

PANEL SOCIO-ECONOMIQUE

"LIEWEN ZU LETZEBUERG"

---

Document PSELL N° 19

DISPOSITIF DES PONDERATIONS  
INDIVIDUELLES ET DES  
PONDERATIONS DES MENAGES  
EN 1985 ET 1986

Première version  
mars 1990



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 33-228

Président: Gaston Schaber

---

1 9 9 0

\*\*\*\*\*

**DISPOSITIF DES PONDERATIONS INDIVIDUELLES  
ET DES PONDERATIONS DES MENAGES  
EN 1985 ET 1986**

**première version**

**Mars 1990**

\*\*\*\*\*

**B.Gailly**

**P.Hausman**

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, F. Hentges, A. Kerger, G. Schmaus et J.J. Wester.

# S o m m a i r e

Introduction.....	3
Chapitre 1 : DEUX RAISONS DE PONDERER L'ECHANTILLON .....	6
Chapitre 2 : ORGANISATION DU FICHIER.....	9
2.1. Les individus membres et non-membres du panel	10
2.2. La variable "diagnostic"	10
2.3. Les variables de pondération	13
Chapitre 3 : LES VARIABLES DE PONDERATION Echantillon 1985 .....	16
3.1. Variable instrumentale "MIREV"	17
3.2. Variable opérationnelle de pondération des ménages "MWGT85A"	19
3.3. Variable opérationnelle de pondération individuelle en 1985 "WGT85"	20
ANNEXE : Modification de la distribution des individus, en fonction du nombre de titulaires principaux dans le ménage.....	23
Chapitre 4 : LES VARIABLES DE PONDERATION Echantillon 1986 .....	27
4.1. Variable instrumentale "TREE"	29
4.2. Variable instrumentale "W"	37
4.3. Variable instrumentale "W86"	40
4.4. Variable opérationnelle "WGT86"	41
4.5. Variable instrumentale "W86B"	42
4.6. Variable opérationnelle "WGT86TOT"	45
4.7. Variable opérationnelle "MWGT86A"	46
ANNEXE : Exemple de l'évolution des individus et des ménages .....	49
Tableau de Synthèse .....	52

\*\*\*\*\*

## DISPOSITIF DES PONDERATIONS INDIVIDUELLES

### ET DES PONDERATIONS DES MENAGES

EN 1985 ET 1986

première version

Mars 1990

\*\*\*\*\*

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.

Le mode de calcul des variables de pondération ne sera pas toujours précisé de manière détaillée. Ce premier document a une vocation essentiellement utilitaire. Les procédures de calcul seront plus documentées dans des versions ultérieures.

Cette première 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 de ce document de base. 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. Il sera mis à jour chaque fois que des nouvelles variables seront calculées afin de prendre en compte un nouvel état de l'échantillon.

## 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 peut être utile de décrire brièvement l'organisation du fichier individuel où les variables de pondération sont répertoriées.



## Chapitre 2

### ORGANISATION DU FICHER

---

Le répertoire des variables de pondération réunit l'ensemble des personnes interrogées au moins une fois au cours du panel (1) et l'ensemble des variables de pondération (2).

#### 1. Les personnes

- 1.1. Différentes catégories de personnes peuvent être distinguées en fonction de leur carrière au sein du panel, mais une première distinction doit être clarifiée: tous les individus ne sont pas MEMBRES du PANEL ("FULL MEMBER PANEL").
- 1.2. Des distinctions supplémentaires permettent de sélectionner les catégories de personnes en fonction de leur place dans le système de pondération. Elles interviennent dans le calcul des variables de pondération en fonction de l'influence qu'elles exercent sur l'évolution de l'échantillon.
- 1.3. La variable DIA(-gnostic) permet d'effectuer différentes sélections et différentes combinaisons de ces catégories de personnes afin de constituer l'échantillon souhaité.

#### 2. Les variables

Les variables de pondération répertoriées dans le fichier peuvent correspondre à des usages différents. On distinguera essentiellement les variables "opérationnelles" et les variables "instrumentales".

## **2.1. Les individus membres et non-membres du panel**

Tous les individus ne sont pas nécessairement membres du panel. Un individu est MEMBRE du PANEL lorsqu'il remplit l'une des deux conditions suivantes:

- il appartient à l'échantillon initial du panel, tiré en 1985 :

6110 personnes sont ainsi des individus-panels ou membres du panel (1ère vague);

- il est entré dans l'échantillon du panel au cours d'une vague d'enquête ultérieure, mais il est un descendant direct d'un individu-panel :

en 1986, 78 individus sont entrés dans l'échantillon au titre de fils ou de fille d'un père ou d'une mère appartenant à l'échantillon initial; ils sont nés en 1986 ou ils sont rentrés dans leur famille après une absence de plus ou moins longue durée.

Tout autre individu n'ayant pas appartenu à l'échantillon initial est considéré comme NON-MEMBRE du panel. Il peut s'agir de l'ascendant d'un individu membre du panel, du conjoint d'un individu-panel ou de tout autre personne appartenant ou non à la famille:

en 1986, 146 personnes sont entrées, à ce titre, dans l'échantillon.

Tous les membres d'un échantillon ne sont donc pas nécessairement membres du panel.

Ces distinctions seront importantes au moment d'attribuer des valeurs de pondération à des personnes dont la probabilité de sélection initiale (en 1985) est inconnue.

## **2.2. La variable "diagnostic"**

Cette variable a une utilité essentiellement pratique.

Toutes les variables de pondération sont répertoriées dans un fichier INDIVIDUEL.

Toutes les variables ne sont pas pertinentes pour toutes les catégories de personnes recensées dans le fichier.

Certaines variables ne sont utilisables qu'après avoir subi quelques transformations.

La variable DIA permet de sélectionner aisément les sous-échantillons que l'on désire analyser et de choisir la variable de pondération adéquate. La pertinence des variables fera l'objet des développements du chapitre suivant.

**TABLEAU 1**  
**Les sous-échantillons: Code, Label et Nombre de personnes**

DIA(-gnostic)	échant.85	échant.86	Total
<b>CODES</b>	<b>N =</b>	<b>N =</b>	<b>N =</b>
0. memb. 85/refus 86	833	-	833
1. memb. 85/memb. 86	5167	5167	5167
2. nouveaux memb.-panel 86		78	78
3. nouveaux memb. non-panel 86		146	146
4. memb. 85 décédés/émigrés 86	110	-	110
TOTAL	6110	5391	6334

Le tableau 1. présente les premiers points de repère: la liste des sous-échantillons disponibles et le nombre total de personnes qui correspond à chaque catégorie.

La variable DIA permet de sélectionner trois types d'échantillons principaux:

- l'ensemble des membres d'un échantillon annuel
- l'ensemble des individus-panels
- ou, encore, toute combinaison de sous-échantillons qui pourrait s'avérer utile.

Le premier type d'échantillons permet d'effectuer des analyses synchroniques, tandis que le second se prête aux analyses longitudinales.

## L'échantillon 1985

La sélection des groupes #0, #1 et #4 réunit l'ensemble des membres de l'échantillon interrogés en 1985, soit 6110 individus MEMBRES du PANEL. Des valeurs de pondération peuvent être CALCULEES pour ces individus, en 1985, parce que leur probabilité de sélection initiale est connue. Cette sélection permet donc de saisir l'échantillon de 1985 et la variable de pondération adéquate.

## L'échantillon 1986

La sélection des groupes #1, #2 et #3 définit l'ensemble des personnes interrogées en 1986. Parmi ces 5391 MEMBRES de L'ECHANTILLON 1986, seuls 5245 sont MEMBRES du PANEL:

Les membres du groupe #1 ont accepté de répondre en 1985 ET en 1986. Ils sont MEMBRES du PANEL: leur probabilité de sélection initiale est connue. La probabilité qu'ils répondent en 1986 peut être calculée. Il est donc possible de CALCULER leur " poids " en 1986.

Les membres du groupe #2 sont les descendants directs d'un père ou d'une mère membre du panel. Ils sont nés ou rentrés au foyer entre les deux périodes d'enquête. Ils sont MEMBRES du PANEL. Leur probabilité de sélection initiale et la probabilité qu'ils entrent dans l'échantillon en 1986 ne peuvent pas être calculées. Un poids leur sera néanmoins ATTRIBUE en 1986 en fonction du principe suivant: leur probabilité d'apparaître dans l'échantillon, au titre de membres du panel, est entièrement déterminée par les probabilités de sélection de leur ascendant direct. Par conséquent, ils reçoivent le poids de leur ascendant (membre du groupe #1).

Les membres du groupe #3 sont des nouveaux individus NON MEMBRES du PANEL. Leur probabilité de sélection en 1985 est inconnue. Leur probabilité d'apparaître dans l'échantillon en 1986 est indéterminée. Elle ne dépend pas de leurs liens avec d'autres membres du ménage. Il est donc malaisé de leur attribuer un poids. Ils reçoivent un poids individuel dont la valeur est '0'. Ils ne seront donc pas pris en compte lorsqu'une analyse portera sur l'échantillon de 1986 pondéré. Ceci ne signifie pas qu'ils seront sans utilité. Ils seront pris en compte:

- lorsque l'analyse des données portera sur l'échantillon individuel de 1986 non-pondéré;

- lorsque l'analyse portera sur des données individuelles temporelles, à condition que le début de la période analysée soit fixé en 1986.

- au moment de calculer les pondérations des ménages.

## **L'échantillon longitudinal**

La sélection du groupe #1 permet de saisir l'ensemble des individus interrogés à deux reprises. Cet échantillon permet d'effectuer des analyses de données individuelles **TEMPORELLES**, ou analyses longitudinales. Il est composé exclusivement de membres du panel.

Des pondérations annuelles peuvent être calculées pour ce groupe puisque la probabilité de sélection initiale de tous ses membres est connue.

## **Les échantillons de ménages**

La sélection des groupes #0 ,#1 et #4 permet de reconstituer l'ensemble des personnes interrogées en 1985.

La sélection des groupes #1, #2 et #3 permet de reconstituer l'échantillon des personnes interrogées en 1986.

Les membres d'un ménage sont désignés par un numéro "ménage" identique. Il est donc possible de procéder à l'agrégation des valeurs individuelles au niveau des ménages et, inversement, de ventiler les valeurs d'un ménage sur chacun de ses membres.

Les variables de pondération calculées pour les ménages ont été reventilées sur les membres du ménage afin de rassembler toutes les valeurs de pondération dans un seul fichier individuel. Il est donc aisé de reconstituer la valeur du poids du ménage en agrégeant au niveau du ménage les valeurs ventilées sur ses membres.

### **2.3. Les variables de pondération**

Toutes les variables qui ont été ou qui seront utiles au calcul des pondérations ou à l'exécution d'une procédure de pondération sont répertoriées dans un seul fichier individuel.

Il faut distinguer trois types de variables:

- les variables individuelles "opérationnelles" peuvent être utilisées sans transformation préalable. Leur usage présuppose que l'échantillon adéquat a été sélectionné. Elles permettent de limiter les effets des biais qui affectent les échantillons;
- les variables "instrumentales" ont permis de calculer les variables opérationnelles;
- les variables opérationnelles de pondération des ménages ne peuvent être utilisées qu'après l'agrégation des valeurs individuelles au niveau des ménages.

Dans cette présentation, chaque variable sera définie par le rôle qu'elle est appelée à jouer, soit dans le système de pondération, soit à l'occasion des analyses de données.

En outre, la structure particulière du fichier individuel TOTAL exige que la ventilation des valeurs sur les différents sous-échantillons soit précisée. En effet, ce fichier regroupe les 6110 membres de l'échantillon initial et les 224 "nouveaux arrivants" de l'échantillon 1986. Les variables de pondération ne s'appliquent pas systématiquement à ces 6334 personnes.

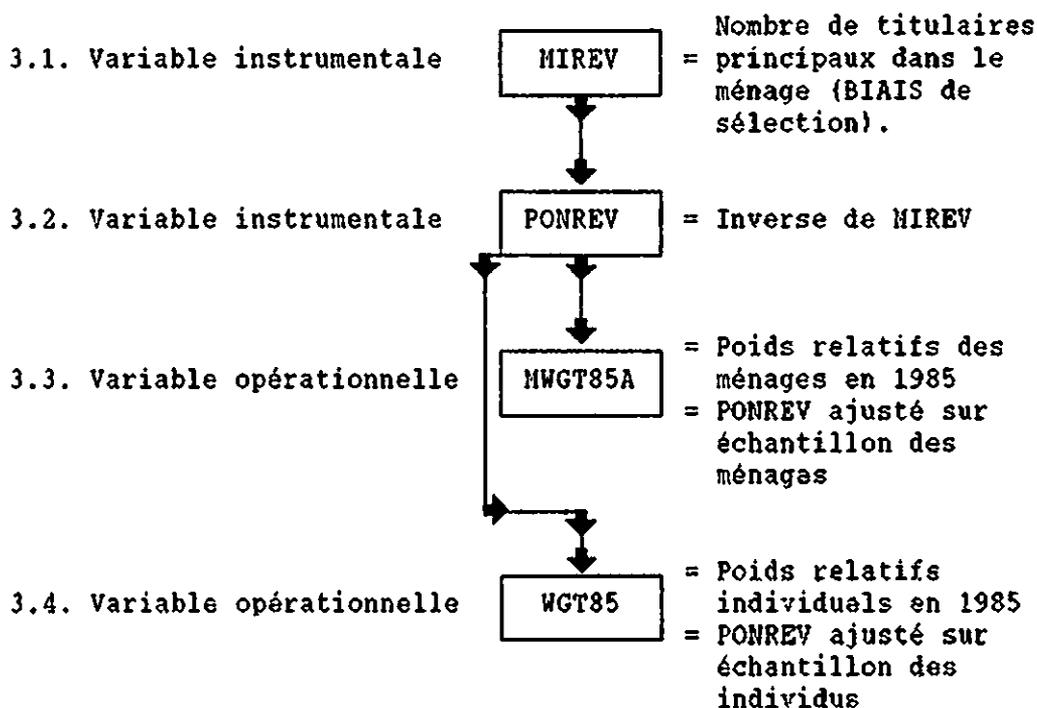
**Chapitre 3**  
**LES VARIABLES DE PONDERATION**

**Echantillon 1985**

---

Encadré 1.

VARIABLES DE PONDERATION 1985



### 3.1. 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.

Il était impossible de tirer l'échantillon d'un fichier national d'individus ou de ménages. Le seul fichier disponible permettait d'identifier les titulaires principaux de revenus, c'est-à-dire les bénéficiaires de la Sécurité sociale ou de la Protection sociale résidant au Luxembourg.

Ces titulaires principaux conduisent chacun à une adresse. 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 au niveau des individus: 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. Chaque ménage doit être affecté d'un facteur de correction qui correspond à l'inverse de sa probabilité d'apparaître dans l'échantillon.

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

**TABLEAU 2.**  
**Distributions des fréquences des ménages et des individus**  
**(en valeurs absolues),**  
**selon "MIREV" et "PONREV"**

"MIREV"	"PONREV"	FREQUENCES	
		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

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 appliqués à ces deux groupes (Tableau 3).

**TABLEAU 3**  
**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

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

Cette variable permet de pondérer correctement l'échantillon des MENAGES observés en 1985. Elle a été calculée directement au niveau de chaque ménage, en appliquant le coefficient de correction décrit ci-dessus et en ajustant les valeurs de pondération afin de ne pas modifier la taille de l'échantillon.

Le poids du ménage est attribué à chacun de ses membres afin de rassembler toutes les variables de pondération dans un seul fichier individuel.

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 peut être utilisée au titre de valeur de pondération du ménage.

**TABLEAU 4**  
**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
<b>TOTAL</b>	<b>6334</b>

Le tableau 4 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.

### **3.3. Variable opérationnelle de pondération individuelle en 1985 "WGTR85"**

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 individus 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 résident à la même adresse.

L'évolution de la structure des ménages rendra cette propriété caduque dans les vagues d'enquêtes ultérieures.

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).

Ce principe sera appliqué dans les vagues d'enquêtes suivantes. L'évolution de la structure du ménage pourra entraîner des variations dans les poids des membres de ce 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 5**  
**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	(78)
3. nouveaux memb. non-panel 86	-
4. memb. 85 décédés/émigrés 86	110
-----	
Total	6188
sysmis	146
-----	
TOTAL	6334

Ces pondérations ont été attribuées à tous les individus présents dans l'échantillon en 1986, soit les groupes #0, #1 et #4 de la variable DIA (-gnostic)(Tableau 5).

Les nouveaux individus-panels, entrés dans l'échantillon en 1986 (groupe #2), ont également reçu ces valeurs de pondération. Celles-ci seront utiles au moment de calculer leur poids, en 1986.

Ces valeurs sont donc ventilées sur 6188 individus, mais elles sont opérationnelles uniquement pour 6110 d'entre eux (6188 moins 78 individus-panels entrés dans l'échantillon en 1986).

Les 146 nouveaux membres non-panel (groupe #3) n'ont reçu aucune valeur sur cette variable (code=SYSMIS).

## 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
<b>TOTAL</b>	<b>6110</b>	<b>100.0</b>	

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:

- par application de la variable "PONREV",

1. 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 ( $1 \times (1 / .712) = 1.40$ ;  $0.5 \times 1.4 = .70$ ; ...).

$$\begin{aligned} \text{Soit: WGT85} &= \text{PONREV} / (4354 / 6110) \\ &= \text{PONREV} \times 1.40 \end{aligned}$$

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 individus 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.

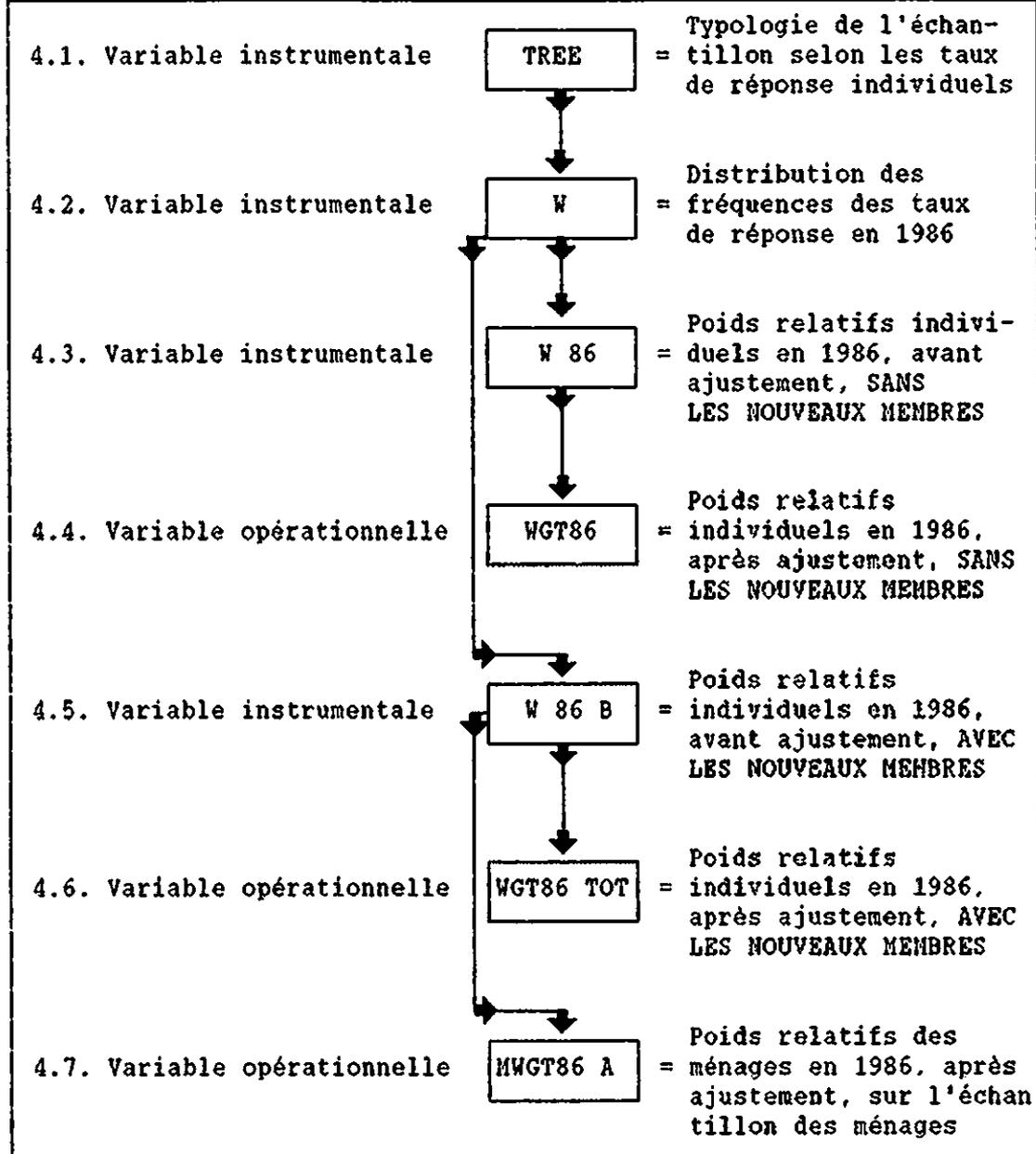
Chapitre 4  
LES VARIABLES DE PONDERATION

Echantillon 1986

---

## Encadré 2.

## VARIABLES DE PONDERATION 1986



#### 4.1. Variable instrumentale "TREE"

Les nouvelles pondérations utiles pour l'échantillon de 1986 (deuxième vague d'enquêtes) doivent prendre en compte différents types de phénomènes qui sont à l'origine de l'évolution de l'échantillon.

La deuxième vague d'enquêtes devait être effectuée auprès de tous les ménages et toutes les personnes interrogées en 1985. En réalité, certains membres de l'échantillon n'ont pas répondu. Les raisons de ces défections sont très diverses (tableau 6).

TABLEAU 6  
Que sont devenues, en 1986,  
les personnes interrogées en 1985?

DIAGNOSTIC INDIVIDUEL	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" <sup>1</sup>	17	.3
6. introuvables	32	.5
7. émigrés	60	1.0
8. décédés	50	.8
TOTAL	6110	100.0

Certaines personnes sont donc " perdues " d'une vague à l'autre: le tableau permet de constater que 15.4% des membres de l'échantillon initial n'ont pas pu être interrogés en 1986. L'absence de ces personnes modifie la structure de l'échantillon. Cette modification peut créer un biais systématique, si ces pertes sont concentrées dans une ou plusieurs catégories homogènes de la population. Ces catégories peuvent être identifiées par leurs traits communs (par exemple, "les femmes isolées de plus de 65 ans").

Avant d'aborder ce problème, il faut distinguer deux catégories d'événements qui se trouvent à l'origine de ces disparitions.

Les décès et l'émigration sont des phénomènes démographiques normaux, observables dans l'ensemble de la population du pays. Ces

---

1. refus individuels explicites de personnes ayant formé un nouveau ménage en 1986.

causes de disparition ne biaisent pas l'échantillon. Des corrections pourront être introduites, après plusieurs vagues d'enquête, s'il devait apparaître que les échantillons successifs ne reflètent pas correctement le taux de mortalité du pays.

Les autres événements sont liés à la vie de l'échantillon. Ils ne reflètent pas nécessairement des phénomènes observables dans la population du pays. Les groupes #2 à #6 (Tableau 6.) répertorient ces différents événements.

Ils seront assimilés à la catégorie des " refus ". puisqu'ils induisent le même problème statistique: 833 personnes, soit 13.6% des membres de l'échantillon initial n'ont pas pu être interrogées en 1986 pour des raisons propres à la vie de l'échantillon.

La méthode utilisée pour détecter l'existence de sous-groupes homogènes dans cet ensemble de " refus " sera plus documentée dans une version ultérieure. En résumé :

- les 5167 personnes ayant participé aux deux enquêtes reçoivent le code " 1 ";
- les 833 personnes assimilées à des "refus" reçoivent le code "0";
- les 110 personnes décédées ou émigrées ne sont pas prises en compte dans l'analyse;
- les personnes appartenant aux deux premières catégories sont caractérisées par une série de variables observées en 1985;
- le taux de réponse observé dans chaque modalité de chaque variable correspond à la proportion de " codes 1 " observés dans cette modalité;
- la recherche vise à identifier des variables dont les modalités présentent des taux de réponse très contrastés (écarts significatifs sur le plan statistique);
- il se pourrait qu'aucune segmentation de l'échantillon ne réponde à cette exigence. Dans ce cas, il faudrait conclure:
  - + soit, que les refus sont distribués aléatoirement dans l'échantillon,
  - + soit, que les variables de segmentation ne sont pas pertinentes.

Dans le cas présent, la recherche montre qu'il existe effectivement des sous-échantillons dont les taux de réponse varient de façon significative.

L'analyse est interrompue :

- soit, lorsque les variables de segmentation ne permettent plus de mettre en évidence des taux de réponse dont les écarts sont significatifs
- soit, lorsque les nouvelles modalités introduites dans l'analyse regroupent moins de 200 personnes.

La variable " TREE " est le résultat de cette démarche. Elle recense les différentes catégories de personnes identifiées au terme de l'analyse.

La présentation de cette variable appelle deux commentaires préalables.

1. L'analyse des taux de réponse est effectuée sur la base de l'échantillon de 1985. Elle doit donc prendre en compte l'inégalité des probabilités de sélection initiale des individus. Ce biais est corrigé en pondérant l'échantillon par la variable WGT85. Le tableau 7 permet d'observer l'évolution de l'échantillon avant et après pondération.

2. D'autre part, l'analyse des taux de réponse ne prend pas en compte les départs naturels car ceux-ci reflètent des phénomènes démographiques observables également dans la population de référence. La troisième colonne du tableau 7 présente l'état de l'échantillon pondéré, après la déduction des départs naturels: l'analyse des taux de réponse ne prend en compte que 5996 unités pondérées.

**TABLEAU 7**  
**Distributions des fréquences des individus membres**  
**de l'échantillon 1985, selon leur situation en 1986,**  
**avant et après pondération**

DIAGNOSTIC INDIVIDUEL	FREQUENCES		
	avant pondé- ration	après pondé- ration (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

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.

Le tableau 8 permet d'observer la " parenté " entre certains sous-groupes et les contrastes entre leurs taux de réponse.

TABLEAU 8 (1)

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

SEGMENTS	"TREE" Code	N = avant pondé- ration	N = après pondé- ration	TAUX DE REPOSE	
<b>1. RESIDENT A LUXEMBOURG-VILLE</b>					
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

TABLEAU 8 (2)  
Taux de survie, dans le panel, entre 1985 et 1986:  
Contrastes entre les taux de réponse de 18 segments de l'échantillon

SEGMENTS	"TREE" Code	N = avant pondé- ration	N = après pondé- ration	TAUX DE REPONSE	
<b>2. NE RESIDENT PAS A LUXEMBOURG-VILLE</b> 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

**TABLEAU 8 (3)**  
**Taux de survie, dans le panel, entre 1985 et 1986:**  
**Contrastes entre les taux de réponse de 18 segments de l'échantillon**

SEGMENTS	"TREE" Code	N = avant pondé- ration	N = après pondé- ration	TAUX DE REPONSE	
<b>3. NE RESIDENT PAS A LUXEMBOURG-VILLE</b> <b>Chef de ménage est</b> . 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.

L'examen du tableau 8 permet d'observer notamment trois faits saillants :

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

- ils habitent en-dehors de Luxembourg-ville, mais ils 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 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 réponse le plus faible (0.71) s'observe dans des ménages résidant à Luxembourg-ville, lorsque le chef de ménage n'a pas d'emploi (chômeur, cherche un emploi, invalide, handicapé, retraité....).

3. Les taux de réponse varient essentiellement en fonction des caractéristiques du ménage, du chef de ménage ou du lieu de résidence du ménage. Au cours de l'analyse, des caractéristiques individuelles ont été mises en concurrence constamment avec les caractéristiques du ménage. A aucun moment, ces caractéristiques individuelles ne se sont révélées aussi discriminantes que les caractéristiques du ménage.

Rem.: La prochaine version de ce document présentera:

- la liste des variables prises en compte au cours de l'analyse
- leur description statistique
  - avant toute pondération (1985)
  - après pondération " 1985 "
  - après pondération " 1986 "
- la segmentation progressive de l'échantillon, sous une forme arborescente.

Toutes les personnes membres de l'échantillon initial sont classées dans cette typologie, excepté les personnes décédées et les émigrés.

**TABLEAU 9**  
**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. 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 10).

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 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.

**TABLEAU 10**  
**Distribution des fréquences de l'échantillon 1986,**  
**selon les taux de réponse observés en 1986**  
**(avant et après correction de l'échantillon initial)**

" W "	"TREE"	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
<b>TOTAL</b>		<b>5167</b>	<b>5192</b>

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 7).

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 10 s'imposait encore à un autre titre:

L'analyse des taux de réponse a été effectuée en partant de l'échantillon initial pondéré. La comparaison des deux dernières colonnes du tableau 10 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 11**  
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	(78)
3. nouveaux memb. non-panel 86	-
4. memb. 85 décédés/émigrés 86	-
----- Total sysmis	5245 1089
----- TOTAL	----- 6334

Le tableau 11 permet de constater que des " taux de réponse " ont été attribués aux NOUVEAUX MEMBRES du PANEL entrés dans l'échantillon en 1986. L'origine et l'utilité de ces valeurs seront précisées plus loin.

### 4.3. 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 réponse 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.

TABLEAU 12  
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 sysmis	5167 1167
-----	
TOTAL	6334

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

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

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

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 12)).

#### 4.4. 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.

$$\text{Soit: } 1 / W \text{ devient } ( 1 / W ) / ( 5997 / 5167 )$$

et W86 devient

$$WGT86 = WGT85 \times ((1 / W ) / ( 5997 / 5167 ))$$

**TABLEAU 13**  
**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
<b>TOTAL</b>	<b>6334</b>

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 13)].

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.

#### **4.5. 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.

## 1. Les nouveaux membres du panel

Les membres du panel interrogés à deux reprises ne sont pas les seuls membres du panel. Chaque année de nouveaux individus viennent régénérer l'échantillon. Ce sont des nouveaux membres du panel, généralement des nouveaux-nés issus d'ascendants membres du panel, mais aussi des fils ou filles rentrés au foyer après une absence plus ou moins longue.

Leur probabilité d'entrer dans le panel dépend directement de deux facteurs:

- la probabilité que leurs parents appartiennent à l'échantillon initial
- la probabilité que leurs parents aient répondu à chacune des enquêtes précédentes.

Un poids relatif peut donc être ATTRIBUE à ces nouveaux membres. Il s'agit bien de leur attribuer un poids et non de le calculer. Ces membres n'étaient pas présents au cours de la première vague d'enquêtes; leur probabilité de sélection en 1985 ne peut donc pas être calculée.

La nature de leurs liens avec d'anciens membres du panel permet d'admettre que la probabilité de leur insertion dans l'échantillon, à un moment donné, est identique à celle de leur ascendant direct. Par convention, ils reçoivent donc un poids relatif égal à celui de leur père au moment où ils entrent dans le panel. En l'absence du père, ils reçoivent le poids relatif calculé pour la mère.

Ce transfert des pondérations peut être réalisé de deux manières différentes (ceci permet de comparer les résultats obtenus et de contrôler le bon fonctionnement des procédures).

Soit, le poids relatif de l'ascendant est attribué directement au nouveau membre panel.

Soit, le poids relatif du nouveau membre est calculé sur la base des valeurs antérieures de l'ascendant: c'est, précisément, pour cette raison que les poids des membres du panel en 1985 (voir " WGT85 " ) et les taux de réponse en 1986 (voir " W " ) ont été attribués anticipativement aux NOUVEAUX MEMBRES du PANEL dans le fichier individuel. Il suffit, dès lors, de calculer la valeur de " W86 " pour ces nouveaux membres, en utilisant les valeurs que leurs ascendants leur ont "légüées" par anticipation.

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).

## 2. Les nouveaux individus non-panel

Les nouveaux individus non-panel sont entrés dans l'échantillon pour des raisons très diverses. Il n'est pas possible d'estimer raisonnablement la valeur de leur probabilité d'insertion dans le panel. Ce calcul serait extrêmement complexe (Cf. Study Design, Procedures, Available Data. P.S.I.D., I.S.R., Ann Arbor, Michigan, 1972, pp.35 et ss.)

Dans ces conditions, il est convenu de leur attribuer un poids relatif de " 0 ". Toute autre solution serait basée sur des postulats qu'il serait impossible d'infirmer ou de démontrer. Les nouveaux individus membres du panel entrent simultanément dans la population et dans l'échantillon. Les nouveaux individus non-membres du panel appartenaient déjà à la population lorsqu'ils sont entrés dans l'échantillon. Ceci justifie la différence de traitement entre ces deux catégories de population.

Lorsqu'une analyse portera sur l'ensemble de l'échantillon individuel pondéré de 1986, les nouveaux individus non-panel n'auront pas d'influence sur les résultats. Sans doute est-ce là un coût important: si des pondérations avaient pu être calculées, l'échantillon aurait bénéficié d'un système de régénération qui se réduit actuellement au seul apport des nouveaux individus membres du panel.

Toutefois, il convient d'observer le fait que les informations liées à ces nouveaux participants seront prises en compte dans les analyses effectuées directement au niveau des ménages et dans les analyses qui utiliseront les caractéristiques du ménage d'appartenance en tant que caractéristiques individuelles.

**TABLEAU 14**  
**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 sysmis	5391 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 (-gnostic)(Tableau 14)].

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

#### **4.6. 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 15**  
**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 sysmis	5391 943
<b>TOTAL</b>	<b>6334</b>

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 (-gnostic)).

#### **4.7. 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'HETEROGENEITE des POIDS INDIVIDUELS au sein du ménage, hétérogénéité qui résulte inexorablement des modifications de la structure du ménage au cours du temps (voir annexe).

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 du fichier individuel 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 16  
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 sysmis	5391 943
----- TOTAL	6334

Le tableau 16 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

### 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:

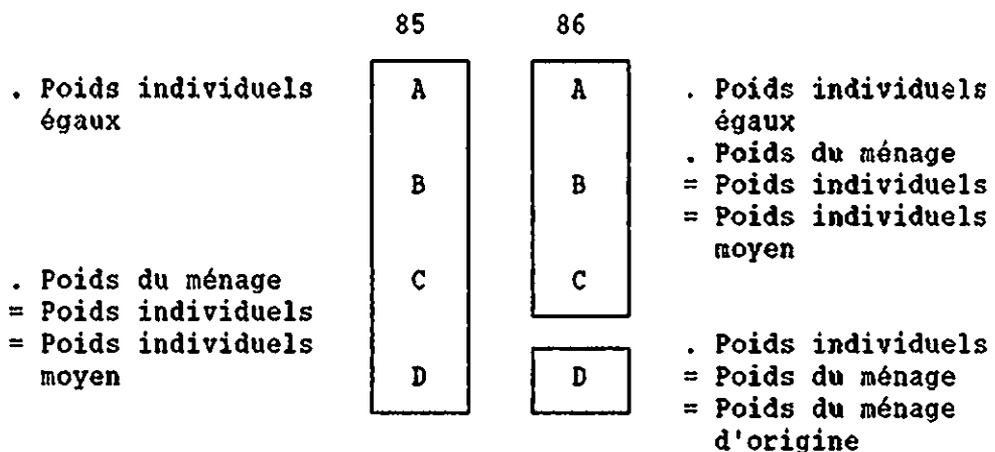
- l'insertion de nouveaux individus non-panel
- l'insertion de nouveaux individus-panel
- les modifications de la structure des ménages.

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

	85	86	
<ul style="list-style-type: none"> <li>. Poids individuels égaux</li>   <li>. Poids du ménage = Poids individuels = Poids individuels moyen</li> </ul>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;"> A  B  C  D </div>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;"> A  B  C  D  E </div>	<ul style="list-style-type: none"> <li>. Poids individuels égaux</li> <li>. Poids du ménage = Poids individuels = Poids individuels moyen</li>   <li>. Poids individuels différents (épouse=0)</li> <li>. Poids du ménage = <math>\frac{1}{2}</math> poids du fils <math>\neq</math> poids du ménage d'origine</li> </ul>

Dans l'hypothèse où 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

	85	86	
<ul style="list-style-type: none"> <li>. Poids individuels égaux</li>   <li>. Poids du ménage = Poids individuels = Poids individuels moyen</li> </ul>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;"> A  B  C  D </div>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;"> F  A  B  C  D  E </div>	<ul style="list-style-type: none"> <li>. Poids individuels différents (aïeul=0)</li> <li>. Poids du ménage = <math>(A + B + C) / 4</math></li>   <li>. Poids individuels différents (épouse=0)</li> <li>. Poids du ménage = <math>\frac{1}{2}</math> poids du fils (D/2) <math>\neq</math> poids du ménage d'origine</li> </ul>

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.

Tableau de synthèse:

Répertoire des variables de pondération  
et sous-échantillons pertinents

DIA Codes	WGTS5	N	W86	WGT86	W86B	WGTS6TOT	MWGT85A	MWGT86A
0. membres 85/refus 86	*						*	
1. membres 85/membres 86	*	*	*	*	*	*	*	*
2. nouveaux membres-panels 86	(*)	(*)			*	*		*
3. nouveaux membres non-panel 86					*	*		*
4. membres 85/décédés, émigrés 86	*						*	

• Les " \* " indiquent les sous-échantillons correspondant à chaque variable de pondération.

• Les " (\*) " indiquent des sous-échantillons ayant reçu des valeurs anticipativement. Les variables de pondération correspondantes ne s'appliquent donc pas directement à ces sous-échantillons.