une caractéristique propre aux individus avait son apparition au terme de la troisième vague d'enquête (l'âge des individus). Cette caractéristique intervient à nouveau au terme de la quatrième vague d'enquête et au terme de cette cinquième vague. En outre, deux nouvelles caractéristiques propres aux individus apparaissent:
   - le lien avec le chef de ménage
   - l'état civil selon le sexe.

   Les variables décrivant le cadre familial ne monopolisent plus l'organisation des taux de réponse.

   Les panels socio-économiques devraient éviter de procéder à des pondérations directes des ménages sans prendre en compte l'hétérogénéité interne de cette unité d'analyse: au fil des années les individus affirment leur autonomie par rapport à "l'unité" -d'analyse- ménage. Leur destin ne se résoud pas dans celui du ménage.

3. L'écart entre le taux de survie maximum et le taux de survie minimum a connu une croissance notable entre la deuxième et la troisième vague,

   - il est resté constant en 1987 et en 1988
   - il s'accroît de nouveau de 1988 à 1989.

	1986	1987	1988	1989
Taux le plus élevé	1.00	.98	.94	.97
Taux le plus bas	.71	.54	.48	.46
Ecart	.29	.44	.46	.51

TABLEAU 7.5.
Ventilation des valeurs de TREE89
dans le fichier individuel

DIA89	TREE89 n =
0. memb. panel refus	1771
1. memb. panel réponse	4220
2. nouv. memb. panel 89	-
3. non-memb. panel réponse	-
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	5991
sysmis	963
FICHIER TOTAL	6954

Cette variable TREE89 décrit le profil des catégories d'appartenance des membres du panel, qu'ils aient répondu ou non en 1989. Elle permet donc de classer tous les membres des classes #0 et #1 de la variable DIA89 (tableau 7.5.).

TREE89 n'exprime pas des probabilités de survie. Elle décrit simplement une typologie de sous-échantillons. Elle ne permet donc pas de calculer le poids relatif des personnes dans le nouvel échantillon. Ce sera le rôle d'une autre variable: RR89.

7.2.2. Variable instrumentale RR89

La variable RR89 est construite de la manière suivante: toute personne **ayant effectivement répondu** en 1989 (DIA89 #1), reçoit la probabilité de sélection observée dans la catégorie de population à laquelle elle appartient (TREE89) (tableau 7.6.).

TABLEAU 7.6.

Distribution des fréquences de l'échantillon 1989, selon les taux de réponses observés en 1989 (avant et après pondération de l'échantillon initial)

RR89 taux de réponses en 1989	TREE89 code	FREQUENCES avant pondération n =	FREQUENCES après pondération (IP8589A) n =
.46	(17)	148	100
.56	(16)	121	152
.61	(15)	279	232
.62	(14)	510	238
.63	(13)	297	319
.67	(12)	218	236
.68	(11)	267	253
.68	(10)	255	267
.70	(9)	204	248
.71	(8)	286	341
.77	(7)	159	184
.78	(6)	150	199
.80	(5)	134	166
.83	(4)	328	410
.84	(3)	282	230
.88	(2)	426	534
.97	(1)	156	205
TOTAL REPONSES		4220	4314

Il n'y a pas lieu de s'étonner si la taille de l'échantillon n'est plus identique avant et après le redressement de l'échantillon. Seule une fraction de l'échantillon des membres du panel a répondu en 1989: 4220 personnes sur 5991.

Ces 4220 personnes ont été pondérées (par IP8589A) dans le cadre de **l'échantillon total,** afin de compenser les variations des probabilités de sélection initiales. Ces poids ont été ajustés: la somme des poids est égale au nombre des individus, soit 5991. Ils n'ont pas été ajustés dans le cadre des 4220 personnes ayant effectivement participé à l'enquête (tableau 7.3.).

La différence entre les tailles de la **fraction de l'échantillon** avant et après pondération est donc parfaitement normale. En outre, la comparaison des deux dernières colonnes du tableau **7.6.** permet de constater le fait que certaines catégories de la population auraient été sur-représentées ou sous-représentées si l'échantillon n'avait pas été pondéré avant d'être soumis à l'analyse des taux de réponses.

TABLEAU 7.7.
Ventilation des valeurs de RR89
dans le fichier individuel

DIA89	RR89 n =
0. memb. panel refus	-
1. memb. panel réponse	4220
2. nouv. memb. panel 89	-
3. non-memb. panel réponse	-
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4220
sysmis	2734
TOTAL	6954

Les valeurs de RR89 ne sont donc attribuées qu'aux seuls membres du panel ayant répondu en 1989 (tableau 7.7.).

7.2.3. Variable instrumentale W89

La variable W89 définit le poids des membres du panel ayant répondu en 1989, **exception faite des nouveaux membres du panel entrés dans l'échantillon en 1989.**

Ces anciens membres appartiennent à différentes catégories de la population et se distinguent par la probabilité qu'ils répondent en 1989.

Lorsque le groupe enregistre un taux de réponses très faible, chaque "survivant" doit représenter un plus grand nombre de personnes présentant les mêmes caractéristiques démographiques. Lorsque le taux de réponses s'élève, l'importance de cet ajustement se réduit.

D'une manière générale, cet ajustement des poids des individus sera inversement proportionnel à leur probabilité de sélection c'est à dire, inversement proportionnel au taux de réponses enregistré dans leur groupe:

Soit: 1 / RR89

Dans le cas présent, ce calcul doit prendre en compte un élément supplémentaire: l'inégalité des probabilités de sélection initiales des membres du panel. Les nouveaux poids ne s'appliquent pas à des personnes "unités" mais à des fractions ou à des multiples de l'unité définis par la variable IP8589.

La nouvelle variable de pondération se calcule donc de la manière suivante:

$$W89 = IP8589 \times (1 / RR89)$$

TABLEAU 7.8.
Ventilation des valeurs de W89 dans le fichier individuel

DIA89	W89 n =
0. memb. panel refus	-
1. memb. panel réponse	4220
2. nouv. memb. panel 89	-
3. non-memb. panel réponse	-
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4220
sysmis	2734
FICHIER TOTAL	6954

Les valeurs de cette variable sont attribuées exclusivement aux MEMBRES DU PANEL présents dans l'échantillon en 1989, exception faite des nouveaux membres entrés dans l'échantillon en 1989 (tableau 7.8.). Le système de pondération doit donc être élargi à ces nouveaux membres.

7.2.4. Variable instrumentale W89B

La variable W89 ne permet pas encore de pondérer tout l'échantillon observé en 1989, parce qu'elle ne prend pas en compte les nouveaux membres de l'échantillon (membres ou non du panel). Il est indispensable d'étendre la portée de cette variable pour pouvoir calculer les poids des ménages.

Dans un premier temps, **les nouveaux membres de l'échantillon, non membres du panel**, reçoivent un poids nul ('0') puisqu'il est impossible de calculer leur probabilité de sélection initiale.

En effet:

- ils étaient déjà présents dans la population de 1985, lorsque l'échantillon initial a été sélectionné,

- ils sont entrés dans l'échantillon par la suite et pour des raisons variées: il est donc difficile de reconstituer l'ensemble des conditions de leur entrée dans l'échantillon et d'en dériver une logique de pondération[2].

Il est indispensable de pondérer ces non-membres avant d'attribuer des poids aux **nouveaux membres de l'échantillon, membres du panel.** En effet, certains nouveaux membres de l'échantillon, **non-membres du panel** peuvent être des ascendants directs **d'enfants nouveaux membres du panel.**

Or, dans un second temps, ces nouveaux membres du panel reçoivent un **poids égal à la moyenne des poids des ascendants directs** tels qu'ils ont été calculés au moment où le nouveau membre est entré dans l'échantillon (et dans la population).

Donc:

- lorsque les ascendants directs ont des poids différents de '0', le nouveau membre reçoit un poids égal à la moyenne de ces deux poids;

- lorsque l'un des ascendants directs a un poids nul, le nouveau membre du panel reçoit logiquement un poids égal à la moitié du poids de l'ascendant direct, membre du panel;

- lorsque les deux ascendants directs ne sont pas membres du panel, le nouveau membre reçoit un poids nul: il n'est pas membre du panel.

Par contre, un nouveau membre reçoit le poids de l'ascendant direct présent dans l'échantillon observé, lorsque l'autre ascendant est ABSENT de l'échantillon au moment où le nouveau entre dans l'échantillon: la probabilité de sélection initiale de ce nouveau membre dépend entièrement de la probabilité de sélection du seul ascendant présent dans l'échantillon. (Cfr. **2.2.1.5.** et **2.2.1.6.**).

2. Le poids relatif de ces individus est certainement différent de '0' mais toute affectation d'un poids différent de '0' reposerait sur un ensemble d'hypothèses comportant des erreurs qu'il est impossible d'estimer.

TABLEAU 7.9.
Ventilation des valeurs de W89B
dans le fichier individuel

DIA89	W89B n =
0. memb. panel refus	-
1. memb. panel réponse	4220
2. nouv. memb. panel 89	71
3. non-memb. panel réponse	464
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4755
sysmis	2199
TOTAL	6954

Ces poids sont attribués à tous les membres de l'échantillon observé en 1989, soit 4755 personnes. Parmi celles-ci, 464 personnes ont un poids égal à 0 (groupe #3 de la variable DIA89) (tableau **7.9.**).

La somme des poids est égale à 4384.6, quel que soit le nombre de personnes prises en compte: 4755 **personnes interrogées** ou 4291 **membres du panel.**

La taille "réelle" de l'échantillon est égale à 4291 membres du panel. Les 464 personnes présentes dans cet échantillon sans être membres du panel ont reçu un poids nul, précisément parce qu'il est impossible d'estimer le poids de leur présence dans cet échantillon individuel.

La somme des poids doit être ajustée à la taille de l'échantillon. La somme des poids doit être égale au nombre de membres du panel afin de sauvegarder les propriétés statistiques de l'échantillon. C'est le rôle de la variable WGT89TOT.

7.2.5. Variable opérationnelle WGT89TOT

La taille de l'échantillon pondéré par la variable W89B serait de 4385 unités. Les poids doivent donc être ajustés par une constante: la moyenne des poids attribués aux individus (**2.2.1.4**). Suite à cette opération, la valeur moyenne des poids sera égale à 1.0 et la somme des poids sera égale au nombre de membres du panel.

Cet ajustement est effectué sur 4291 personnes **membres du panel.** Il s'en suit que les analyses des données peuvent être réalisées de manière équivalente sur l'échantillon de 4755

personnes observées en 1989 ou sur l'échantillon de 4291 **membres du panel** observés en 1989. Dans les deux cas, la taille réelle de l'échantillon sera respectée (4291 unités).

Dans le premier cas le nombre total de personnes prises en compte sera **apparemment** égal à 4755, mais la somme des poids sera égale à 4291 unités. Seuls, les membres du panel sont pris en compte au cours des analyses, puisque 464 personnes ont un poids égal à 0: les valeurs de ces personnes sur les variables analysées sont annulées.

Dans le second cas, seuls les membres du panel sont pris en compte dans l'analyse et la mise à l'écart des non-membres du panel n'affecte en rien les résultats (**2.2.1.7.**).

La variable W89B devient:

WGT89TOT = W89B / (4384.69 / 4291)

TABLEAU 7.10.
Ventilation des valeurs de WGT89TOT dans le fichier individuel

DIA89	WGT89TOT n =
0. memb. panel refus	-
1. memb. panel réponse	4220
2. nouv. memb. panel 89	71
3. non-memb. panel réponse	464
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4755
sysmis	2199
TOTAL	6954

Les valeurs de la variable WGT89TOT sont attribuées à l'ensemble des personnes ayant participé à l'enquête de 1989, soit les groupes #1, #2 et #3 de la variable DIA89 (tableau **7.10.**).

7.2.6. Variable opérationnelle de pondération des ménages en 1989 MWGT89A

La variable MWGT89A permet de pondérer correctement l'échantillon des **ménages** observé en 1989.

En 1985, le biais d'échantillonnage apparaît au niveau des ménages. La correction de ce biais s'effectue, en premier lieu, au niveau des ménages. Elle est répercutée sur les membres du ménage. Ceux-ci reçoivent le même poids que le ménage auquel ils appartiennent.

Il existe donc un rapport constant entre le poids du ménage et le poids des individus: le poids du ménage est égal à la valeur moyenne des poids des personnes qui le composent.

Depuis 1986, l'échantillon subit l'influence de facteurs d'évolution qui introduisent des biais aux deux niveaux d'analyse.

Tantôt un ménage complet disparaît et la raison du refus est commune à l'ensemble des membres du ménage.

Tantôt, un seul membre disparaît pour des raisons strictement individuelles.

Des nouveaux membres entrent dans les ménages. Les uns sont pris en compte au titre de membres du panel, les autres ne sont pas reconnus comme membres du panel.

De nouveaux ménages se créent et, généralement, provoquent l'inclusion d'une nouvelle personne étrangère à l'échantillon initial. Des ménages se défont et provoquent tantôt l'apparition de deux nouveaux ménages, tantôt la création d'un nouveau ménage et le retour d'une personne dans son milieu d'origine.

Cette liste n'est pas exhaustive. Elle illustre suffisamment à quel point les trajectoires des personnes et les trajectoires des ménages se différencient progressivement. Elle invite à fonder l'analyse de l'évolution de l'échantillon au niveau des trajectoires individuelles plutôt qu'au niveau des ménages.

Cependant, le rapport entre la valeur des poids relatifs individuels et la valeur des poids des ménages garde sa propriété initiale: le poids du ménage est égal à la valeur moyenne des poids des individus.

La procédure est simple.

* Les valeurs des poids individuels sont agrégées au niveau du ménage.

* La valeur moyenne des poids individuels (WGT89TOT) détermine la valeur du poids du ménage (MWGT89T).

  Le poids du ménage est calculé en fonction du **nombre de personnes ayant répondu** et non en fonction du nombre de personnes présentes dans le ménage.

* Cette valeur peut être ajustée en fonction de la taille de l'échantillon des ménages (MWGT89A).

La taille "réelle" de l'échantillon des ménages correspond à **1637 ménages composés d'au moins un membre du panel.**

En 1989, 26 ménages sont entièrement composés par des personnes qui ne sont pas membres du panel: ils reçoivent logiquement un poids de '0' (valeur moyenne des poids de leurs membres).

L'ajustement des poids des ménages s'effectue donc sur un nombre de ménages égal à 1637 unités et non sur 1637 + 26 ménages. De cette façon, les analyses seront toujours effectuées sur 1637 unités, indépendamment du fait que les ménages composés exclusivement de non-membres du panel sont pris en compte ou sont mis à l'écart (2.2.1.7.). A défaut d'un tel ajustement, l'échantillon pondéré par MWGT89T compterait 1588 unités. D'où:

MWGT89A = MWGT89T / (1588.31 / 1637)

TABLEAU 7.11.
Ventilation des valeurs de MWGT89A dans le fichier individuel

DIA89	MWGT89A n =
0. memb. panel refus	-
1. memb. panel réponse	4220
2. nouv. memb. panel 89	71
3. non-memb. panel réponse	464
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4755
sysmis	2199
TOTAL	6954

Les valeurs des poids des ménages (MWGT89T et MWGT89A) sont reventilées sur les individus. Chaque membre du ménage reçoit la valeur du poids du ménage auquel il appartient. Les non-membres du panel reçoivent, cette fois, un poids différent de 0, à condition qu'ils n'appartiennent pas à un ménage composé exclusivement de non-membres du panel. Cette procédure, purement technique, permet de regrouper l'ensemble des poids dans un seul fichier individuel.

Le tableau 7.11. présente les sous-échantillons individuels qu'il convient de sélectionner dans le fichier au moment de reconstituer l'échantillon des ménages observés en 1989. Les poids des ménages s'obtiennent aisément, par exemple, en agrégeant les valeurs de MWGT89A au niveau des ménages et en prenant la moyenne comme valeur de pondération. (Certains logiciels permettent de prendre la valeur du premier membre du ménage comme valeur de référence (SPSSX)).

Il peut être utile de rappeler que TOUS les membres appartenant à l'échantillon observé en 1989 doivent être pris en compte au moment de reconstituer l'échantillon des ménages. En l'absence des non-membres du panel, il serait impossible de calculer certaines variables propres au ménage (par exemple, la taille du ménage).

ANNEXE 7 A

TABLEAU COMPARATIF DES CARACTERISTIQUES DEMOGRAPHIQUES DES MEMBRES DU PANEL DE 1985 A 1989

(après pondération)

- caractéristiques personnelles
- caractéristiques du ménage
- caractéristiques du chef de ménage

Caractéristiques Personnelles	1985 (N=6110)	1986 (N=5245)	1987 (N=4648)	1988 (N=4477)	1989 (N=4291)
Sexe					
1. homme	48.8	48.7	48.2	48.0	48.1
2. femme	51.2	51.3	51.8	52.0	51.9
Age					
1. < 18 ans	24.1	23.5	23.6	22.9	22.5
2. 18 - 24 ans	11.2	11.2	11.1	11.0	11.4
3. 25 - 34 ans	15.6	15.1	15.2	14.6	15.0
4. 35 - 44 ans	14.3	14.3	14.6	15.1	14.4
5. 45 - 54 ans	12.7	12.8	12.6	13.0	13.4
6. 55 - 64 ans	10.4	10.9	10.5	11.5	10.9
7. > 64 ans	11.7	12.1	12.4	11.9	12.5
Age par Sexe					
1. homme < 18 ans	12.7	12.4	12.0	11.6	11.2
2. homme 18 - 64 ans	31.7	31.8	31.5	32.0	32.4
3. homme > 64 ans	4.4	4.6	4.7	4.4	4.5
4. femme < 18 ans	11.4	11.2	11.6	11.2	11.3
5. femme 18 - 64 ans	32.6	32.6	32.5	33.2	32.6
6. femme > 64 ans	7.2	7.5	7.7	7.5	8.0
Etat civil					
1. célibataire	39.8	39.9	40.0	40.0	40.6
2. marié	49.6	49.2	48.8	48.9	48.5
3. séparé	1.0	0.8	0.8	0.9	0.7
4. divorcé	1.9	1.9	2.3	2.0	2.0
5. veuf	7.9	8.0	8.1	8.2	8.2
6. N.R.	0.0	-	-	-	-

Nationalité

1. Luxembourgeois	76.1	76.4	77.9	78.7	79.4
2. autres C.E.	21.3	21.3	20.3	19.7	18.8
3. hors C.E.	2.5	2.3	1.9	1.6	1.3
4. N.R.	0.1	-	-	-	-

Emploi

1. oui	38.1	39.8	37.1	40.3	39.8
2. non	35.4	35.3	34.2	34.6	35.1
3. enfants	26.5	24.9	28.7	25.1	25.1

Adulte/enfant

1. adulte	73.5	75.1	74.7	74.9	74.9
2. enfant	26.5	24.9	25.3	25.1	25.1

Lien avec le chef de ménage

1. chef de ménage	36.4	37.2	36.6	37.8	37.1
2. épouse	24.2	24.1	23.9	23.8	23.7
3. "épouse"	1.1	1.0	0.9	1.1	1.0
4. ami, amie	0.3	0.1	0.2	0.1	0.1
5. fils, fille	33.4	33.6	34.2	33.5	34.2
6. autres	4.6	4.0	4.2	3.7	4.0

Caractéristiques du ménage

Type d'habitation

1. maison rurale	11.4	10.7	9.3	7.5	8.2
2. unifamiliale individuelle	33.7	34.7	34.7	33.5	33.0
3. unifamiliale jumelée	13.5	8.8	9.1	10.3	10.4
4. unifamiliale en série	19.5	21.4	25.3	26.5	26.6
5. petit collectif	12.6	13.1	13.1	12.4	11.3
6. collectif moyen	7.1	7.3	5.8	6.2	7.0
7. grand collectif	1.1	1.4	1.1	0.8	1.1
8. précaire	0.3	0.0	0.1	0.1	0.0
9. N.R.	0.8	2.6	1.7	-	1.8
0. maison de retraite	-	-	-	0.3	0.5

Canton

1. Luxembourg-ville	17.6	17.6	17.7	16.2	15.9
2. Luxembourg-campagne	8.9	8.8	8.8	9.3	9.2
3. Capellen	6.9	7.0	6.9	7.1	7.2
4. Esch sur Alzette	33.5	33.6	34.3	33.7	33.4
5. Mersch	4.6	4.3	3.7	4.2	4.6
6. Clervaux	2.3	2.3	2.5	2.5	2.5
7. Diekirch	6.2	6.2	6.7	6.7	6.5
8. Redange	3.0	3.0	2.8	3.0	2.9
9. Vianden	0.8	0.7	0.8	0.8	0.8
10. Wiltz	3.0	3.1	3.0	3.0	3.1
11. Echternach	3.5	3.5	3.5	3.4	3.9
12. Grevenmacher	5.8	5.8	5.7	6.3	5.9
13. Remich	4.0	4.0	3.6	3.9	4.0

Nombre de personnes dans le ménage

1.	7.6	8.0	8.0	8.5	8.4
2.	21.0	20.6	19.6	21.1	20.7
3.	22.5	23.1	22.8	22.1	22.7
4.	27.6	27.6	28.7	28.6	28.0
5.	12.3	12.7	12.9	12.7	12.9
6.	5.8	5.3	4.9	5.0	4.6
7.	2.1	1.8	2.1	1.2	1.7
8.	0.7	0.6	0.7	0.7	1.0
9.	0.3	0.3	0.2	-	-
11.	0.0	-	0.1	-	-
12.	-	0.1	-	0.1	-

Nombre d'enfants dans le ménage

0.	39.2	41.9	41.7	43.4	43.4
1.	24.3	23.2	23.6	21.2	22.9
2.	22.9	23.2	22.8	24.1	21.6
3.	9.9	9.5	9.1	8.7	9.8
4.	3.5	2.2	2.7	2.5	2.0
5.	0.1	0.1	0.1	0.1	0.3
6.	0.1	-	-	-	-

Nombre d'adultes dans le ménage

0.	0.1	-	-	-	-
1.	9.9	10.2	10.4	11.0	10.7
2.	63.0	59.5	58.4	60.3	60.3
3.	16.2	18.9	18.9	17.4	16.5
4.	7.7	6.9	7.8	7.8	8.3
5.	2.1	3.0	2.8	2.4	2.5
6.	0.8	1.2	1.4	0.8	1.2
7.	0.2	0.2	0.2	0.3	0.3
9.	-	0.1	0.1	-	0.1

Nombre d'emplois dans le ménage

0.	17.1	17.5	17.4	17.3	18.1
1.	46.9	43.1	42.1	43.4	42.7
2.	29.2	29.8	30.0	28.9	28.6
3.	4.9	6.8	6.9	7.1	6.9
4.	1.4	2.1	2.5	2.6	2.7
5.	0.5	0.7	0.9	0.6	0.8
6.	0.0	0.0	0.1	0.1	0.4

Caractéristiques du chef de ménage

Sexe

1. homme	87.2	86.0	86.3	83.6	85.6
2. femme	12.8	14.0	13.7	16.4	14.4

Age

1. < 24 ans	2.6	2.3	1.8	2.1	2.1
2. 25 - 34 ans	18.3	18.1	18.7	17.7	17.0
3. 35 - 44 ans	27.6	27.1	27.3	26.6	26.8
4. 45 - 54 ans	24.4	23.7	23.0	23.7	23.9
5. 55 - 64 ans	14.8	15.9	15.8	16.9	16.6
6. > 64 ans	12.2	12.9	13.4	13.1	13.5

Etat civil

1. célibataire	6.0	6.1	6.5	6.2	6.2
2. marié	80.6	80.0	79.6	79.2	79.5
3. séparé	1.4	1.7	1.2	1.3	0.9
4. divorcé	2.5	2.7	2.7	3.0	3.1
5. veuf	9.5	9.6	10.0	10.3	10.3
6. N.R.	0.0	-	-	-	-

Emploi

1. oui	74.1	73.3	73.3	72.0	70.9
2. non	25.9	26.7	26.7	28.0	29.1

Etat civil par Sexe

1. homme non-marié	7.0	6.7	7.0	6.2	6.5
2. homme marié	80.2	79.3	79.3	77.4	79.1
3. femme non-mariée	12.4	13.4	13.4	14.6	13.9
4. femme mariée	0.4	0.6	0.3	1.8	0.4

Emploi par Sexe

1. homme sans emploi	17.5	17.4	17.6	19.0	20.0
2. homme avec emploi	69.8	68.6	68.7	64.6	65.6
3. femme sans emploi	8.5	9.3	9.1	9.0	9.1
4. femme avec emploi	4.3	4.7	4.6	7.4	5.3

ÉCHANTILLON 1990

ET VARIABLES DE PONDÉRATION

Encadré 8.1. **VARIABLES DE PONDERATION 1990**

8.2.1.	Variable instrumentale	TREE90	= Typologie de l'échantillon selon les % de réponse
8.2.2.	Variable instrumentale	RR90	= Distribution des fréquences des % de réponse
8.2.3.	Variable instrumentale	W90	= Poids individuels 1990, avant ajustement, SANS LES NOUVEAUX MEMBRES
8.2.4.	Variable instrumentale	W90B	= Poids individuels 1990, avant ajustement, AVEC LES NOUVEAUX MEMBRES
8.2.5.	Variable opérationnelle	WGT90TOT	= Poids individuels 1990, après ajustement, AVEC LES NOUVEAUX MEMBRES
8.2.6.	Variable opérationnelle	MWGT90A	= Poids des MENAGES 1990 après ajustement

Chapitre 8

ÉCHANTILLON 1990

ET VARIABLES DE PONDÉRATION

L'échantillon observé en 1990 n'est plus conforme à l'échantillon observé en 1985. Phénomènes démographiques et phénomènes propres à la vie de l'échantillon conjuguent leurs effets et provoquent une dérive progressive de l'échantillon.

Les trajectoires des personnes dans l'étude se diversifient de plus en plus. Ces trajectoires déterminent la position des personnes dans la procédure de pondération.

La première partie du chapitre sera consacrée à la description de cette évolution de l'échantillon.

La seconde partie présentera les différentes variables qui ont jalonné l'élaboration du nouveau système de pondérations.

Pour chaque variable, un tableau récapitule les sous-échantillons pertinents et la taille de ces sous-échantillons.

8.1. L'échantillon 1990

8.1.1. Les trajectoires individuelles: version courte

TABLEAU 8.1.
Trajectoires individuelles 1985 - 1990

VARIABLE DE SÉLECTION: DIA90	n =
0. memb. panel refus	1900
1. memb. panel réponse	4108
2. nouv. memb. panel 90	51
3. non-memb. panel réponse	576
4. non-memb. panel refus	119
5. émigrés / décédés	398
FICHIER TOTAL	7152

Le nouveau fichier contient désormais 7152 personnes classées en 6 catégories principales. Le contenu de ces catégories évolue d'année en année.

Le nouvel échantillon compte 4735 personnes réparties en 3 classes:

- les membres du panel ayant répondu (4108),
- les nouveaux membres du panel entrés dans l'échantillon en 1990 (51)
- et les membres de l'échantillon qui ne sont pas membres du panel (576).

Le taux de réponses est calculé en rapportant le nombre des réponses des *membres du panel* (4108) au nombre total des membres du panel (6008, soit: 1900 + 4108).

Les nouveaux membres du panel entrés dans l'échantillon en 1990 (51) et les personnes émigrées ou décédées ne sont pas prises en compte dans le calcul du taux de réponse.

Ces 6 classes sont définies en fonction du point initial et du point final des trajectoires des individus dans l'ensemble de l'étude. Le point initial détermine le statut de la personne au moment ou elle est entrée dans l'étude: membre ou non-membre du panel. Le point final correspond à la présence ou à l'absence de la personne dans l'échantillon considéré.

Ceci signifie, par exemple, que les membres du panel

* entrés dans l'échantillon initial en 1985

* ayant refusé de répondre en 1986, en 1987, en 1988 et en 1989

* peuvent être classés, en 1990

- soit parmi les refus (groupe #0)

- soit parmi les réponses (groupe #1) s'ils ont accepté de renouveler leur coopération, s'ils ont rejoint un ménage d'origine toujours présent dans l'échantillon, etc ...

L'ensemble de toutes ces trajectoires est présenté au tableau **8.2.**. Chaque trajectoire est définie par la succession des états occupés par les personnes au cours des années écoulées.

Le tableau **8.1.** rassemble ces trajectoires en 6 classes principales parce que ces classes

* sont nécessaires et suffisantes pour permettre le calcul des nouvelles pondérations

* permettent de gérer facilement les sous-échantillons principaux qui forment l'ensemble du fichier individuel.

Le tableau **8.2.** permet de recomposer d'autres échantillons en fonction des besoins de l'analyste.

La variable DIA90 comporte six grandes modalités qui peuvent être décrites de la manière suivante.

8.1.1.1. *Les refus des membres du panel*

La première catégorie est formée par les membres du panel ayant refusé de prendre part à l'enquête. Cet ensemble ne correspond plus à la première catégorie de la variable DIA utilisée en 1986 ni à la variable DIA89 utilisée en 1989. Elle réunit des membres du panel présents en 1985 ET des nouveaux membres entrés en 1986, en 1987, en 1988 ou en 1989.

Tous ces membres ont un point commun: **ils n'ont pas répondu en 1990.**

8.1.1.2. *Les membres du panel*

Le groupe #1 réunit:
- des membres du panel présents en 1985,
- des nouveaux membres entrés en 1986,
- des nouveaux membres entrés en 1987,
- des nouveaux membres entrés en 1988,
- des nouveaux membres entrés en 1989.

Ils ont un point commun: **ils ont répondu en 1990.**

Tous ces membres sont réunis dans cette catégorie, même s'ils n'ont pas répondu au cours de la seconde, de la troisième ou de la quatrième vague d'enquêtes.

8.1.1.3. *Les nouveaux membres du panel*

Le groupe #2 isole les nouveaux membres du panel entrés dans l'échantillon en 1990. Ce sont des nouveaux-nés descendants directs de membres du panel.

Ce groupe ne contient pas les nouveaux membres du panel entrés entre 1986 et 1989. Ces derniers font désormais partie de l'ensemble des membres du panel. Ils sont traités à ce titre dans la procédure de pondération. Ils sont répartis entre le groupe #0, le groupe #1 et le groupe #5 en fonction de la trajectoire qu'ils ont suivie entre 1986 et 1990.

Ce groupe #2 permet d'identifier les nouveaux membres de l'échantillon annuel qui recevront un poids individuel lié à celui de leurs ascendants. Les membres de ce groupe sont donc différents chaque année.

8.1.1.4. *Les non-membres du panel*

Les groupes #3 et #4 sont formés par l'ensemble des nouvelles personnes entrées dans les échantillons successifs sans accéder au titre de membres du panel.

Ces personnes sont classées dans le groupe #3 lorsqu'elles appartiennent à l'échantillon annuel. Les nouveaux non-membres du panel entrés dans l'échantillon en 1990 sont donc classés dans ce groupe au même titre que les non-membres du panel entrés dans l'étude au cours de vagues antérieures. Ils ont un point commun: **ils ont répondu en 1990.**

Les non-membres du panel, entrés dans les échantillons précédents, sont classés dans le groupe #4 **lorsqu'ils n'ont pas participé à l'enquête en 1990.**

Cette distinction entre les groupes #3 et #4 permet de reconstituer aisément l'échantillon annuel, lorsqu'il y a lieu de prendre en compte les membres du panel ET les non-membres du panel présents dans l'échantillon: il suffit de sélectionner les groupes #1, #2 et #3.

Cette distinction est également très utile au moment de regrouper les individus appartenant à un même ménage ou à un même groupe de revenu.

Elle est indispensable au moment de calculer le poids individuel des nouveaux membres du panel: par exemple, le poids du nouveau membre est différent selon que l'un des ascendants directs est absent de l'échantillon ou non-membre du panel (**8.2.4.**).

8.1.1.5. *Les émigrés et les décédés*

Le groupe #5 réunit toutes les personnes qui ont été présentes dans un échantillon au moins une fois, et qui ne sont pas présentes dans l'échantillon de l'année considérée pour une raison d'ordre *démographique*.

Cette distinction s'impose puisque les émigrés et les personnes décédées ne sont pas prises en compte dans le système de pondération annuel.

Leur classement dans une catégorie particulière facilite également la sélection de l'échantillon annuel et la reconstitution des ménages.

Il faut rappeler qu'une personne émigrée peut rejoindre l'échantillon ultérieurement. Elle sera reclassée dans son groupe d'origine, c'est à dire dans le groupe #1 si elle était membre du panel ou dans le groupe #3 si elle n'était pas membre du panel.

8.1.2. Les trajectoires individuelles: version complète

TABLEAU 8.2.
Liste des trajectoires individuelles: 1985-1990

VARIABLE: DIA6	
CODES	TRAJECTOIRES
101	'F/85..../90'
105	'F/85..../E'
106	'F/85..../D'
107	'F/85..../.'
111	'F/85...89/90'
201	'F8586.../90'
207	'F8586.../.'
210	'F8586..89/.'
211	'F8586..89/90'
307	'F.86.../.'
406	'F85.87../D'
501	'F858687../90'
507	'F858687../.'
510	'F858687.89/.'
511	'F858687.89/90'
607	'F.8687../.'
707	'F85E87../.'
807	'F..87../.'
811	'F..87.89/90'
1011	'F85..8889/90'
1107	'F8586.88./.'
1110	'F8586.8889/.'
1111	'F8586.8889/90'
1115	'F8586.8889/E'
1137	'F8586.88D/-'
1211	'F.86.8889/90'
1311	'F85.878889/90'
1401	'F85868788./90'
1407	'F85868788./.'
1410	'F8586878889/.'
1411	'F8586878889/90'
1415	'F8586878889/E'
1416	'F8586878889/D'
1421	'F85868788E/90'
1427	'F85868788E/-'
1437	'F85868788D/-'
1507	'F.868788./.'
1510	'F.86878889/.'
1511	'F.86878889/90'
1515	'F.86878889/E'
1527	'F.868788E/-'

VARIABLE: DIA6 (suite)	
CODES	TRAJECTOIRES
1607	'F..8788./.'
1610	'F..878889/.'
1611	'F..878889/90'
1711	'F8586E8889/90'
1715	'F8586E8889/E'
1807	'F...88./.'
1810	'F...8889/.'
1811	'F...8889/90'
1827	'F...88E/-'
1830	'F....89/.'
1831	'F....89/90'
1832	'F.....90'
2007	'N.86.../.'
2011	'N.86..89/90'
2107	'N.8687../.'
2201	'N..87../90'
2207	'N..87../.'
3007	'N.86.88./.'
3011	'N.86.8889/90'
3107	'N.868788./.'
3110	'N.86878889/.'
3111	'N.86878889/90'
3116	'N.86878889/D'
3137	'N.878788D/-'
3201	'N..8788./90'
3207	'N..8788./.'
3210	'N..878889/.'
3211	'N..878889/90'
3227	'N..8788E/-'
3237	'N..8788D/-'
3301	'N...88./90'
3307	'N...88./.'
3310	'N...8889/.'
3311	'N...8889/90'
3315	'N...8889/E'
3316	'N...8889/D'
3321	'N...88E/90'
3327	'N...88E/-'
3337	'N...88D/-'
3340	'N....89/.'
3341	'N....89/90'

.../..

VARIABLE: DIA6 (suite)	
CODES	TRAJECTOIRES
3342	'N...../90'
4007	'F85E---/-'
4107	'F8586E--/-'
4115	'F8586E-89/E'
4207	'F858687E-/-'
4307	'F..87E-/-'
5007	'F85D---/-'

VARIABLE: DIA6 (suite)	
CODES	TRAJECTOIRES
5107	'F8586D--/-'
5207	'F8586.D-/-'
5307	'F858687D-/-'
5407	'F.86D--/-'
6007	'N.86D--/-'
6107	'N.8687D-/-'
6207	'N..87D-/-'

Le tableau **8.2.** est composé de la manière suivante.

La variable est nommée **DIA6** parce qu'elle correspond à l'état du fichier après la sixième vague d'enquêtes.

Les 7152 personnes recensées dans le fichier total après six vagues d'enquêtes ont suivi 96 trajectoires différentes entre 1985 et 1990 (contre 63, entre 1985 et 1989).

Chaque trajectoire est un vecteur défini par une série d'états successifs. Chaque état correspond au statut occupé par une personne dans le panel ET dans l'échantillon suite à une vague d'observation annuelle.

Chaque vecteur est désigné par un code. Ces codes ont été attribués de manière arbitraire et peuvent être modifiés.

Les vecteurs 511, 1111, 1311, 1401, 1511 désignent des membres du panel absents pendant une année et présents dans l'échantillon en 1990.

Les vecteurs sont décrits de la manière suivante:

* **F** signifie que ce vecteur contient exclusivement des **membres du panel** (Full Member).
  N signifie que le vecteur contient exclusivement des personnes **non membres du panel** (Non Full member).

* **85** signifie que ces personnes sont **membres de l'échantillon** observé en 1985.
  86 signifie que ces personnes sont membres de l'échantillon observé en 1986.
  L'année est remplacée par le signe '.' lorsqu'une personne est **absente de l'échantillon** suite à un refus ou lorsqu'il n'est pas avéré que cette absence est liée à une raison d'ordre démographique.

  Par exemple: F8586.../90 (VECTEUR # 201) désigne le vecteur des personnes

- entrées dans l'échantillon au cours de la première vague d'enquête effectuée en 1985
- toujours présentes dans l'échantillon observé en 1986
- absentes des échantillons suivants pour des raisons assimilées à des refus
- mais présentes dans l'échantillon en 1990, suite aux efforts d'un(e) enquêteur(trice).

* **E** signifie que ces personnes ont **émigré** et sont absentes de l'échantillon correspondant à la vague d'enquête de cette année. Dans certains cas cette absence est provisoire.

Le vecteur #707 est composé de la manière suivante:

- **F85E87../.** soit les membres du panel entrés dans l'échantillon en 1985, émigrés en 1986, revenus dans l'échantillon en 1987 et refusant de répondre en 1988, en 1989 et en 1990.

* **D** signifie que ces personnes sont **décédées.**

Le vecteur #6207 est composé de la manière suivante:

- **N..87D-/-** soit les personnes entrées dans l'étude en 1987 sans accéder au titre de membres du panel et décédées au cours de l'année 1988.

Le passage de la variable DIA6 à la variable DIA90 s'opère aisément. Les classes générales de trajectoires sont composées à partir du point initial et du point final de chaque vecteur.

Le **groupe #1** de la variable DIA89 regroupe l'ensemble des membres du panel entrés dans l'échantillon en 1985, en 1986, en 1987, en 1988 ou en 1989 et toujours présents dans l'échantillon en 1990, quelles que soient les étapes intermédiaires ayant jalonné leur trajectoire.

Le vecteur #1411 regroupe l'ensemble des membres du panel présents tous les ans dans l'échantillon. Ce groupe est important lorsque l'analyste souhaite analyser uniquement les personnes qui ont participé à l'enquête chaque année.

Le regroupement des vecteurs désignés par la lettre **F** et se terminant par **"/90"** équivaut au regroupement des classes #1 et #2 de la variable DIA90: il désigne l'ensemble des membres du panel présents dans l'échantillon en 1990 soit 4159 personnes (dont 51 nouveaux membres entrés dans l'échantillon en 1990).

8.2. Les variables de pondération

8.2.1. Variable instrumentale TREE90

Les refus de répondre constituent le facteur principal de la déformation de l'échantillon (2.2.1.1.).

1. La notion de "taux de réponses" a pris désormais un sens particulier. Elle correspond davantage à la notion de "taux de survie dans le panel" (8.2.1.1).

2. Les refus de répondre seraient sans effet s'ils étaient répartis aléatoirement dans l'échantillon. Il convient donc de rechercher les concentrations éventuelles des refus dans des groupes démographiques homogènes afin d'en corriger les effets (8.2.1.3).

3. Au préalable, l'inégalité des probabilités de sélection initiales des membres du panel doit être annulée en pondérant les individus par l'inverse de cette probabilité. La recherche des groupes démographiques atteints par les taux de refus les plus contrastés ne doit pas être influencée par ce biais lié au mode de tirage de l'échantillon initial (8.2.1.2.).

8.2.1.1. *Des taux de survie*

Il est important de rappeler que la notion de taux de réponses est entendue désormais dans un sens particulier.

Le terme "taux de survie" est plus correct que le terme "taux de réponses". En effet, le **taux de survie** prend en compte le **cumul** des refus enregistrés au cours des enquêtes successives à condition que ces personnes n'aient pas rejoint l'échantillon en 1990.

Le taux de réponses correspond, en fait, au nombre des réponses enregistrées au cours d'une vague d'enquête, rapporté au nombre cumulé des membres du panel.

> *Il n'est pas calculé par référence au nombre des membres du panel présents dans l'échantillon précédent. Il est calculé en rapportant le nombre de réponses des membres du panel au nombre total des membres du panel entrés dans les échantillons successifs depuis la première vague d'enquête.*

Cette procédure permet de prendre en compte la mobilité des membres du panel: ils peuvent entrer et sortir de l'échantillon. Chaque fois qu'ils rejoignent l'échantillon, leur

nouvelle probabilité de sélection peut être calculée puisque les deux éléments essentiels du calcul sont connus:

* leur probabilité de sélection initiale (**8.2.1.2.**)

* et les caractéristiques démographiques du sous-groupe auquel ils appartenaient au moment ou ils sont entrés dans l'échantillon (**8.2.1.3.**).

8.2.1.2. *Correction du biais initial*

L'analyse des taux de réponses prend en compte les 6008 personnes regroupées dans les classes #0 et #1 de la variable DIA90. Elle ne prend en compte, ni les nouveaux membres du panel entrés dans l'échantillon au cours de la dernière vague d'enquête, ni les non-membres du panel, ni les personnes émigrées ou décédées.

Parmi les 6008 membres du panel, entrés dans l'échantillon de 1985 à 1989, 4108 ont répondu en 1990, soit 68.4%.

Ces valeurs ne prennent pas en compte l'inégalité des probabilités de sélection initiales des individus. Il convient donc d'annuler cette inégalité avant d'entreprendre toute autre opération.

Les membres entrés dans l'échantillon en 1985 doivent être pondérés par l'inverse de leur probabilité de sélection soit, par la variable PONREV.

La probabilité de sélection des nouveaux membres correspond à la valeur moyenne des probabilités de sélection de leurs parents *au moment ou ces nouveaux membres entrent dans le panel.* Ces nouveaux membres ont déjà reçu ce poids au moment ou ils sont entrés dans l'échantillon: c'est la valeur de la variable W86B attribuée en 1986 aux nouveaux membres de l'année; c'est la valeur de W87B attribuée en 1987 aux nouveaux membres entrés en 1987; ce sont, encore, la valeur de W88B attribuée en 1988 aux nouveaux membres entrés en 1988 et la valeur de W89B attribuée en 1989 au nouveaux membres entrés en 1989.

La synthèse de ces valeurs donne lieu à la création de la variable **IP8590.** Cette variable est sauvegardée dans le fichier global parce qu'elle sera réutilisée l'année prochaine dans le cadre de la même procédure. (Cette variable doit être ajustée en fonction de la taille de l'échantillon, soit IP8590A).

Cette première procédure modifie la valeur du taux de survie. Le tableau **8.3.** permet d'observer les glissements des proportions avant et après la correction des biais de sélection.

TABLEAU 8.3.

Distribution des fréquences des individus membres du panel, en 1990, selon leur position dans l'échantillon, avant et après pondération

DIA90	FREQUENCES			
	avant pondération		après pondération (IP8590A)	
	n	%	n	%
0. memb.pan. refus	1900	31.6	1810.	30.1
1. memb.pan. réponses	4108	68.4	4198.	69.9
TOTAL	6008	100.0	6008	100.0

Le taux de survie passe de 68.4% à 69.9% lorsque les biais de sélection initiaux sont contrôlés.

8.2.1.3. *Les nouvelles probabilités de sélection*

La qualité de l'échantillon n'est pas affectée par les refus de répondre aussi longtemps qu'ils se distribuent de manière aléatoire dans l'ensemble de l'échantillon.

La méthode permettant de repérer des concentrations éventuelles de ces refus dans certains sous-groupes homogènes de l'échantillon a été présentée dans le second chapitre (2.2.1.1.).

La démarche met en concurrence le pouvoir discriminant de 16 variables:

CARACTERISTIQUES DU MÉNAGE	DU CHEF DE MÉNAGE	DE LA PERSONNE
Canton de résidence	Age	Age
Type d'habitat	Sexe	Sexe
Taille du ménage	Etat civil	Etat civil
Nombre d'adultes	Emploi (oui/non)	Emploi
Nombre d'enfants		Adulte/Enfant
Nombre d'emplois		Lien avec le CM

Toutes ces caractéristiques sont attribuées aux personnes. Les caractéristiques du ménage et les caractéristiques du chef de ménage sont des éléments de l'environnement de la personne.

Les membres entrés dans l'échantillon en 1985 sont caractérisés par les valeurs qu'ils présentaient sur ces variables **en 1985**, au moment où ils sont entrés dans l'échantillon.

Les nouveaux membres, entrés de 1986 à 1989, sont caractérisés par les valeurs qu'ils présentaient sur ces différentes variables au moment de leur entrée dans l'échantillon. Ces variables décrivent donc le contexte du ménage dans lequel ils sont nés et *au moment où ils sont nés.*

Il est donc fort probable qu'il existe un décalage systématique entre les caractéristiques des nouveaux membres et les caractéristiques de leurs parents.

Si la taille du ménage dans lequel ils font leur entrée s'est modifiée entre 1985 et 1989, les nouveaux membres du panel entrés en 1989 ne sont pas caractérisés par la même valeur que leurs parents sur cette variable.

L'âge du chef de ménage fournit un exemple simple: en règle générale, l'âge du chef de ménage sera différent pour les parents et pour le nouveau membre du panel entré dans la famille cinq ans plus tard.

La synthèse de ces valeurs donne lieu à la création de nouvelles variables. Ces variables sont sauvegardées dans le fichier global parce qu'elles seront réutilisées l'année suivante dans le cadre de la même procédure.

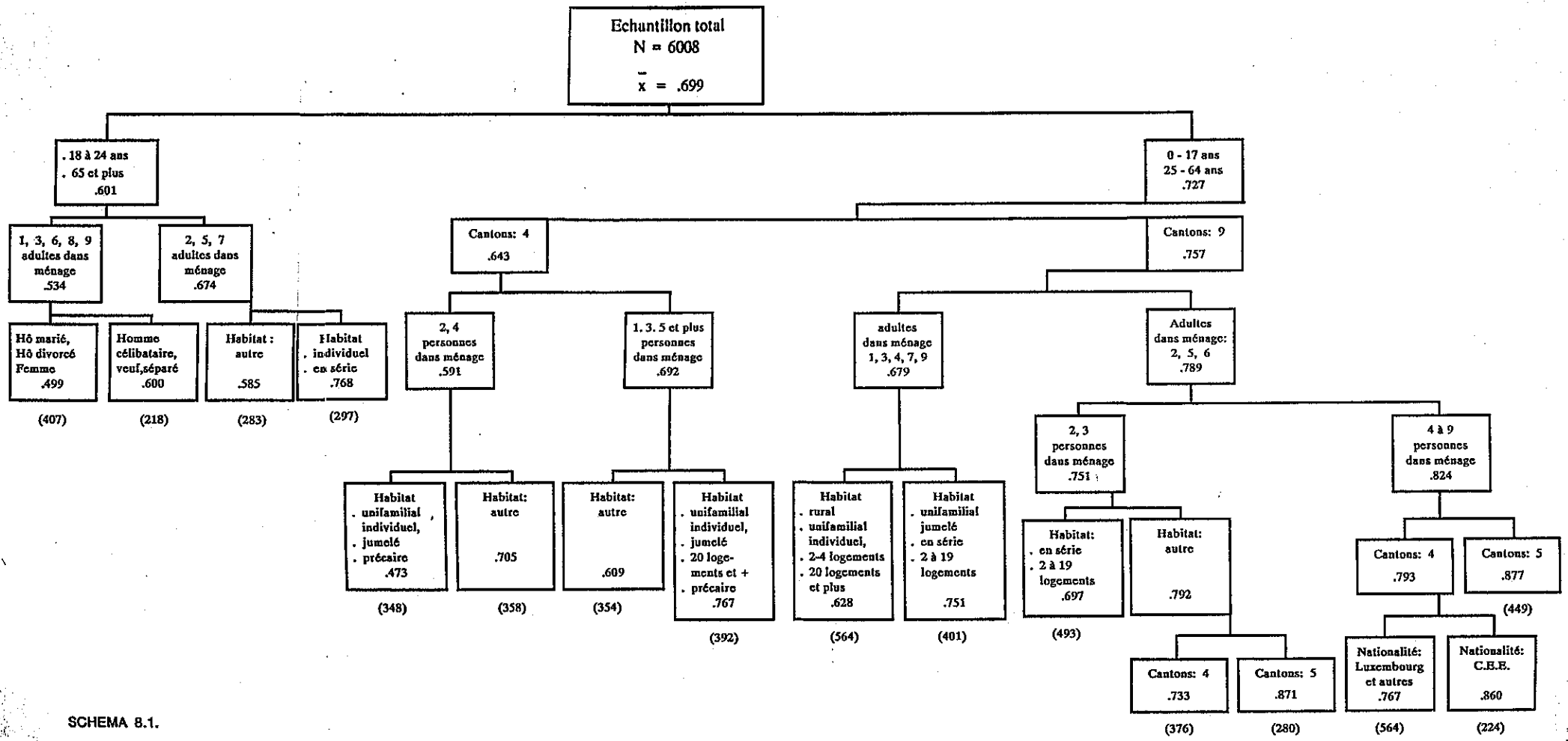
(La variable de correction des probabilités de sélection initiales a été actualisée selon la même logique (**8.2.1.2.**)).

Lorsque l'ensemble des variables choisies ont été actualisées, la recherche des catégories démographiques les plus discriminantes peut être entreprise (**2.2.1.1.**).

En résumé, la procédure consiste à diviser l'échantillon en sous-groupes mutuellement exclusifs par une succession de divisions binaires. L'enchaînement de ces divisions successives crée des filières ou séries de modalités. Les variables qui produisent ces divisions sont choisies parce qu'elles créent les contrastes les plus significatifs entre les taux de réponses des deux sous-groupes. Ces divisions sont interrompues dès qu'un sous-groupe ne peut plus être divisé en deux sous-groupes d'une taille supérieure ou égale à 5% de l'échantillon. (Cf. Procedures and Tape Codes, 1978 Interviewing Year, Wave XI, a Supplement. P.S.I.D., I.S.R., Ann Arbor, Michigan, 1979, p.10).

Le résultat de cette analyse est présenté dans le schéma **8.1.**. Il reproduit sous une forme arborescente la progression de l'analyse vers des sous-groupes de plus en plus homogènes.

Analyse des taux de réponses des sous-groupes entre 1985 et 1990
(après pondération initiale)



SCHEMA 8.1.

Seize sous-groupes sont identifiés à l'aide de 9 variables.

La variable dépendante est une variable binaire (0,1) qui rend compte du **taux de survie** dans le panel longitudinal:

$$\text{Taux de survie} = \frac{\text{Nombre de codes 1}}{\text{Nombre de cas dans le sous-groupe}}$$

Les 16 sous-groupes "n'expliquent" que 6.4% de la variance de la variable dépendante.

L'objectif de cette procédure n'était pas de prendre en compte toute la variance des taux de survie. Des mesures ont été prises en vue d'éviter que l'analyse identifie des groupes marginaux de très petite taille. La variance des poids des individus est maintenue dans des limites acceptables[1]

Le faible pouvoir explicatif de la typologie constituée par cette procédure provient également du fait que les modalités initiales des variables démographiques divisent l'échantillon en segments dont les moyennes sont peu contrastées (tableau **8.4.**).

1. La variance des poids des individus (ou variance de l'échantillon) a évolué de la manière suivante: 1985:'.18'; 1986:'.18'; 1987:'.19'; 1988: '.20'; 1989: '.18'; 1990: '.20'.

TABLEAU 8.4.

Taux de survie des membres du panel en 1990

CARACTÉRISTIQUES DES SOUS-GROUPES	TAUX DE REPONSE $\bar{x}$ = .699	VARIANCE "EXPLIQUÉE" (ETA)	VARIANCE EXPLIQUÉE PAR LA MEILLEURE DICHOTOMISATION
Caractéristiques individuelles			
Sexe			
1. masculin	.70	.01	.01
2. féminin	.69		
Etat civil			
1. célibataire	.70		
2. marié	.72		
3. séparé	.59	.08	.08
4. divorcé	.61		
5. veuf	.57		
6. D.M.	(.00)*		
Age			
1. < 17 ans	.76		
2. 18 à 24 ans	.60		
3. 25 à 34 ans	.69		
4. 35 à 44 ans	.71	.11	.10
5. 45 à 54 ans	.70		
6. 55 à 64 ans	.69		
7. 65 ans et plus	.59		
Nationalité			
1. Luxembourgeois	.70		
2. CEE	.70		
3. hors CEE	.59	.05	.00
4. D.M.	(.16)		

* Les chiffres entre parenthèses correspondent à des catégories de moins de 20 personnes.

Situation professionnelle

1. Enfants	.75		
2. Travail ou maladie	.68		
3. Demandeur d'emploi	.58	.07	.07
4. Pensionné, invalide	.69		
5. Retraité	.65		
6. Ménage, milice	.68		

Lien avec le chef de ménage

1. Chef de ménage	.69		
2. épouse	.72		
3. "épouse"	.57	.07	.06
4. ami(e)	(.65)		
5. fils, fille	.71		
6. autre	.57		

Adultes

1. Enfant	.75	.07	.07
2. Adulte	.68		

Etat civil selon le sexe

1. Homme célibataire	.70		
2. marié	.72		
3. séparé	.54		
4. divorcé	.68		
5. veuf	.66		
6. M.D.	(.00)	.09	.03
7. Femme célibataire	.70		
8. mariée	.72		
9. séparée	.63		
10. divorcée	.56		
11. veuve	.56		

Caractéristiques du ménage

Nombre de personnes dans le ménage

1.	.59		
2.	.69		
3.	.71		
4.	.70		
5.	.73	.08	.06
6.	.73		
7.	.66		
8.	.80		
9.	(.89)		
11.	(.09)		
12.	(.00)		

Nombre d'enfants dans le ménage

0.	.66		
1.	.69		
2.	.74		
3.	.72	.09	.07
4.	.84		
5.	(1.00)		
6.	(.78)		

Nombre d'adultes dans le ménage

0.	(.00)		
1.	.62		
2.	.73		
3.	.64	.10	.10
4.	.62		
5.	.70		
6.	.61		
7.	(.72)		
9.	(.00)		

Nombre d'emplois dans le ménage

0.	.67		
1.	.71		
2.	.71		
3.	.60	.07	.05
4.	.62		
5.	(.79)		
6.	(.00)		

Type d'habitation

1. maison rurale	.72		
2. maison unifamiliale individuelle	.69		
3. -"- jumelée	.68		
4. -"- en série	.69	.03	.02
5. petit collectif	.70		
6. collectif moyen	.70		
7. grand collectif	.76		
8. précaire	(.71)		
9. Non réponse	.63		

Canton de résidence

1. Luxembourg-Ville	.62		
2. Luxembourg-Campagne	.76		
3. Capellen	.63		
4. Esch-sur-Alzette	.72		
5. Mersch	.73		
6. Clervaux	.82	.11	.09
7. Diekirch	.69		
8. Rédange	.64		
9. Vianden	.74		
10 Wiltz	.77		
11 Echternach	.63		
12 Grevenmacher	.71		
13 Remich	.76		

Caractéristiques du chef de ménage

Age du chef de ménage

1. < 24 ans	.66		
2. 25 à 34 ans	.75		
3. 35 à 44 ans	.71		
4. 45 à 54 ans	.70	.09	.07
5. 55 à 64 ans	.66		
6. 65 et plus	.61		

Sexe du chef de ménage

1. masculin	.71	.07	.07
2. féminin	.60		

Etat civil du chef de ménage

1. Célibataire	.66		
2. marié	.72		
3. séparé	.59		
4. divorcé	.58	.09	.08
5. veuf	.61		
6. Non réponse	(.00)		

Situation professionnelle du C.M.

1. sans emploi	.66	.05	.05
2. travaille	.71		

Etat civil selon le sexe du C.M.

1. Homme non marié	.64		
2. marié	.72		
3. Femme non mariée	.61	.09	.08
4. mariée	.46		

Situation professionnelle du C.M. selon le sexe

1.	Homme	sans emploi	.68		
2.		au travail	.72		
3.	Femme	sans emploi	.60	.08	.07
4.		au travail	.61		

Age du C.M. selon le sexe

1.	Homme	< 35 ans	.73		
2.		35 à 44 ans	.73		
3.		45 à 54 ans	.71		
4.		55 à 64 ans	.67		
5.		65 ans et +	.66	.11	.09
6.	Femme	< 35 ans	.79		
7.		35 à 44 ans	.52		
8.		45 à 54 ans	.61		
9.		55 à 64 ans	.60		
10.		65 ans et +	.55		

La lecture détaillée de l'arbre (schéma 8.1) apporte quelques informations.

Trois points peuvent être mis en évidence.

1. Un contraste très prononcé entre la métropole et les autres cantons du pays avait été observé au terme de la deuxième vague d'enquête. Il s'est effacé.

2. Une caractéristique propre aux individus avait fait son apparition au terme de la troisième vague d'enquête (l'âge des individus). Cette caractéristique intervient à nouveau au terme de la quatrième et de la cinquième vague d'enquête. Au terme de la sixième vague, elle s'impose comme la première source de contrastes entre les taux de réponse. En outre, deux caractéristiques propres aux individus, apparues en 1989, réapparaissent en 1990: l'**état civil** et le **sexe**.

   Les variables décrivant le cadre familial ne monopolisent plus l'organisation des taux de réponse. Par contre, les variables "contextuelles" ont pris une place déterminante: le **type d'habitat** et le **canton de résidence**.

   Les panels socio-économiques devraient éviter de procéder à des pondérations directes des ménages sans prendre en compte l'hétérogénéité interne de cette unité d'analyse: au fil des années les individus affirment leur autonomie par rapport à l'unité d'analyse "ménage". Leur destin ne se résoud pas dans celui du ménage.

3. L'écart entre le taux de survie maximum et le taux de survie minimum a connu une croissance notable entre la deuxième et la troisième vague,

   - il est resté constant en 1987 et en 1988
   - il s'accroît de nouveau de 1988 à 1989
   - il diminue entre 1989 et 1990.

	1986	1987	1988	1989	1990
Taux le plus élevé	1.00	.98	.94	.97	.88
Taux le plus bas	.71	.54	.48	.46	.47
Ecart	.29	.44	.46	.51	.41

TABLEAU 8.5.
Ventilation des valeurs de TREE90 dans le fichier individuel

DIA90	TREE90 n =
0. memb. panel refus	1900
1. memb. panel réponse	4108
2. nouv. memb. panel 90	-
3. non-memb. panel réponse	-
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	6008
sysmis	1144
FICHIER TOTAL	7152

Cette variable TREE90 décrit le profil des catégories d'appartenance des membres du panel, qu'ils aient répondu ou non en 1990. Elle permet donc de classer tous les membres des classes #0 et #1 de la variable DIA90 (tableau **8.5.**).

TREE90 n'exprime pas des probabilités de survie. Elle décrit simplement une typologie de sous-échantillons. Elle ne permet donc pas de calculer le poids relatif des personnes dans le nouvel échantillon. Ce sera le rôle d'une autre variable: RR90.

8.2.2. Variable instrumentale RR90

La variable RR90 est construite de la manière suivante: toute personne **ayant effectivement répondu** en 1990 (DIA90 #1), reçoit la probabilité de sélection observée dans la catégorie de population à laquelle elle appartient (TREE90) (tableau **8.6.**).

TABLEAU 8.6.

Distribution des fréquences de l'échantillon 1990, selon les taux de réponses observés en 1990 (avant et après pondération de l'échantillon initial)

RR90 taux de réponses en 1990	TREE90 code	FREQUENCES avant pondération n =	FREQUENCES après pondération (IP8590A) n =
.88	(16)	325	394.
.87	(15)	188	244.
.86	(14)	147	193.
.77	(13)	160	174.
.77	(12)	378	433.
.77	(11)	283	301.
.75	(10)	401	301.
.73	(9)	217	276.
.71	(8)	214	253.
.70	(7)	273	344.
.63	(6)	489	354.
.61	(5)	226	216.
.60	(4)	199	131.
.58	(3)	207	218.
.50	(2)	255	203.
.47	(1)	146	165.
TOTAL REPONSES		4108	4198

Il n'y a pas lieu de s'étonner si la taille de l'échantillon n'est plus identique avant et après le redressement de l'échantillon. Seule une fraction de l'échantillon des membres du panel a répondu en 1990: 4108 personnes sur 6008.

Ces 4108 personnes ont été pondérées (par IP8590A) dans le cadre de **l'échantillon total**, afin de compenser les variations des probabilités de sélection initiales. Ces poids ont été ajustés: la somme des poids est égale au nombre des individus, soit 6008. Ils n'ont pas été ajustés dans le cadre des 4108 personnes ayant effectivement participé à l'enquête (tableau **8.3.**).

La différence entre les tailles de la **fraction de l'échantillon** avant et après pondération est donc parfaitement normale. En outre, la comparaison des deux dernières colonnes du tableau **8.6.** permet de constater le fait que certaines catégories de la population auraient été sur-représentées ou sous-représentées si l'échantillon n'avait pas été pondéré avant d'être soumis à l'analyse des taux de réponses.

TABLEAU 8.7.
Ventilation des valeurs de RR90
dans le fichier individuel

DIA90	RR90 n =
0. memb. panel refus	-
1. memb. panel réponse	4108
2. nouv. memb. panel 90	-
3. non-memb. panel réponse	-
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4108
sysmis	3044
TOTAL	7152

Les valeurs de RR90 ne sont donc attribuées qu'aux seuls membres du panel ayant répondu en 1990 (tableau **8.7.**).

8.2.3. Variable instrumentale W90

La variable W90 définit le poids des membres du panel ayant répondu en 1990, **exception faite des nouveaux membres du panel entrés dans l'échantillon en 1990.**

Ces anciens membres appartiennent à différentes catégories de la population et se distinguent par la probabilité qu'ils répondent en 1990.

Lorsque le groupe enregistre un taux de réponses très faible, chaque "survivant" doit représenter un plus grand nombre de personnes présentant les mêmes caractéristiques démographiques. Lorsque le taux de réponses s'élève, l'importance de cet ajustement se réduit.

D'une manière générale, cet ajustement des poids des individus sera inversement proportionnel à leur probabilité de sélection c'est à dire, inversement proportionnel au taux de réponses enregistré dans leur groupe:

Soit: 1 / RR90

Dans le cas présent, ce calcul doit prendre en compte un élément supplémentaire: l'inégalité des probabilités de sélection initiales des membres du panel. Les nouveaux poids ne s'appliquent pas à des personnes "unités" mais à des fractions ou à des multiples de l'unité définis par la variable IP8590.

La nouvelle variable de pondération se calcule donc de la manière suivante:

$$W90 = IP8590 \times (1 / RR90)$$

TABLEAU 8.8.
Ventilation des valeurs de W90 dans le fichier individuel

DIA90	W90 n =
0. memb. panel refus	-
1. memb. panel réponse	4108
2. nouv. memb. panel 90	-
3. non-memb. panel réponse	-
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4108
sysmis	3044
FICHIER TOTAL	7152

Les valeurs de cette variable sont attribuées exclusivement aux MEMBRES DU PANEL présents dans l'échantillon en 1990, exception faite des nouveaux membres entrés dans l'échantillon en 1990 (tableau 8.8.). Le système de pondération doit donc être élargi à ces nouveaux membres.

8.2.4. Variable instrumentale W90B

La variable W90 ne permet pas encore de pondérer tout l'échantillon observé en 1990, parce qu'elle ne prend pas en compte les nouveaux membres de l'échantillon (membres ou non du panel). Il est indispensable d'étendre la portée de cette variable pour pouvoir calculer les poids des ménages.

Dans un premier temps, **les nouveaux membres de l'échantillon, non membres du panel**, reçoivent un poids nul ('0') puisqu'il est impossible de calculer leur probabilité de sélection initiale.

En effet:

- ils étaient déjà présents dans la population de 1985, lorsque l'échantillon initial a été sélectionné,

- ils sont entrés dans l'échantillon par la suite et pour des raisons variées: il est donc difficile de reconstituer l'ensemble des conditions de leur entrée dans l'échantillon et d'en dériver une logique de pondération[2].

Il est indispensable de pondérer ces non-membres avant d'attribuer des poids aux **nouveaux membres de l'échantillon, membres du panel**. En effet, certains nouveaux membres de l'échantillon, **non-membres du panel** peuvent être des ascendants directs d'**enfants nouveaux membres du panel.**

Or, dans un second temps, ces nouveaux membres du panel reçoivent un **poids égal à la moyenne des poids des ascendants directs** tels qu'ils ont été calculés au moment où le nouveau membre est entré dans l'échantillon (et dans la population).

Donc:

- lorsque les ascendants directs ont des poids différents de '0', le nouveau membre reçoit un poids égal à la moyenne de ces deux poids;

- lorsque l'un des ascendants directs a un poids nul, le nouveau membre du panel reçoit logiquement un poids égal à la moitié du poids de l'ascendant direct, membre du panel;

- lorsque les deux ascendants directs ne sont pas membres du panel, le nouveau membre reçoit un poids nul: il n'est pas membre du panel.

Par contre, un nouveau membre reçoit le poids de l'ascendant direct présent dans l'échantillon observé, lorsque l'autre ascendant est ABSENT de l'échantillon au moment où le nouveau entre dans l'échantillon: la probabilité de sélection initiale de ce nouveau membre dépend entièrement de la probabilité de sélection du seul ascendant présent dans l'échantillon. (Cfr. **2.2.1.5.** et **2.2.1.6.**).

2. Le poids relatif de ces individus est certainement différent de '0' mais toute affectation d'un poids différent de '0' reposerait sur un ensemble d'hypothèses comportant des erreurs qu'il est impossible d'estimer.

TABLEAU 8.9.
Ventilation des valeurs de W90B
dans le fichier individuel

DIA90	W90B n =
0. memb. panel refus	-
1. memb. panel réponse	4108
2. nouv. memb. panel 90	51
3. non-memb. panel réponse	576
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4735
sysmis	2417
TOTAL	7152

Ces poids sont attribués à tous les membres de l'échantillon observé en 1990, soit 4735 personnes. Parmi celles-ci, 576 personnes ont un poids égal à 0 (groupe #3 de la variable DIA90) (tableau **8.9.**).

La somme des poids est égale à 4382.0, quel que soit le nombre de personnes prises en compte: 4735 **personnes interrogées** ou 4159 **membres du panel.**

La taille "réelle" de l'échantillon est égale à 4159 membres du panel. Les 576 personnes présentes dans cet échantillon sans être membres du panel ont reçu un poids nul, précisément parce qu'il est impossible d'estimer le poids de leur présence dans cet échantillon individuel.

La somme des poids doit être ajustée à la taille de l'échantillon. La somme des poids doit être égale au nombre de membres du panel afin de sauvegarder les propriétés statistiques de l'échantillon. C'est le rôle de la variable WGT90TOT.

8.2.5. Variable opérationnelle WGT90TOT

La taille de l'échantillon pondéré par la variable W90B serait de 4382 unités. Les poids doivent donc être ajustés par une constante: la moyenne des poids attribués aux individus (2.2.1.4). Suite à cette opération, la valeur moyenne des poids sera égale à 1.0 et la somme des poids sera égale au nombre de membres du panel.

Cet ajustement est effectué sur 4159 personnes **membres du panel.** Il s'en suit que les analyses des données peuvent être réalisées de manière équivalente sur l'échantillon de 4735 **personnes observées** en 1990 ou sur l'échantillon de 4159 **membres du panel** observés en 1990. Dans les deux cas, la taille réelle de l'échantillon sera respectée (4159 unités).

Dans le premier cas le nombre total de personnes prises en compte sera **apparemment** égal à 4735, mais la somme des poids sera égale à 4159 unités. Seuls, les membres du panel sont pris en compte au cours des analyses, puisque 576 personnes ont un poids égal à 0: les valeurs de ces personnes sur les variables analysées sont annulées.

Dans le second cas, seuls les membres du panel sont pris en compte dans l'analyse et la mise à l'écart des non-membres du panel n'affecte en rien les résultats (**2.2.1.7.**).

La variable W90B devient:

WGT90TOT = W90B / (4382.008 / 4159)

TABLEAU 8.10.
Ventilation des valeurs de WGT90TOT dans le fichier individuel

DIA90	WGT90TOT n =
0. memb. panel refus	-
1. memb. panel réponse	4108
2. nouv. memb. panel 90	51
3. non-memb. panel réponse	576
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4735
sysmis	2417
TOTAL	7152

Les valeurs de la variable WGT90TOT sont attribuées à l'ensemble des personnes ayant participé à l'enquête de 1990, soit les groupes #1, #2 et #3 de la variable DIA90 (tableau **8.10.**).

8.2.6. Variable opérationnelle de pondération des ménages en 1990 MWGT90A

La variable MWGT90A permet de pondérer correctement l'échantillon des **ménages** observé en 1990.

En 1985, le biais d'échantillonnage apparaît au niveau des ménages. La correction de ce biais s'effectue, en premier lieu, au niveau des ménages. Elle est répercutée sur les membres du ménage. Ceux-ci reçoivent le même poids que le ménage auquel ils appartiennent.

Il existe donc un rapport constant entre le poids du ménage et le poids des individus: le poids du ménage est égal à la valeur moyenne des poids des personnes qui le composent.

Depuis 1986, l'échantillon subit l'influence de facteurs d'évolution qui introduisent des biais aux deux niveaux d'analyse.

Tantôt un ménage complet disparaît et la raison du refus est commune à l'ensemble des membres du ménage.

Tantôt, un seul membre disparaît pour des raisons strictement individuelles.

Des nouveaux membres entrent dans les ménages. Les uns sont pris en compte au titre de membres du panel, les autres ne sont pas reconnus comme membres du panel.

De nouveaux ménages se créent et, généralement, provoquent l'inclusion d'une nouvelle personne étrangère à l'échantillon initial. Des ménages se défont et provoquent tantôt l'apparition de deux nouveaux ménages, tantôt la création d'un nouveau ménage et le retour d'une personne dans son milieu d'origine.

Cette liste n'est pas exhaustive. Elle illustre suffisamment à quel point les trajectoires des personnes et les trajectoires des ménages se différencient progressivement. Elle invite à fonder l'analyse de l'évolution de l'échantillon au niveau des trajectoires individuelles plutôt qu'au niveau des ménages.

Cependant, le rapport entre la valeur des poids relatifs individuels et la valeur des poids des ménages garde sa propriété initiale: le poids du ménage est égal à la valeur moyenne des poids des individus.

La procédure est simple.

* Les valeurs des poids individuels sont agrégées au niveau du ménage.

* La valeur moyenne des poids individuels (WGT90TOT) détermine la valeur du poids du ménage (MWGT90T).

  Le poids du ménage est calculé en fonction du **nombre de personnes ayant répondu** et non en fonction du nombre de personnes présentes dans le ménage.

* Cette valeur peut être ajustée en fonction de la taille de l'échantillon des ménages (MWGT90A).

En 1990, 40 ménages sont entièrement composés par des personnes qui ne sont pas membres du panel: ils reçoivent logiquement un poids de '0' (valeur moyenne des poids de leurs membres).

L'ajustement des poids des ménages s'effectue donc sur un nombre de ménages égal à 1636 unités et non sur 1636 + 40 ménages. De cette façon, les analyses seront toujours effectuées sur 1636 unités, indépendamment du fait que les ménages composés exclusivement de non-membres du panel sont pris en compte ou sont mis à l'écart (**2.2.1.7.**). A défaut d'un tel ajustement, l'échantillon pondéré par MWGT90T compterait 1554 unités. D'où:

MWGT90A = MWGT90T / (1554.0 / 1636)

TABLEAU 8.11.
Ventilation des valeurs de MWGT90A dans le fichier individuel

DIA90	MWGT90A n =
0. memb. panel refus	-
1. memb. panel réponse	4108
2. nouv. memb. panel 90	51
3. non-memb. panel réponse	576
4. non-memb. panel refus	-
5. émigrés / décédés	-
total	4735
sysmis	2417
TOTAL	7152

Les valeurs des poids des ménages (MWGT90T et MWGT90A) sont reventilées sur les individus. Chaque membre du ménage reçoit la valeur du poids du ménage auquel il appartient. Les non-membres du panel reçoivent, cette fois, un poids différent de 0, à condition qu'ils n'appartiennent pas à un ménage composé exclusivement de non-membres du panel. Cette procédure, purement technique, permet de regrouper l'ensemble des poids dans un seul fichier individuel.

Le tableau **8.11.** présente les sous-échantillons individuels qu'il convient de sélectionner dans le fichier au moment de reconstituer l'échantillon des ménages observés en 1990. Les poids des ménages s'obtiennent aisément, par exemple, en agrégeant les valeurs de MWGT90A au niveau des ménages et en prenant la moyenne comme valeur de pondération. (Certains logiciels permettent de prendre la valeur du premier membre du ménage comme valeur de référence (SPSSX)).

Il peut être utile de rappeler que TOUS les membres appartenant à l'échantillon observé en 1990 doivent être pris en compte au moment de reconstituer l'échantillon des ménages. En l'absence des non-membres du panel, il serait impossible de calculer certaines variables propres au ménage (par exemple, la taille du ménage).

ANNEXE 8 A

TABLEAU COMPARATIF DES CARACTERISTIQUES DEMOGRAPHIQUES DES MEMBRES DU PANEL DE 1985 A 1990

(après pondération)

- caractéristiques personnelles
- caractéristiques du ménage
- caractéristiques du chef de ménage

Caractéristiques Personnelles	1985 (N=6110)	1986 (N=5245)	1987 (N=4648)	1988 (N=4477)	1989 (N=4291)	1990 (N=4159)
Sexe						
1. homme	48.8	48.7	48.2	48.0	48.1	48.1
2. femme	51.2	51.3	51.8	52.0	51.9	51.9
Age						
1. < 18 ans	24.1	23.5	23.6	22.9	22.5	22.4
2. 18 - 24 ans	11.2	11.2	11.1	11.0	11.4	10.7
3. 25 - 34 ans	15.6	15.1	15.2	14.6	15.0	14.8
4. 35 - 44 ans	14.3	14.3	14.6	15.1	14.4	14.5
5. 45 - 54 ans	12.7	12.8	12.6	13.0	13.4	13.0
6. 55 - 64 ans	10.4	10.9	10.5	11.5	10.9	10.8
7. > 64 ans	11.7	12.1	12.4	11.9	12.5	13.7
Age par Sexe						
1. homme < 18 ans	12.7	12.4	12.0	11.6	11.2	11.1
2. homme 18 - 64 ans	31.7	31.8	31.5	32.0	32.4	32.0
3. homme > 64 ans	4.4	4.6	4.7	4.4	4.5	5.0
4. femme < 18 ans	11.4	11.2	11.6	11.2	11.3	11.2
5. femme 18 - 64 ans	32.6	32.6	32.5	33.2	32.6	31.9
6. femme > 64 ans	7.2	7.5	7.7	7.5	8.0	8.7
Etat civil						
1. célibataire	39.8	39.9	40.0	40.0	40.6	40.3
2. marié	49.6	49.2	48.8	48.9	48.5	48.3
3. séparé	1.0	0.8	0.8	0.9	0.7	1.0
4. divorcé	1.9	1.9	2.3	2.0	2.0	2.1
5. veuf	7.9	8.0	8.1	8.2	8.2	8.3
6. N.R.	0.0	-	-	-	-	-

Nationalité

1. Luxembourgeois	76.1	76.4	77.9	78.7	79.4	79.7
2. autres C.E.	21.3	21.3	20.3	19.7	18.8	18.5
3. hors C.E.	2.5	2.3	1.9	1.6	1.3	1.8
4. N.R.	0.1	-	-	-	-	-

Emploi

1. oui	38.1	39.8	37.1	40.3	39.8	39.3
2. non	35.4	35.3	34.2	34.6	35.1	35.6
3. enfants	26.5	24.9	28.7	25.1	25.1	25.1

Adulte/enfant

1. adulte	73.5	75.1	74.7	74.9	74.9	74.9
2. enfant	26.5	24.9	25.3	25.1	25.1	25.1

Lien avec le chef de ménage

1. chef de ménage	36.4	37.2	36.6	37.8	37.1	37.4
2. épouse	24.2	24.1	23.9	23.8	23.7	23.9
3. "épouse"	1.1	1.0	0.9	1.1	1.0	0.9
4. ami, amie	0.3	0.1	0.2	0.1	0.1	0.1
5. fils, fille	33.4	33.6	34.2	33.5	34.2	33.8
6. autres	4.6	4.0	4.2	3.7	4.0	3.9

Caractéristiques du ménage

Type d'habitation

1. maison rurale	11.4	10.7	9.3	7.5	8.2	7.7
2. unifamiliale individuelle	33.7	34.7	34.7	33.5	33.0	34.0
3. unifamiliale jumelée	13.5	8.8	9.1	10.3	10.4	10.3
4. unifamiliale en série	19.5	21.4	25.3	26.5	26.6	25.3
5. petit collectif	12.6	13.1	13.1	12.4	11.3	13.3
6. collectif moyen	7.1	7.3	5.8	6.2	7.0	6.3
7. grand collectif	1.1	1.4	1.1	0.8	1.1	0.7
8. précaire	0.3	0.0	0.1	0.1	0.0	0.1
9. N.R.	0.8	2.6	1.7	-	1.8	2.1
0. maison de retraite	-	-	-	0.3	0.5	0.2

Canton

1. Luxembourg-ville	17.6	17.6	17.7	16.2	15.9	15.4
2. Luxembourg-campagne	8.9	8.8	8.8	9.3	9.2	9.0
3. Capellen	6.9	7.0	6.9	7.1	7.2	7.7
4. Esch sur Alzette	33.5	33.6	34.3	33.7	33.4	33.7
5. Mersch	4.6	4.3	3.7	4.2	4.6	4.2
6. Clervaux	2.3	2.3	2.5	2.5	2.5	2.7
7. Diekirch	6.2	6.2	6.7	6.7	6.5	5.8
8. Redange	3.0	3.0	2.8	3.0	2.9	3.3
9. Vianden	0.8	0.7	0.8	0.8	0.8	0.7
10. Wiltz	3.0	3.1	3.0	3.0	3.1	3.3
11. Echternach	3.5	3.5	3.5	3.4	3.9	3.8
12. Grevenmacher	5.8	5.8	5.7	6.3	5.9	6.1
13. Remich	4.0	4.0	3.6	3.9	4.0	4.1

Nombre de personnes dans le ménage

1.	7.6	8.0	8.0	8.5	8.4	8.9
2.	21.0	20.6	19.6	21.1	20.7	20.1
3.	22.5	23.1	22.8	22.1	22.7	23.6
4.	27.6	27.6	28.7	28.6	28.0	27.7
5.	12.3	12.7	12.9	12.7	12.9	12.7
6.	5.8	5.3	4.9	5.0	4.6	5.0
7.	2.1	1.8	2.1	1.2	1.7	1.3
8.	0.7	0.6	0.7	0.7	1.0	0.7
9.	0.3	0.3	0.2	-	-	-
11.	0.0	-	0.1	-	-	-
12.	-	0.1	-	0.1	-	-

Nombre d'enfants dans le ménage

0.	39.2	41.9	41.7	43.4	43.4	43.8
1.	24.3	23.2	23.6	21.2	22.9	22.3
2.	22.9	23.2	22.8	24.1	21.6	22.4
3.	9.9	9.5	9.1	8.7	9.8	8.5
4.	3.5	2.2	2.7	2.5	2.0	2.7
5.	0.1	0.1	0.1	0.1	0.3	0.3
6.	0.1	-	-	-	-	-

Nombre d'adultes dans le ménage

0.	0.1	-	-	-	-	-
1.	9.9	10.2	10.4	11.0	10.7	11.4
2.	63.0	59.5	58.4	60.3	60.3	59.5
3.	16.2	18.9	18.9	17.4	16.5	17.4
4.	7.7	6.9	7.8	7.8	8.3	8.2
5.	2.1	3.0	2.8	2.4	2.5	2.2
6.	0.8	1.2	1.4	0.8	1.2	0.7
7.	0.2	0.2	0.2	0.3	0.3	0.5
9.	-	0.1	0.1	-	0.1	0.0

Nombre d'emplois dans le ménage

0.	17.1	17.5	17.4	17.3	18.1	18.2
1.	46.9	43.1	42.1	43.4	42.7	43.1
2.	29.2	29.8	30.0	28.9	28.6	28.5
3.	4.9	6.8	6.9	7.1	6.9	7.1
4.	1.4	2.1	2.5	2.6	2.7	2.5
5.	0.5	0.7	0.9	0.6	0.8	0.3
6.	0.0	0.0	0.1	0.1	0.4	0.3

Caractéristiques du chef de ménage

Sexe

1. homme	87.2	86.0	86.3	83.6	85.6	85.9
2. femme	12.8	14.0	13.7	16.4	14.4	14.1

Age

1. < 24 ans	2.6	2.3	1.8	2.1	2.1	1.7
2. 25 - 34 ans	18.3	18.1	18.7	17.7	17.0	16.2
3. 35 - 44 ans	27.6	27.1	27.3	26.6	26.8	26.5
4. 45 - 54 ans	24.4	23.7	23.0	23.7	23.9	23.9
5. 55 - 64 ans	14.8	15.9	15.8	16.9	16.6	17.1
6. > 64 ans	12.2	12.9	13.4	13.1	13.5	14.7

Etat civil

1. célibataire	6.0	6.1	6.5	6.2	6.2	6.4
2. marié	80.6	80.0	79.6	79.2	79.5	79.1
3. séparé	1.4	1.7	1.2	1.3	0.9	1.3
4. divorcé	2.5	2.7	2.7	3.0	3.1	3.1
5. veuf	9.5	9.6	10.0	10.3	10.3	10.1
6. N.R.	0.0	-	-	-	-	-

Emploi

1. oui	74.1	73.3	73.3	72.0	70.9	70.6
2. non	25.9	26.7	26.7	28.0	29.1	29.4

Etat civil par Sexe

1. homme non-marié	7.0	6.7	7.0	6.2	6.5	7.3
2. homme marié	80.2	79.3	79.3	77.4	79.1	78.6
3. femme non-mariée	12.4	13.4	13.4	14.6	13.9	13.6
4. femme mariée	0.4	0.6	0.3	1.8	0.4	0.4

Emploi par Sexe

1. homme sans emploi	17.5	17.4	17.6	19.0	20.0	20.3
2. homme avec emploi	69.8	68.6	68.7	64.6	65.6	65.6
3. femme sans emploi	8.5	9.3	9.1	9.0	9.1	9.1
4. femme avec emploi	4.3	4.7	4.6	7.4	5.3	5.0

TABLEAUX DE SYNTHESE

DES VARIABLES DE PONDERATION ET DES ECHANTILLONS PERTINENTS

1985

Variable de sélection	Variables de pondération			
DIA	MIREV	PONREV	WGT85	MWGT85A
0. memb. 85/refus 86	*	*	*	*
1. memb. 85/memb. 86	*	*	*	*
2. nouv. memb.-panel 86				
3. nouv. memb. non-panel 86				
4. memb. 85/émigrés, décédés 86	*	*	*	*

1986

Variable de sélection	Variables de pondération opérationnelles		
DIA	WGT86	WGT86TOT	MWGT86A
0. memb. 85/refus 86			
1. memb. 85/memb. 86	*	*	*
2. nouv. memb.-panel 86		*	*
3. nouv. memb. non-panel 86		*	*
4. memb. 85/émigrés, décédés 86			

1987

Variable de sélection	Variables de pondération opérationnelles	
DIA87	WGT87TOT	MWGT87A
0. membres panel refus		
1. membres panel réponses	*	*
2. nouv. memb.-panel 87	*	*
3. non membres panel réponse	*	*
4. non membres panel refus		
5. émigrés / décédés		

1988

Variable de sélection	Variables de pondération opérationnelles	
DIA88	WGT88TOT	MWGT88A
0. membres panel refus		
1. membres panel réponses	*	*
2. nouv. memb.-panel 88	*	*
3. non membres panel réponse	*	*
4. non membres panel refus		
5. émigrés / décédés		

1989

Variable de sélection	Variables de pondération opérationnelles	
DIA89	WGT89TOT	MWGT89A
0. membres panel refus		
1. membres panel réponses	*	*
2. nouv. memb.-panel 89	*	*
3. non membres panel réponse	*	*
4. non membres panel refus		
5. émigrés / décédés		

1990

Variable de sélection	Variables de pondération opérationnelles	
DIA90	WGT90TOT	MWGT90A
0. membres panel refus		
1. membres panel réponses	*	*
2. nouv. memb.-panel 90	*	*
3. non membres panel réponse	*	*
4. non membres panel refus		
5. émigrés / décédés		

Liste des publications du panel socio-économique des ménages

"Liewen zu Lëtzebuerg"

CEPS/Instead, Walferdange, Grand-Duché de Luxembourg

- **Méthodologie générale & répertoire des variables** - Année d'enquête: 1985 (Première vague). Document PSELL n° 1 - P. DICKES, P. HAUSMAN, A. KERGER -1987.

- **L'état de la pauvreté au Grand-Duché de Luxembourg.** Document PSELL n° 2 - F. COURTOIS, P. HAUSMAN (1987).

- **Description des niveaux de vie et de bien-être économique dans les ménages résidant au Luxembourg - Année 1985-1987.** Document PSELL n° 3 P. HAUSMAN (1987).

- **Niveaux de vie et de bien-être économique des ménages en 1985: principaux résultats en 1985.** Document PSELL n° 4 P. HAUSMAN (1987).

- **Un indicateur pour mesurer la pauvreté objective: théorie et application dans la première vague du panel socio-économique Luxembourgeois - 1985.** Document PSELL n° 5, P. DICKES (1987).

- **Un indicateur pour mesurer la pauvreté subjective. Théorie et application dans la première vague du panel socio-économique luxembourgeois : année d'enquête 1985 (première vague).** Document PSELL n° 6 P. DICKES (1987).

- **Pratique de l'échelonnement multidimensionnel.** Document PSELL n° 7 P. DICKES, J. TOURNOIS (1989).

- **Pauvreté et conditions d'existence: théories, modèles et mesures.** Document PSELL n° 8. P. DICKES (1989).

- **Logistique & documentation - Principes d'organisation de la documentation dans le panel.** Document PSELL n° 9 J. TOURNOIS (1988).

- **Documentation transversale des variables 1985: première vague.** Document PSELL n° 10. J. TOURNOIS (1988).

- **Evolution d'un groupe de ménages pauvres entre 1985 et 1987.** Document PSELL n° 11. A. WAGNER (1989).

- **Description statistique des variables du questionnaire -1986-** (deuxième vague). Document PSELL n° 12. A. KERGER, R. DE WEVER (1988).

- **Activité féminine, isolement et prestations familiales: un premier parallèle Luxembourg-Lorraine - Annexes.** Document PSELL n° 13. J.C.RAY, B.JEANDIDIER, S.CARVOYEUR (1990).

- **Le mode d'échantillonnage du panel "Liewen zu Lëtzebuerg" - Bilan des deux premières vagues.** Document PSELL n°14. P.HAUSMAN (1990).

- **Analyse des données irlandaises (enquête pilote -1987) pour construire une échelle de pauvreté.** Document PSELL n° 15. P.DICKES (1988).

- **La collecte des données en 1986** - Elaboration du questionnaire, déroulement de l'enquête, opérations de chiffrement. Document PSELL n° 16. A.KERGER (1989).

- **Organisation der Daten des Luxemburger Haushaltspanels.** (Eingabe, Speicherung und Analyse von Paneldaten). Document PSELL n° 17. G.SCHMAUS (1990) - (version anglaise: 17a).

- **MNDr, partition valuée selon la méthode de Roubens et Libert.**
Document PSELL n° 18. B.GAILLY (1989).

- **Dispositif des pondérations individuelles et des pondérations des ménages en 1985 et 1986.**
Document PSELL n° 19; B.GAILLY, P.HAUSMAN (1990).

- **Les modes de garde des jeunes enfants.**
Document PSELL n° 20. A.AUBRUN, P.HAUSMAN (1990).

- **Les indicateurs sociaux de pauvreté:** Tableaux de base et documentation.
Document PSELL n° 21. P.HAUSMAN (1990).

- **Les personnes âgées et/ou retraitées au Luxembourg: leur environnement familial et leurs réseaux de solidarité.** Document PSELL n° 22. P.HAUSMAN, G.SCHABER (1991).

- **Examen des effets du phénomène d'attrition sur l'étude des revenus et de l'emploi - Années de références: 1985, 1986 et 1987.** Document PSELL n° 23 P.HAUSMAN, B.GAILLY (1990).

- **La constitution des fichiers de référence, nécessaire à l'étude du phénomène d'attrition.**
Document PSELL n° 24 R.DE WEVER (1990).

- **Dispositif des pondérations individuelles et des pondérations des ménages en 1985 et 1987.**
Document PSELL n° 25. B.GAILLY, P.HAUSMAN (1990).

- **Bilan de l'attrition au cours des trois premières vagues d'enquêtes:1985/1986/1987.**
Document PSELL n° 26 B.GAILLY, P.HAUSMAN (1990).

- **Imputation des revenus manquants dans le panel socio-économique luxembourgeois.**
Document PSELL n°27 P.HAUSMAN (1990).

- **"PSELLDOC" Système documentaire pour le panel Luxembourgeois.**
Document PSELL n°28. J.J.WESTER avec la collaboration de A.AUBRUN (1990).

- **Le déroulement de la collecte en 1987. Elaboration du questionnaire, déroulement de l'enquête, opération de chiffrement.** Document PSELL n°29. A.KERGER (1990).

- **La production des données: Vague 1988-1990.** Document PSELL n° 30. A.KERGER (1990) .

- **Description statistique des variables du questionnaire -1987-** Troisième vague.
Document PSELL n° 31. R.DE WEVER, A.KERGER (1991).

- **Les ménages de retraités et les ménages d'actifs** - Comparaison des niveaux de vie et des niveaux de dépenses. Document PSELL n° 34. P.HAUSMAN, A.AUBRUN, A.KERGER (1991).

- **Situation der Arbeitslosen in Luxemburg.** Document PSELL n° 35. G.SCHMAUS(1991).

- **La recherche sur la pauvreté au Grand-Duché de Luxembourg.**
Document PSELL 36. A.WAGNER (1990).

- **La distribution des revenus entre ménages en 1986** - Une comparaison de statistiques Luxembourg - Lorraine - Belgique. Document PSELL n° 38. B.DELVAUX (1991).

- **Efficacité de la sécurité sociale dans la lutte contre la pauvreté.**
Document PSELL n°39. HAUSMAN P.

- **Dispositif des pondérations individuelles et des pondérations des ménages de 1985 à 1988.**
Document PSELL n° 40. B.GAILLY (1991).

- **Projet de recommandation du conseil portant sur les critères communs relatifs à des ressources et prestations suffisantes dans les systèmes de protection sociale. Rapport préparatoire relatif au dispositif de suivi.** Document PSELL n° 42. J.C.BROWN, A.WAGNER avec la collaboration de: P. HAUSMAN, A.KERGER, G.MENARD (1991).

- **La loi sur le Revenu Minimum Garanti. Quelques avis du public.**
Document PSELL n° 43. A.WAGNER (1992).

- **L'endettement des ménages au Luxembourg 1985 -1989.**
Document PSELL n° 45. B.GAILLY, P.HAUSMAN (1992).

- **Les femmes au Grand-Duché de Luxembourg: Démographie-Famille I.**
Document PSELL n°46. P.HAUSMAN avec la collaboration de M. Langers (STATEC) et du Ministère de la Famille et de la Solidarité (1992).

- **Les femmes au Grand-Duché de Luxembourg: Revenus-Conditions de vie II.**
Document PSELL n° 47. P.HAUSMAN, J.VECERNIK avec la collaboration du Ministère de la Famille et dela Solidarité (1992).

- **Dispositif des pondérations individuelles et des pondérations des ménages de 1985 à 1989.**
Document PSELL n°48. B.GAILLY (1992).

- **Luxembourg, 1985 à 1989, une vague d'endettement.**
Document PSELL n°49. B.GAILLY, P.HAUSMAN (1993).

- **Les Effets de la Protection Sociale dans la Communauté Européenne.** Contribution au programme "Convergence en matière de Protection Sociale" (Commission C.E., DG V).
Document PSELL n°50. P.HAUSMAN (1993) .

- **Situation démographique de l'Europe des Douze -EUR12.**
Document PSELL n°51. P.HAUSMAN (1993).

- **Les phénomènes associés au vieillissement de la population.**
Document PSELL N°52. P.HAUSMAN (1993).

- **Dispositif des pondérations individuelles et des pondérations des ménages de 1985 à 1990.**
Document PSELL n°53. B.GAILLY (1993).

- **Insérer des nouveaux membres dans un panel longitudinal de ménages et d'individus: simulations.** Document PSELL n°54. B.GAILLY (CEPS/Instead) avec la collaboration de P.LAVALLÉE (Statistics-Canada) (1993).

PANEL SOCIO-ECONOMIQUE
"LIEWEN ZU LËTZEBUERG"

Document PSELL No 19

DISPOSITIF DES PONDERATIONS INDIVIDUELLES ET DES PONDERATIONS DES MENAGES EN 1985 ET 1986

Première version
mars 1990

B.Gailly
P.Hausman

PANEL SOCIO-ECONOMIQUE
"LIEWEN ZU LËTZEBUERG"

Document PSELL No 25

DISPOSITIF DES PONDERATIONS INDIVIDUELLES ET DES PONDERATIONS DES MENAGES DE 1985 A 1987

B.Gailly
P.Hausman

PANEL SOCIO-ECONOMIQUE
"LIEWEN ZU LËTZEBUERG"

Document PSELL No 40

DISPOSITIF DES PONDERATIONS INDIVIDUELLES ET DES PONDERATIONS DES MENAGES DE 1985 A 1988

B.Gailly

PANEL SOCIO-ECONOMIQUE
"LIEWEN ZU LËTZEBUERG"

Document PSELL No 48

DISPOSITIF DES PONDERATIONS INDIVIDUELLES ET DES PONDERATIONS DES MENAGES DE 1985 A 1989

B.Gailly

Document produit par le

CENTRE D'ETUDES DE POPULATIONS, DE PAUVRETE
ET DE POLITIQUES SOCIO-ECONOMIQUES

C.E.P.S. / INSTEAD
B.P.65, L-7201 Walferdange
Tél. (352) 33 32 33 - 1

Président: Gaston Schaber

1992

PANEL SOCIO-ECONOMIQUE
"LIEWEN ZU LËTZEBUERG"

Document PSELL No 23

EXAMEN DES EFFET DU PHENOMENE D'ATTRITION SUR L'ETUDE DES REVENUS ET DE L'EMPLOI

Années de référence: 1985, 1986 et 1987

B.Gailly
P.Hausman

Document produit par le

CENTRE D'ETUDES DE POPULATIONS, DE PAUVRETE
ET DE POLITIQUES SOCIO-ECONOMIQUES

C.E.P.S. / INSTEAD
B.P.65, L-7201 Walferdange
Tél. (352) 33 32 33 - 1

Président: Gaston Schaber

1990

PANEL SOCIO-ECONOMIQUE
"LIEWEN ZU LËTZEBUERG"

Document PSELL No 26

BILAN DE L'ATTRITION AU COURS DES TROIS PREMIERE VAGUES D'ENQUETE: 1985, 1986, 1987

B.Gailly
P.Hausman

Document produit par le

CENTRE D'ETUDES DE POPULATIONS, DE PAUVRETE
ET DE POLITIQUES SOCIO-ECONOMIQUES

C.E.P.S. / INSTEAD
B.P.65, L-7201 Walferdange
Tél. (352) 33 32 33 - 1

Président: Gaston Schaber

1990