UN ECHANTILLON PROBABILISTE DE A à Z :

L'EXEMPLE DE L'ENQUETE PEUPLEMENT ET DEPEUPLEMENT DE PARIS

INED (1986)

B.RIANDEY

Chef du Service des Enquêtes à l'INED

INTRODUCTION

Les ouvrages traitant des sondages probabilistes abordent souvent le travail d'échantillonnage d'une façon théorique et à partir d'exemples partiels. C'est au contraire une vision globale et concrète que nous recherchons à travers l'exemple traité, l'enquête Peuplement et dépeuplement de Paris de l'INED.

L'interprétation rigoureuse des résultats d'une enquête nécessite une connaissance précise des modalités de l'échantillonnage. C'est particulièrement vrai pour la démographie car les principes de l'analyse démographique conduisent souvent à s'intéresser à des sous-populations d'accès difficile dans les bases de sondage. La définition de la population enquêtée mérite donc une grande attention. Ainsi traiterons-nous d'abord de l'ajustement difficile des objectifs de l'enquête aux contraintes des bases de sondage.

Trois facteurs confèrent une marge d'incertitude aux enquêtes par sondage : les biais de sélection induits par les échecs de collecte, la taille limitée de l'échantillon et les erreurs de mesure (dans la présente enquête, les omissions d'événements ou leur datation inexacte). Nous allons évaluer les données de ces trois points de vue, examiner les méthodes de correction applicables et leur impact sur la validité des résultats. Ce faisant, nous aborderons les modèles (ou les mécaniques !) de redressement et la mesure des incertitudes d'échantillonnage.

Cette présentation assez complète de la technique du sondage probabiliste évite la théorie et le formalisme nécessaire aux spécialistes. Ce texte devrait ainsi être accessible à toute personne désireuse d'aller au fond des choses.

I - DE L'ECHANTILLON THEORIQUE A L'ECHANTILLON REEL

Les comptes rendus d'enquêtes évoquent rarement la démarche qui conduit à retenir un plan de sondage de préférence à des solutions alternatives qui tombent vite dans l'oubli. Celle qui est mise en oeuvre est alors revêtue d'une évidence trompeuse. Notre évaluation démographique du présent

échantillon comprend donc une discussion qualitative de la méthode d'échantillonnage. Dans ce but, nous avons exhumé les notes techniques préliminaires à l'enquête. Nous y avions même avancé des hypothèses sur le rendement de l'échantillon que nous pourrons confronter au bilan de la collecte. Nous regarderons ensuite la structure géographique de la population et réévaluerons notre optimisation spatiale de l'échantillon.

I.1 - Définition de la population sondée et méthode de sondage

Cette étude s'intéresse à l'impact des changements urbains sur l'histoire résidentielle d'un groupe de générations nées entre 1926 et 1935. Installées au lendemain de la guerre, ces générations vivent aujourd'hui en région parisienne l'étape d'émancipation de leurs enfants.

Deux enquêtes de l'INED sur l'agglomération de Paris :

L'enquête «Peuplement de Paris» de Guy Pourcher (1961) et l'enquête «Peuplement et dépeuplement de Paris» de Catherine Bonvalet (1986)

En 1961, G. POURCHER réalisait une innovation importante dans le domaine de la collecte des données démographiques : avec l'enquête Peuplement de Paris, il inaugurait la longue série des enquêtes longitudinales sur les migrations. Il apportait au domaine des études biographiques les outils de la statistique. Ainsi nous permet-il de suivre l'itinéraire des parisiens de son époque depuis leur naissance jusqu'à la date de l'enquête : leurs changements de résidence, la constitution progressive de leur famille, leur carrière professionnelle [1]. POURCHER a donné figure humaine à cet appel de population qu'a constitué l'expansion parisienne de l'après-guerre.

Vingt-cinq ans plus tard, C. BONVALET poursuit cette fresque à l'aide d'une enquête assez similaire. Ainsi nous révèle-t-elle le changement d'origine des immigrants qui ont nourri l'expansion périphérique des quinze années suivantes [2], tandis que, telle une géode, le coeur de la ville se vidait de sa population [3]. Elle observe la constitution de leur famille en pleine crise de logement de l'après-guerre, puis leur déserrement avec l'accroissement du parc immobilier et la politique d'accession à la propriété [4]. On pourra mesurer dans une prochaine publication l'évolution du nombre de pièces ou de mètres carrés par habitant au fil des mariages, des naissances et des entrées ou sorties de membres du ménage. La grande originalité du matériau apparaît ainsi que l'exigence d'une solide technique de questionnaire : l'entretien requiert un effort important de mémoire pour reconstituer la liste des logements et des co-résidents de l'enquêté,retrouver ses dates d'emménagement et celles des personnes avec qui il a partagé chaque logement.

Cette enquête rétrospective a été menée en Ile de France seulement ; son champ démographique exclut donc tous les Parisiens des générations 1926-1935 qui ont quitté la région au cours de leur cycle familial. C. BONVALET et E. LELIEVRE estiment à partir de l'enquête Triple biographie de l'INED que 25% de cette génération a connu au moins une résidence parisienne. Or la population soumise à l'enquête ne constitue que 19 % de celle-ci (en fait de sa fraction survivante résidant en France métropolitaine). De ce point de vue, l'enquête ne décrit donc pas exactement l'itinéraire-logement de la génération des jeunes Parisiens d'après-guerre, mais celui de près des trois-quart de celle-ci, de sa fraction stable, aux racines plus profondément franciliennes. L'échantillon subit d'ailleurs un léger biais de sélection : les hommes de 50 ans ont subi un risque d'émigration vers la province ou l'étranger moindre que ceux de 59 ans et surtout que les femmes mariées de 59 ans dont les maris sont généralement retraités. La population sondée présente donc des contours exacts légèrement artificiels.

Il ne s'agit pas d'un échantillon classique de ménages, mais d'individus. Le recours à une méthode probabiliste de préférence à la méthode empirique des quotas supposait l'accès à une base de sondage de qualité. En 1961, G. POURCHER a eu recours aux listes électorales permettant aisément de tirer des générations d'individus, mais limitées aux Français. Nous avons abandonné cette solution pour une raison plus technique : les Parisiens inscrits dans la commune de leur résidence secondaire seraient absents de l'échantillon et de nombreuses adresses, périmées de longue date, ne permettraient pas d'accéder à l'enquêté. L'ancienneté du recensement de 4 ans ne plaidait pas pour cette seconde source ; la dernière voie probabiliste envisageable supposait une enquête-filtre (reposant par exemple sur un échantillon de logements) dont on aurait retenu la génération concernée par l'étude. Cette solution s'annonçait coûteuse, à moins de disposer à cet usage d'une grande enquête déjà collectée. En fait, malgré son ancienneté, le recensement s'est avéré une base adaptée à l'échantillonnage de ces générations grâce aux faibles taux migratoires à ces âges :

- les individus partis en province depuis le recensement ont quitté notre champ et ne risquent pas d'être indûment enquêtés. Cette situation prévaut de même pour les personnes décédées. On doit néanmoins les repérer pour les déduire de l'effectif des échecs ; nous évaluons ces deux catégories à moins de 10 % de l'échantillon initial[1] malgré le développement récent des pré-retraites.

- les individus qui ont migré depuis le recensement à l'intérieur de «l'espace parisien» doivent être retrouvés et enquêtés ; ils constituent environ 20% de l'effectif de la tranche d'âge pendant la période intercensitaire (tableau 1), soit un peu moins de 15 % de l'échantillon de l'enquête ; Cette recherche supposait un effort de collecte important mais elle nous paraissait facilitée par la rareté des migrations multiples à ces âges et par leur bonne insertion dans les relations de voisinage. Ces conditions favorables étaient supposées compenser les difficultés de suivi inhérentes à notre grande agglomération.

- Le cas des immigrants est plus délicat, mais leur importance numérique est marginale : pour cette tranche d'âge, l'immigration parisienne intercensitaire voisine 2 % : le biais dû à l'omission de cette sous-population est donc absolument minime ; il ne justifie pas d'en constituer un sous-échantillon problématique à partir d'une base complémentaire .

Ainsi, en 1985, nous avons avancé le pronostic suivant pour cent individus tirés :

- 10 personnes auraient quitté l'aire de l'enquête ;

- 90 personnes seraient dans le champ de l'enquête et se répartiraient en

  . 75 non migrants
  . 15 migrants internes à l'aire.

Si les enquêteurs retrouvent les 2/3 des adresses de migrants internes, 85 % de l'échantillon pourraient être effectivement **contactés** (hypothèse

1 - La transformation des flux intercensitaires (tableau 1) en flux quadriannuels est délicate, mais nous ne recherchons que des ordres de grandeur. Pour plus de précision, voir D. COURGEAU [5] que nous remercions de ses conseils et de son aide ("Migrants et migrations", Population n° 2, 1973). Les estimations portant sur l'agglomération parisienne sont très voisines de celles relatives à la région.

TABLEAU 1 - POPULATION TOTALE DE L'ILE DE FRANCE EN 1982 PAR CATEGORIE DE MIGRANT

Age atteint en 1982 (générations)	Population totale en 1982	Hors ménages ordinaires	Même logement en 1975 *CM2 = 1*	Immigrants			Estimation population totale en 1975	Estimation émigration décès depuis 1975	Emigrants *tab MI2070*
				Autre région *CM2 = 7*	Hors métropole en 1975	en Ile de France en 1975			
45-49 ans (1933-1937)	610.560	2.360	398.040 60,00%	18.020 -	16.640 -	175.600 -	663.940 *(3/5B+2/5A)*	88.030 13,30%	43.080
50-54 ans (1928-1932)	617.640	2.120	447.820 67,82%	15.760 -	10.940 -	141.000 -	660.380 *(3/5C+2/5B)*	69.440 10,50%	43.660
55-59 ans (1923-1927)	545.360	1.620	420.780 67,50%	10.360 -	8.340 -	104.260 -	623.314 *(3/5D+2/5C)*	96.600 15,50%	56.620
50-59 ans en 1982 Générations 23-32	1.163.000	3.740	868.600 67,66%	45.400		245.260 19,10%	1.283.694 *(2/5B+C+3/5D)*	166.040 12,90%	100.280

Les pourcentages sont rapportés aux effectifs de 1975
Source : tableau MI2090 du recensement de 1982

Population totale d'Ile de France en 1975

A 1936-1940	654.225
B 1931-1935	670.400
C 1930-1926	653.700
D 1925-1921	602.890

en 7 ans les 2/3 des personnes ont conservé leur logement (67,66 %)
13% des personnes sont décédées ou ont quitté la région ;
19,1% d'entre elles ont changé de résidence dans la région ;
(la différence tient aux hors ménages 3 %)

Remarque sur l'émigration de la région parisienne

L'INSEE publie au tableau MI2070 (reproduit en dernière colonne du tableau 1) des estimations d'émigration par âge très inférieures à nos chiffres (en particulier pour les 45-49 ans, peu influencées par les décès).
Ces chiffres correspondent probablement aux résidents en Ile de France en 1975 (survivants résidant en métropole en 1982).
L'écart correspondrait à la mortalité et aux migrations de retour vers l'étranger et les DOM-TOM.

H1) ; mais s'ils ne retrouvent que la moitié de leurs adresses, ce serait seulement le cas de 82,5 % de l'échantillon (hypothèse B1).

Supposons enfin un taux d'échec (refus, absents de longue durée) de 15% (hypothèse H2) ou de 20 % (hypothèse B2). Ces hypothèses déterminent quatre éventualités ; le tableau 2 fournit pour chacune d'entre elles une estimation de la taille de l'échantillon effectif en fonction du déroulement d'une collecte menée auprès de 3000 individus initiaux. Nous allons maintenant examiner si la collecte a correspondu à l'une de ces hypothèses.

TABLEAU 2
RENDEMENT DE LA COLLECTE ET TAILLE FINALE DE L'ECHANTILLON SELON PLUSIEURS HYPOTHESES DE COLLECTE

		Recherche des adresses	
		H1 (67%)	B1 (50%)
SUCCES DE L'INTERVIEW	H2 (85%)	72% 2160	70% 2100
	B2 (80%)	68% 2040	66% 1980

I.2 - Le bilan global de la collecte

Le tableau 3 fournit le bilan du terrain. Près de 3000 adresses ont abouti à 1994 questionnaires remplis, dont 1987 sont exploitables. En définitive, les échecs s'élèvent à 25% des adresses non reconnues hors-champ ; ils se répartissent en 6,1% adresses non retrouvées, 14,1% de refus, 1,6% d'interviews impossibles à réaliser et 3,2% des personnes impossibles à joindre, soit 19% d'échecs de l'entretien lui-même. Les échecs liés aux adresses s'avèrent au

TABLEAU 3
LE BILAN DE LA COLLECTE

Adresses tirées	2993	
Adresses confiées aux enquêteurs	2939	
Questionnaires hors champ	280	(9,5%)
dont :		
- hors-champ géographique	202	
- décès	61	
- autre hors-champ (erreur d'âge)	17	
Questionnaires dans le champ	2659	(90,5%)
Questionnaires remplis	1994	(75,0%)
Echecs	665	(25,0%)
dont :		
- Adresses inconnues	161	(6,1%)
- Refus	377	(14,1%)
- Impossibles à réaliser	42	(1,6%)
- Impossibles à joindre	85	(3,2%)

contraire mineurs. C'est là un résultat heureux car ils conduisent par définition à un biais (en faveur des sédentaires).

L'ancienneté de 4 ans de notre base nous avait fait prévoir des **dossiers hors-champ** qu'on doit décompter des échecs de collecte. Nous avions envisagé 10 % de personnes hors-champ géographique ou décédées (294). Nous nous situons à 31 unités en deçà de notre seuil, malgré le mouvement d'anticipation des retraites depuis février 1982.

Avec 1987 questionnaires exploitables, notre bilan semble légèrement plus favorable que l'**hypothèse basse B1-B2**. En fait, ce point de vue est

pessimiste car le terrain n'a porté que sur 2939 adresses [1]. Ainsi un total de 3000 adresses valables aurait conduit à 2048 questionnaires **remplis**, soit à une situation très voisine de l'hypothèse plus favorable H1B2 .

La recherche d'adresse : hypothèse H1 ou B1 ?

Au total, les enquêteurs ont dénombré 528 changements de domicile, intervenus depuis le recensement (18 % des adresses confiées). Ceux-ci se répartissent en 202 départs en province (hors-champ géographique), 165 adresses identifiées en Ile de France et 161 adresses indéterminées. Ces dernières ne peuvent être ventilées entre les "hors-champ" (départ en province) et les échecs (migrations internes) qu'en formulant des hypothèses. Selon le cas, la mobilité interne à l'Ile de France s'établirait entre 6 % (165) et 11 % (326) des adresses confiées aux enquêteurs (décès déduits); elle resterait de toutes façons inférieure aux 15 % prudemment envisagés. **Dans l'hypothèse la plus défavorable**, ces 161 dossiers correspondraient toutes à des migrations dans le champ; le suivi ne se traduirait alors que par 51 % de succès, conformément à l'hypothèse B1.

L'hypothèse H1 requièrerait au contraire que ces 161 adresses proviennent de 84 migrations dans le champ et 75 hors de la zone d'enquête ; ainsi, 27 % des migrations hors-zone et 33 % des migrations internes échapperaient aux enquêteurs. Cette hypothèse est **irréaliste** car on repère beaucoup plus aisément un départ vers la province (ou l'étranger) qu'on ne localise **avec précision** une nouvelle adresse parisienne.

Retenons une hypothèse plus raisonnable : les enquêteurs auraient échoué pour 40 % des migrations dans le champ (111 dossiers), et donc pour 20 % des migrations hors champ (environ 50). Ce résultat demeure inférieur à l'hypothèse H1, pourtant atteinte quelques mois plus tôt, grâce à un rattrapage, pour l'échantillon parisien de l'enquête "Divorce" de l'INED.

---

[1] Il a fallu éliminer quelques dizaines de personnes (1,9%) déjà enquêtées par l'INSEE, ou dont le bulletin était mal classé dans la base de sondage. On a encore dû enlever 18 enquêtés hors-champ, tirés à la suite d'une erreur de saisie . L'échantillon tiré par l'INSEE s'est donc avéré un peu faible, mais les statisticiens avertis savent bien qu'avec 2,5% de déchet, une base de sondage doit être considérée comme excellente.

Les échecs de l'entretien : L'hypothèse B2

Les échecs de l'interview valident notre hypothèse basse B2, malgré un retour sur le terrain en avril. Destinée à rattraper des refus et des absents de longue durée, cette résorption par un nouvel enquêteur avait porté sur 575 adresses ; elle avait permis de rattraper 164 questionnaires (soit 6 % des personnes dans le champ) et de repérer 4 hors-champ et 4 décès.

Cette résorption s'est donc avérée assez difficile et a bien mis en valeur les difficultés d'accueil rencontrées par cette enquête : la période pré-électorale, que nous n'avons pu éviter, est défavorable à la collecte ; d'autre part le thème du logement ne s'est pas montré aussi porteur que nous l'imaginions : les personnes âgées de 50 à 60 ans l'associent peut-être plus au patrimoine immobilier qu'aux conditions de vie, car ils vivent leur phase de déserrement familial lié au départ des enfants. Cette enquête récapitulait le cheminement de leur génération ; cette contribution à l'histoire des Parisiens n'a pas suscité leur enthousiasme ; ils auraient mieux compris qu'on interrogeât leurs enfants qui connaissent aujourd'hui leurs difficultés d'installation d'antan . Les enquêteurs s'attendaient à un taux de refus élevé dans cette tranche d'âge, peut-être plus méfiante que les générations plus jeunes. Les faits semblent leur donner raison.

Nous attendions bien sûr les difficultés de collecte qu'occasionnent les grandes métropoles urbaines. Citons à nouveau l'enquête divorce de l'INED ; les échecs de l'interview y sont en Ile de France supérieurs de moitié à ceux de la province (13 % et 21 %) tandis que les remises à jour des adresses y subissent deux fois plus d'échecs (31 % et 16 %) [6].

Les refus se sont avérés la principale cause d'échecs, contrairement à notre attente d'un accueil amélioré pour cette enquête régionale. L'absence de cet effet dénie un sentiment régionaliste dans une agglomération incontournable qui, selon les provinciaux, s'identifierait trop au pays entier.

En définitive, si le bilan de la collecte ne valide pas nos hypothèses les plus optimistes, il demeure encore satisfaisant, car conforme aux fourchettes annoncées. Au chapitre suivant, il nous restera à contrôler que ces refus ne sont pas -trop- concentrés dans quelques segments de la population -par

exemple, les propriétaires. Alors sera justifié le choix de la méthode probabiliste de préférence aux quotas.

I.3 - Structure géographique de la population

Nous n'avons pas encore précisé exactement le champ géographique retenu pour l'enquête ; en général, la définition de l'aire d'enquête repose sur des critères administratifs. On facilite ainsi la diffusion des résultats et les comparaisons aux recensements et aux autres enquêtes. Mais un autre type de découpage peut s'imposer pour constituer un espace géographique «complet», c'est-à-dire un espace qui évite de tronquer le phénomène étudié : dans notre cas, il s'agit de la migration parisienne.

Quatre niveaux géographiques se présentaient a priori :

- Paris intra-muros ;
- l'agglomération parisienne ;
- la ZPIU de Paris (zone de peuplement industriel ou urbain) ;
- la région d'Ile de France.

Les deux premiers niveaux sont trop restrictifs pour prendre en compte le phénomène péri-urbain avec ses navettes quotidiennes. La ZPIU se distingue peu de la région, mais elle englobe un grand nombre de communes rurales et de petites agglomérations trop distantes des lieux d'implantation du réseau d'enquêteurs.

Un choix plus satisfaisant s'appuie sur la typologie de l'Ile de France élaborée à l'IAURIF par A. FOUCHER [7] (tableau 4 et carte 1(voir fin)) : cette typologie distingue les zones «sous influence urbaine» de celles qui ne le sont pas. Elle est en harmonie avec notre notion large d'espace urbain ; par exemple, la distinction entre «les axes et vallées» et «les villes petites et moyennes moins bien desservies» rend compte du processus historique d'urbanisation le long des voies de communication, phase qui précède l'urbanisation dense concentrique. Autre avantage, l'IAURIF mène sur cette base des exploitations du recensement qui permettent des comparaisons éclairantes avec nos résultats.

Nous avons donc écarté du champ de l'étude les "villes petites et moyennes moins bien desservies" et bien sûr les communes rurales éloignées, mais nous avons dû aussi renoncer aux communes rurales proches

TABLEAU 4

APPLICATION A L'ENQUETE DE LA TYPOLOGIE DE L'IAURIF CONCERNANT LES COMMUNES DE L'ILE DE FRANCE

	Nombre de communes * *	Part de territoire urbanisée	Superficie en ha	Population en 1982	Population âgée de 50-59 ans en 1986 * * *	Nombre d'aires dans l'échantillon * * * *	Taille de l'échantillon initial * * * *
AGGLOMERATION URBAINE DENSE							
Paris	20	82 %	10.540	2.176.000	254.000	8-12	512-507
Banlieue intérieure	70	83 %	35.470	2.930.000	376.000	17-17	386-392
Partie urbanisée de la banlieue extérieure	120	67 %	54.030	2.530.000	323.000	14-14	334-336
VILLE NOUVELLE	58	18 %	56.930	414.000	37.000	5	97
ZONE EXTERIEURE							
* Franges de l'agglomération	114	24 %	87.260	726.000	91.000	4-4	115-121
* Villes petites et moyennes bien desservies (axes et vallées)	127	14 %	130.540	677.000	77.000	6	200
SOUS-TOTAL	409	-	374.770	9.453.000	1.158.000	101	3000
* Villes petites et moyennes moins bien desservies	84	10 %	70.150	160.000			
* Communes rurales proches		5 %	259.280			-	-
	705			369.000	-		
Communes rurales éloignées		5 %	497.030		-	-	-
ENSEMBLE	1.198	-	1.201.230	9.972.000	-	-	-

Source : A. FOUCHER, Annuaire de l'IAURIF

* Ces quatre secteurs constituent l'ensemble péri-urbain proprement dit

** En comptant les arrondissements de Paris.

*** En fait, population recensée en 1982, âgée de 46 à 55 ans au 01/01/1982.

**** En distinguant les deux sous-strates (communes «populaires» communes «bourgeoises»)

dont le nombre des migrations et des navettes avec l'agglomération ne justifie pas une dispersion coûteuse de la collecte. En définitive, nous avons étendu notre aire aux communes rurales des villes nouvelles et aux "axes et vallées". La zone retenue pour l'enquête recouvre donc les 6 premiers types de communes totalisant 94,8 % de la population de l'Ile de France mais seulement 32 % du territoire. Les densités de population y sont respectivement de 25 200 habitants/ha pour l'aire de l'enquête et 620/ha sur la zone écartée.

L'Histoire a légué à la capitale une morphologie du bâti fortement structurée, mais la carte du peuplement n'en est pas moins contrastée : la ségrégation sociale mesurée par L.LEBART et N. TABARD[1] [8] en serait le trait dominant, avant même les déséquilibres de la structure par âge.

Appliquant au recensement de 1975 la méthode expérimentée au CREDOC, nous avons résumé cette information en classant chaque commune de l'agglomération comme "bourgeoise" ou "ouvrière".

Cette dichotomie sociale se superpose donc au classement morphologique de l'IAURIF pour caractériser 18 types de communes ou d'arrondissements[2] en Ile de France (voir tableau 5). Dans toute enquête régionale en Ile de France, le plan de sondage se doit de restituer fidèlement cette image contrastée du "grand Paris".

I.4 - Optimisation géographique de l'échantillon

Un échantillon uniforme de 3000 personnes correspondrait à un taux de sondage de 1/386 ème. Nous avons préféré un échantillon stratifié géographiquement à taux de sondage multiples.

---

1 - Elle repose sur une analyse factorielle des correspondances du tableau de la population de l'agglomération répartie par catégorie socio-professionnelle et par commune ou arrondissement, et produit un indicateur ordonnant les communes selon la prédominance relative des classes riches ou populaires au recensement de 1968.

2 - Rappelons que 6 strates sont d'emblée exclues du champ de l'enquête en zone rurale et péri-urbaine.

TABLEAU 5
LES UNITES PRIMAIRES DE L'ECHANTILLON

Strate 1 : le Paris ouvrier (8 arrondissements retenus d'office)	**Strate 2 : le Paris bourgeois** (12 arrondissements retenus d'office)	**Strate 3 : la banlieue intérieure ouvrière** 7 communes retenues d'office 10 communes tirées au sort		**Strate 4 : la banlieue intérieure bourgeoise** 4 communes retenues d'office 13 communes tirées au sort	
PARIS 2EME	PARIS 1ER	DRANCY	93	RUEIL-MALMAISON	92
PARIS 10EME	PARIS 4EME	CHAMPIGNY-SUR-MARNE	94	COLOMBES	92
PARIS 12EME	PARIS 3EME	NANTERRE	92	ST-MAUR-DES-FOSSES	94
PARIS 11EME	PARIS 8EME	VITRY-SUR-SEINE	94	BOULOGNE-BILLANCOURT	92
PARIS 19EME	PARIS 6EME	SAINT-DENIS	93		
PARIS 13EME	PARIS 5EME	MONTREUIL	93	KREMLIN-BICETRE	94
PARIS 20EME	PARIS 9EME	ARGENTEUIL	95	LILAS	93
PARIS 18EME	PARIS 7EME	BONNEUIL-SUR-MARNE	94	VANVES	92
	PARIS 14EME	ROMAINVILLE	93	FONTENAY-AUX-ROSES	92
	PARIS 17EME	NOISY-LE-SEC	93	PERREUX-SUR-MARNE	94
	PARIS 16EME	CHOISY-LE-ROI	94	SURESNES	92
	PARIS 15EME	BOBIGNY	93	VINCENNES	94
		PANTIN	93	FONTENAY-SOUS-BOIS	94
		EPINAY-SUR-SEINE	93	MAISONS-ALFORT	94
		VILLEJUIF	94	LEVALLOIS-PERRET	92
		IVRY-SUR-SEINE	94	COURBEVOIE	92
		AUBERVILLIERS	93	NEUILLY-SUR-SEINE	92

Strate 5 : la banlieue extérieure ouvrière 1 commune retenue d'office 13 communes tirées au sort		**Strate 6 : la banlieue extérieure bourgeoise** 1 commune retenue d'office 13 communes tirées au sort		**Strate 7 : France ouvrière de l'agglomération** 4 communes tirées au sort		**Strate 8 : France bourgeoise de l'agglomération** 4 communes tirées au sort		**Strate 9 : villes nouvelles** 5 villes nouvelles retenues d'office	**Strate 10 : Axes et vallées** 1 agglomération retenue d'office 5 agglomérations tirées au sort	
AULNAY S/BOIS	93	VERSAILLES	78	OLLAINVILLE	91	MAREIL-MARLY	78	EVRY (2 communes au sort)	MANTES LA JOLIE	78
EPINAY S/SENART	91	VAUCRESSON	92	BOISSY-ST-LEGER	94	CHAMBOURCY	78	MELUN-SENART (3 communes au sort)	EPONE	78
LES ULIS	91	ARNOUVILLE-LES-GONNESSE	95	BRETIGNY-ST-LEGER	91	VILLEPREUX	78	ST QUENTIN YVELINES (3 communes au sort) CERGY (4 communes au sort) MARNE LA VALLEE (5 communes au sort)	SAINT FARGEAU	77
LONGJUMEAU	91	VIROFLAY	78	PLAISIR	78	HERBLAY	95		BEAUMONT SUR OISE	95
PIERREFITTE SUR SEINE	93	VESINET	78						MEAUX	77
MONTFERMEIL	93	TAVERNY	95						MELUN	77
RIS ORANGIS	91	LE CHESNAY	78							
NEUILLY S/MARNE	93	ERMONT	95							
HOUILLES	78	YERRES	91							
CORBEIL-ESSONNES	91	CHATENAY-MALABRY	92							
STAINS	93	PALAISEAU	91							
CHELLES	77	LIVRY-GARGAN	93							
SARTROUVILLE	78	GAGNY	93							
SARCELLES	95	ANTONY	92							

Selon la pratique habituelle, le sondage est réalisé à deux degrés :

- dans un premier temps, on tire dans chaque strate des aires géographiques (communes, petites agglomérations, villes nouvelles) constituant ainsi autant d'échantillons indépendants que de strates ;

- puis on tire des individus dans ces aires.

Cette procédure nous assure de l'exacte représentation de chaque strate et concentre la collecte sur un nombre limité d'aires. Nous nous sommes ramenés à dix strates car la population encore réduite des villes nouvelles et des axes et vallées ne justifiait pas qu'on les scinde en sous-secteurs "bourgeois" et "ouvriers".

La procédure de tirage des communes est simple :

1) toute commune comportant 9000 personnes nées entre 1926 et 1935 est retenue d'office ainsi que tous les arrondissements parisiens ;

2) toute commune en comportant moins est tirée selon une probabilité proportionnelle à la taille de cette population selon la procédure de tirage systématique[1]. Les communes retenues sont indiquées au tableau 5.

Taux de sondage uniforme ou différencié ?

La stratification est une technique efficace, même pour un plan à taux de sondage uniforme : elle garantit un effectif exact à chaque strate, réduisant en cela l'effet de grappe dû au tirage à plusieurs degrés. Mais, en choi-

1 - Dans la pratique, on classe les communes de la strate par population croissante (générations 1926-1935) ; on cumule cette population et on tire selon une progression arithmétique des individus fictifs dont la commune est retenue (tirage systématique). Cette procédure réalise dans chaque strate une quasi-stratification des communes selon leur taille, ce qui améliore encore l'efficacité du sondage.

sissant des taux de sondage différenciés entre strates, on assure une représentation optimisée pour l'un des trois objectifs suivants : analyser une sous-population de petite dimension, établir des différentielles entre sous-populations d'effectifs très inégaux, réaliser des estimations sur une variable dont la dispersion est inégale entre sous-groupes (sondage de Neyman en situation d'hétéroscédasticité).

Nous nous étions fixé un seuil minimal de 600 questionnaires collectés dans la commune de Paris pour en permettre une analyse isolée ; nous souhaitions aussi pouvoir isoler les différentes strates du sondage dans les distributions simples.

Nous avons dans ce but envisagé trois répartitions par strate des taux de sondage (tableau 6). Le taux de sondage uniforme de 1/386 conduisait à 657 fiches-adresses à Paris. Après déduction des hors-champ, des adresses per-

TABLEAU 6
TROIS PROPOSITIONS DE PLAN DE SONDAGE

STRATES	Taux unique	2 taux		3 taux		Echantillon
		effectif	variation	effectif	variation	effectif *
1-2 Paris	657	857	+ 200	1019	+ 362	618
3 Banlieue intérieure ouvrière	483	386	- 97	386	- 97	245
4 Banlieue intérieure bourgeoise	490	392	- 98	392	- 98	275
5 Banlieue extérieure ouvrière	417	334	- 83	334	- 83	221
6 Banlieue extérieure bourgeoise	420	336	- 84	336	- 84	251
SOUS-TOTAL 3-6	1810	1448	- 362	1448	- 362	992
7 Frange ouvrière	115	150	+ 35	115	-	71
8 Frange bourgeoise	121	158	+ 37	121	-	96
9 Villes nouvelles	97	126	+ 29	97	-	66
10 Axes et vallées	200	261	+ 61	200	-	144
SOUS-TOTAL 7-10	533	695	+ 162	533	-	377
TOTAL	3000	3000	-	3000	-	1987

* par strate de résidence à la date de l'enquête

dues, des refus et du solde migratoire Paris-banlieue, on n'aurait pu espérer 600 questionnaires. Avec cette taille d'échantillon constante, la surreprésentation de Paris s'avérait nécessaire. Les effectifs abondants des banlieues intérieure et extérieure suggéraient d'y prélever l'appoint recherché, sans affaiblir les petites strates (7-10). Cet appoint pouvait être ventilé sur les 6 autres strates ou réservé à la capitale. La solution médiane améliorait les estimations inter-strates, contrairement à la dernière qui avec une distribution plus complexe des poids donnait une priorité à l'objectif

Ces trois plans de sondage se traduisent par le jeu suivant de taux de sondage :

- un taux unique : t = 1/386
- deux taux dans la seconde hypothèse :
  . t1 = 0,8 t = 1/483 pour la banlieue intérieure ou extérieure,
  . t2 = 1,3 t = 1/297 pour Paris, les franges, les villes nouvelles et les axes et vallées.
- trois taux dans l'hypothèse retenue
  . t'1 = 0,8 t = 1/483 pour la banlieue intérieure ou extérieure,
  . t'2 = t = 1/386 pour les franges, les villes nouvelles et les axes et vallées.
  . t'3 = 1,55 t = 1/249

Notre choix devait anticiper sur les résultats de la collecte estimés à 72 % dans l'hypothèse optimiste H1H2 et à 66 % dans l'hypothèse prudente, éventuellement aggravée par un solde migratoire négatif Paris-périphérie (tableau 7). La troisième solution répondait à notre stratégie prudente. Elle s'est révélée judicieuse puisque l'objectif des 600 enquêtés à Paris n'a été dépassé que de 18 unités car le rendement de l'échantillon s'est avéré particulièrement faible à Paris et s'améliore du centre vers la périphérie (tableau 8).

Par ailleurs, il se montre toujours plus élevé (de 6 à 17 %) dans une strate bourgeoise que dans la strate ouvrière associée, sous l'effet cumulé des différentielles de hors-champ et d'échec. La ventilation par strate des hors-champ et des échecs éclaire ces inégalités de rendement. Plusieurs phénomènes se conjuguent :

- Les échecs de l'interview décroissent du centre vers la périphérie, mais sont toujours plus marqués dans les communes populaires.

TABLEAU 7

PREVISIONS DE LA TAILLE DE L'ECHANTILLON PARISIEN *

Hypothèse de rendement	Taux unique	2 taux	3 taux
H1H2 72 %	473	617	734
B1B2 66 %	434	566	673

* Les enquêtés sont ventilés selon leur strate de résidence en 1982

TABLEAU 8

TAUX DE RENDEMENT DE L'ECHANTILLON PAR STRATE DE TIRAGE *

Paris		Banlieue intérieure		Banlieue extérieure		Frange		Villes nouvelles	Axes et vallées	Ensemble
Ouvrier 1	Bourgeois 2	Ouvrière 3	Bourgeoise 4	Ouvrière 5	Bourgeoise 6	Ouvrière 7	Bourgeoise 8	9	10	
58%	64%	64%	70%	66%	75%	62%	79%	68%	72%	66%
61%		69%				71 %				66%

- Les déménagements se raréfient très régulièrement du centre vers la périphérie sans distinction entre strates bourgeoises et ouvrières.

Cette section illustre la masse d'informations nécessaire à la construction d'un plan de sondage efficace. Selon les sujets traités, une fraction plus ou moins considérable de celle-ci fait défaut : le statisticien doit lui substituer des fourchettes d'hypothèses, rendant le résultat aléatoire. Il n'en est que plus nécessaire que les objectifs soient clairement définis et hiérarchisés. C'est ce qui, dans cette enquête, nous a permis de développer une démarche bien finalisée.

II - EVALUATION DE LA REPRESENTATIVITE DE L'ECHANTILLON

Nous allons maintenant comparer la structure de l'échantillon recueilli à celle de la population entière (supposée connue grâce à des statistiques exhaustives). Les écarts statistiquement significatifs sont révélateurs du biais introduit par les échecs différentiels. De tels biais doivent être corrigés. C'est l'objet du redressement présenté en partie III.

Si au contraire la collecte avait été parfaitement réalisée (100 % de succès, pas d'erreur de mesure), les écarts entre estimations et statistiques ne devraient qu'exceptionnellement être significatifs : ces écarts résultant seulement du hasard mériteraient cependant d'être éliminés par une pondération efficace ; c'est la technique de la stratification a posteriori, très semblable au redressement dans sa mise en oeuvre[1].

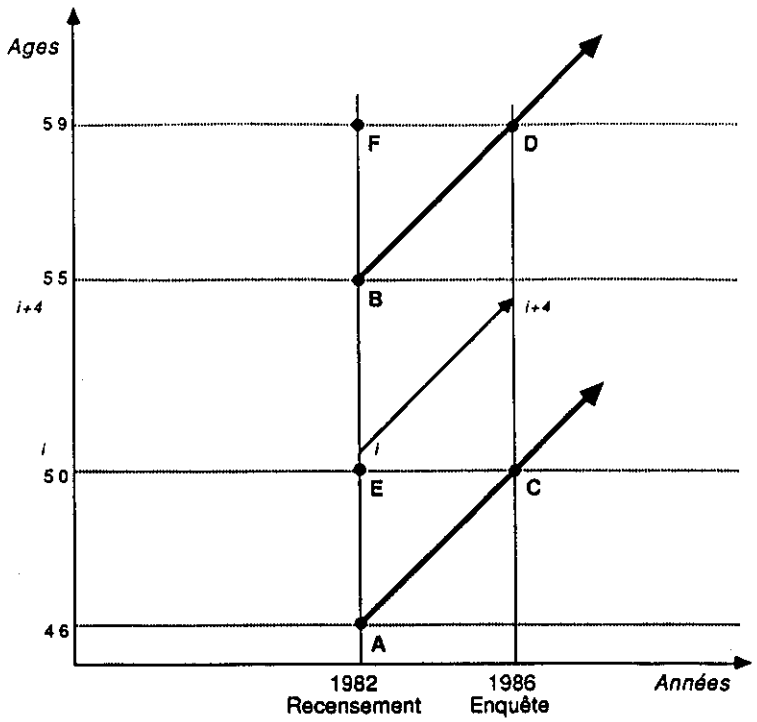
II.1 - Les sources de comparaison

D'une façon inhabituelle, les termes de la comparaison ne s'imposent pas a priori. Il n'existe pas de statistiques relatives strictement à notre population (les personnes âgées de 46 à 55 ans en 1982 résidant encore en 1986 dans le champ de l'enquête). Nous avons répertorié trois populations "voisines" permettant une comparaison au prix de certaines hypothèses. C'est donc à un **choix d'hypothèses théoriques** que nous sommes confrontés.

---

1 - Toutefois, les écarts pourraient cependant être révélateurs d'une procédure erronée de tirage ou de la mauvaise qualité de la base de sondage (si les statistiques sur la population ne sont pas issues elles-mêmes de notre base de sondage). Notre attitude critique peut donc également se porter sur les statistiques de référence.

FIGURE 1
Les populations de référence



LES POPULATIONS DE REFERENCE

(I) - La population des 46-55 ans à la date du recensement (AB)

C'est bien la population **sondée** (y **compris** les hors-champ) décrite avec ses caractéristiques **à la date du tirage**. La comparaison des structures s'avère donc valide, mais irréalisable et vaine : irréalisable car faute d'interview, nous ignorons la plupart des caractéristiques individuelles des hors-champ de notre échantillon (émigrants ou décédés à la date de l'enquête) comme celle des échecs. Elle est vaine puisque cette méthode ne tient pas compte du comportement récent des enquêtés, par exemple de leur récent parcours logement.

(II) - La population âgée de 50 à 59 ans au 1er janvier 1986 résidant à cette date dans la zone de l'enquête (CD) : c'est notre population dans laquelle les immigrés de la période 1982-1985 seraient réintroduits. Faute de recensement à cette date, elle n'est connue que par l'enquête Emploi.

(III) - La population âgée de 50 à 59 ans au 1er mars 1982 recensée dans la zone de l'enquête (EF). Cette population est formée de générations de 4 ans antérieures à celles de l'enquête. Leur histoire n'a pas été identique : ils ont vécu la guerre en adultes, se sont installés avant la Libération, mais surtout n'ont pas connu la baisse de l'âge des retraites (et pré-retraites) comme leurs successeurs. Ils n'ont donc pas connu le même régime de migrations tardives. Les statistiques en sont disponibles, y compris dans la typologie de l'IAURIF.

Le diagramme de Lexis (figure 1) permet d'illustrer ces trois populations (voir encadré). Les dates du recensement et de la collecte sont portées sur l'axe horizontal. On peut lire sur l'axe vertical l'âge atteint par une personne à toute date de sa vie. Ainsi la population **sondée** (I) est représentée par le segment AB. Agée de 45 à 55 ans en 1982, elle se retrouve en 1986 diminuée des individus décédés ou émigrés pendant cette période (et âgée de 4 ans de plus). Ces résidents de 1982 à 1986 constituent la population **à enquêter** (segment CD) mais se distinguent de la population résidente en 1986 (II) par l'absence des individus immigrés pendant cette période.

Si l'enquête fournit les caractéristiques des enquêtés en 1982, on peut comparer leurs distributions à celles du recensement. Conformément à la méthode longitudinale, on se réfère aux mêmes générations, mais en négligeant l'impact sélectif de l'émigration et de la mortalité. Les supposer négligeables serait une hypothèse forte, cependant tempérée par la brièveté de la période (4 ans).

On peut au contraire confronter les caractéristiques de l'échantillon en 1986 à celles observées au recensement pour les générations de 4 ans leurs aînées (III) (segment EF) car le groupe des 50-59 ans a, en effet, déjà été soumis à ces mêmes risques de mortalité et d'émigration. Ce serait supposer entre ces deux cohortes une homogénéité que l'Histoire et la conjoncture économique récente ont contrariée.

De ces trois populations de référence, nous choisissons la seconde car les 2 % d'immigrants ne peuvent aucunement perturber les distributions observées en 1986. Or nous disposons des résultats de l'enquête Emploi de Mars 1986. Son échantillon, proche de 12000 enquêtés habitant dans l'agglomération parisienne, est très suffisant pour confronter les distributions marginales des deux enquêtes. Contrairement à notre choix, l'enquête Emploi ne tient pas compte de la typologie des communes de l'IAURIF pour le tirage des communes-échantillons. Elle ne permet donc pas de ventiler nos distributions par strate, ni même d'en évaluer les populations. En conséquence, nous ne tiendrons pas compte de la stratification pour le redressement (voir partie III). Pour la même raison, nous devons limiter notre comparaison à l'agglomération parisienne, sans tenir compte de la classe des axes et vallées, inaccessible à l'enquête Emploi.

II.2 - Quelques écarts dans les distributions

Pour chaque sexe séparément, nous avons comparé les structures relatives aux critères suivants : âge, état matrimonial, activité économique, catégories socio-professionnelles, indicateur de nationalité, statut d'occupation du logement. Mais nous n'avons pas contrôlé la taille du ménage, partiellement liée à l'état matrimonial. La répartition entre immeuble collectif ou individuel aurait également mérité un contrôle.

La répartition par sexe de l'échantillon est parfaite (50,1 % et 49,6 % d'hommes dans les deux enquêtes). La distribution par âge est très satisfaisante : les écarts entre années successives ne présentent aucune variation systématique, même aux âges de la pré-retraite (Tableau 9).

TABLEAU 9

REPARTITION PAR AGE ET SEXE

AGE	51	52	53	54	55	56	57	58	59	60
Hommes Emploi	11,0	10,7	11,4	11,8	11,7	11,1	7,1	9,0	8,6	7,7
Hommes PDP	10,5	10,2	10,7	12,1	10,2	10,2	9,7	7,0	9,2	10,2
Femmes Emploi	10,2	9,8	11,9	10,0	11,2	9,4	11,0	7,7	9,3	9,6
Femmes PDP	8,7	8,7	11,0	11,7	10,0	12,7	10,0	9,3	9,0	8,8
TOTAL Emploi	10,6	10,3	11,6	10,9	11,4	10,2	9,1	8,3	9,0	8,6
TOTAL PDP	9,6	9,5	10,9	11,9	10,1	11,5	9,8	8,2	9,1	9,5

L'état matrimonial présente de sérieuses perturbations avec un déficit de l'enquête PDP en célibataires des deux sexes et en femmes divorcées. Plusieurs interprétations de ce constat se présentent : s'agit-il d'erreurs de déclaration ou de biais de réalisation de l'échantillon ? A quelle enquête imputer cette distorsion (Tableau 10) ?

TABLEAU 10

REPARTITION PAR ETAT MATRIMONIAL ET PAR SEXE

Etat matrimonial	Célibataires	Mariés	Divorcés	Veufs
Hommes Emploi	11,1	81,3	5,5	1,9
Hommes PDP	6,2	87,2	5,0	1,6
Femmes Emploi	9,9	68,3	10,9	10,7
Femmes PDP	7,2	74,8	7,1	10,9
TOTAL Emploi	10,5	74,7	8,3	6,3
TOTAL PDP	6,7	81,0	6,1	6,3

Plusieurs principes peuvent nous guider. L'échantillonnage aréolaire et l'accueil excellent réservé au thème de l'emploi évitent à l'enquête Emploi les biais d'échantillonnage habituels ou les atténuent très fortement. Nous les imputerons donc à l'enquête PDP. Les erreurs de déclaration proviennent surtout d'une précision insuffisante du questionnaire pour la question traitée. L'enquête Emploi serait donc la plus précise sur le thème de l'activité et l'enquête PDP, sur les points démographiques ou relatifs au logement.

Le **déficit de célibataires** ne relève pas d'erreurs de déclarations : il est exclu que les célibataires en union libre se soient plus souvent déclarés mariés dans l'enquête PDP, plus précise à ce sujet. C'est donc surtout un biais de l'échantillon, déjà noté dans l'enquête 3B (Biographie familiale, professionnelle et migratoire) : des célibataires peu enclins à parler de leur "non-vie" familiale refuseraient l'enquête. Nous préférons cette interprétation à la sous-représentation habituelle des personnes seules, liée à la difficulté de les contacter car les veufs et les veuves ne sont pas sous-représentés dans l'enquête PDP.

Les **femmes divorcées** sont-elles en sur-nombre dans l'enquête Emploi ? Par exemple, les femmes en instance de divorce s'y seraient-elles déclaré déjà divorcées ? Cet argument vaudrait aussi pour les hommes ; elle ne justifie pas une sous-évaluation du tiers de l'effectif. Par ailleurs, il serait incohérent de supposer que dans l'enquête PDP des femmes divorcées cachent davantage leur état matrimonial que dans l'enquête Emploi. Comme les

célibataires, les femmes divorcées auraient davantage refusé l'enquête PDP, contrairement aux hommes qui, se remettant en couple plus facilement, dissimuleraient moins la situation. En conclusion, nous acceptons pour vraie la distribution de l'enquête Emploi et soumettrons sur ce point l'enquête PDP au redressement.

La **population étrangère** est mal prise en compte par l'enquête PDP dont elle ne représente que 9,0 % de l'échantillon contre 13,6 % dans l'enquête Emploi. Pour plusieurs raisons, nous avons hésité à redresser l'enquête sur ce critère ; la population étrangère est très hétérogène ; la sous-représentation des étrangers est probablement assez sélective, ce qui rend peu fiable ce sous-échantillon limité (180 individus) et sensible à l'effet de grappe. La fréquence des étrangers et leur répartition ethnique est très variable d'une commune à l'autre, ce qui nuit à nos estimations malgré notre stratification fine. L'étude de la population étrangère à partir de cette enquête suppose un examen préalable rigoureux. Nous avons cependant pris en compte le critère de la nationalité actuelle dans le redressement (contrairement à notre choix pour l'enquête Triple Biographie déjà citée). Le taux général d'étrangers est réévalué mais la répartition des étrangers par nationalité a fort peu de chances d'en être améliorée. Les précautions préliminaires à toute étude de ce sous-échantillon n'en sont donc pas diminuées.

La représentation globale de **l'activité** est excellente (la différence entre taux d'activité est limitée à 3 pour 1000), mais au niveau de chaque sexe, l'écart n'est pas négligeable (Tableau 11).

TABLEAU 11

REPARTITION SELON LE SEXE ET L'ACTIVITE

	Actifs	Inactifs
Hommes Emploi	85,5	14,5
Hommes PDP	83,1	16,9
Femmes Emploi	59,2	40,8
Femmes PDP	61,9	38,1
TOTAL Emploi	72,2	27,8
TOTAL PDP	72,5	27,5

La **répartition professionnelle** des actifs se caractérise dans l'enquête PDP par une surreprésentation inattendue des indépendants des deux sexes et des professions intermédiaires et employés masculins ; en compensation, on y observe une sous-représentation des cadres supérieurs et surtout des ouvriers. Ce dernier point corrobore le moindre rendement des strates ouvrières du plan de sondage (Tableau 12).

TABLEAU 12

REPARTITION DES ACTIFS PAR SEXE ET CATEGORIES SOCIO-PROFESSIONNELLES

	Indépendants	Cadres Sup.	Professions intermédiaires	Employés	Ouvriers
Hommes Emploi	6,7	26,8	20,3	8,9	32,9
Hommes PDP	8,4	23,2	25,2	10,8	26,3
Femmes Emploi	11,0	12,2	21,7	47,1	12,1
Femmes PDP	14,1	11,5	22,8	47,8	9,5
TOTAL Emploi	9,3	20,8	20,9	24,7	24,3
TOTAL PDP	11,6	18,2	24,2	26,6	19,1

La répartition de la population par **statut d'occupation du logement** exige également un certain effort d'interprétation. Nous n'avons retenu que trois catégories de statut : propriétaires, locataires et logés gratuitement. L'enquête PDP présente un important déficit de locataires par rapport à l'enquête Emploi. En particulier, l'enquête de l'INED dénombre davantage de propriétaires que l'INSEE parmi les hommes (Tableau 13).

TABLEAU 13

REPARTITION PAR SEXE ET STATUT D'OCCUPATION DU LOGEMENT

Statut d'occupation	Propriétaire	Locataire	Logé gratuitement
Hommes Emploi	48,1	47,3	4,4
Hommes PDP	53,5	37,8	6,6
Femmes Emploi	50,7	44,8	4,3
Femmes PDP	53,8	39,6	6,7
TOTAL Emploi	49,4	46,0	4,3
TOTAL PDP	53,7	39,6	6,7

L'enquête Emploi est probablement peu précise dans sa distinction entre location et logement gratuit ou loué **à prix réduit**. En particulier, tout logement fourni par l'employeur entre dans cette dernière catégorie, même en présence d'un loyer. Cette imprécision se comprend dans un questionnaire non spécialisé. Au contraire, la confusion entre propriétaire et locataire est peu vraisemblable ; le taux de 49 % de propriétaires de l'enquête Emploi en est confirmé. L'enquête PDP présente donc un léger biais en faveur des propriétaires qui auraient mieux accepté l'enquête que les locataires. Ce serait une explication à la surreprésentation des indépendants qui auraient ainsi réagi en propriétaires intéressés au logement. En définitive, nous retenons le taux de propriétaires de l'enquête Emploi et la ventilation entre locataires et logés gratuitement de l'enquête PDP qui peuvent ainsi être confondus au stade du redressement.

L'enquête Emploi ne permet pas une comparaison précise de la **répartition géographique** de notre échantillon. Nous ne disposons que d'une ventilation entre Paris et la banlieue, excluant les axes et vallées. Nous ignorons donc la répartition de notre population entre les 10 strates en 1986. Avec un jeu d'hypothèses, nous aurions pu la reconstruire à partir de l'enquête Emploi. Nous avons renoncé à cette étape superflue pour le redressement, sans nous rendre compte de son utilité pour les calculs de variance (chapitre IV). Nous nous sommes limités à la population des strates regroupées par taux de sondage : Paris, la banlieue intérieure ou extérieure, enfin la frange, les villes nouvelles et les axes et vallées (Tableau 14). Cette

TABLEAU 14

REPARTITION DES GENERATIONS 1926-1935
DANS L'AGGLOMERATION PARISIENNE EN 1982 ET 1986

	Recensement 1982	Enquête Emploi 1986	Taux de départ
Paris	254 000	254 000	0 %
Banlieue	827 000	725 000	-12,3 %
ENSEMBLE	1 081 000	979 000	-9,4 %

construction repose sur des hypothèses implicites d'"émigration" différentielle par strate. Mais ces hypothèses touchent à l'**objet même de l'enquête**, ce qui est conceptuellement regrettable.

A Paris, l'immigration aurait compensé les décès et l'émigration parmi nos générations, essentiellement par un retour au centre des banlieusards, tandis que la banlieue aurait perdu 12 % de cette génération. Faute de mieux, nous devons supposer que les axes et vallées ont évolué comme la banlieue. Nous détachons ensuite les franges et les villes nouvelles de la banlieue pour les agréger aux axes et vallées. Pour ce faire, nous supposons, qu'elles ont eu une évolution homogène à celle de la banlieue, et homogène dans sa répartition par sexe (malgré les différentielles d'âge aux migrations de retraite) : ces hypothèses ne sont pas anodines, bien qu'elles passent pour de simples "règles de trois". Nous en déduisons le tableau 15.

TABLEAU 15

REPARTITION SPATIALE PAR SEXE

	Paris	Banlieue int. et ext.	Grande Banlieue *	TOTAL
Hommes Emploi	23,6	59,3	17,1	100,0
Hommes PDP	18,1	62,4	19,5	100,0
Femmes Emploi	25,0	57,9	17,1	100,0
Femmes PDP	21,4	60,8	17,8	100,0
TOTAL Emploi	24,3	58,6	17,1	100,0
TOTAL PDP	19,7	61,6	18,7	100,0

* Y compris les "Axes et vallées".

La sous-représentation de Paris dans l'enquête de l'INED se manifeste clairement; c'est la conséquence des échecs plus fréquents au centre de l'agglomération. Elle est plus marquée pour les hommes : les déficits relatifs s'établissent à 23 % et 14 % pour chacun des sexes. Le comportement voisin de la banlieue et de la grande banlieue nous rassure quant à l'effet limité des hypothèses précédentes.

En définitive, les écarts repérés ne sont pas considérables sauf pour la catégorie numériquement faible des étrangers, défaillante d'un tiers de l'effectif attendu. Une technique simple de redressement réduira ces irrégularités sans risque .

Du point de vue de la méthode, retenons la critique démographique des sources et l'interprétation réfléchie des différences qui suppose une connaissance spécialisée de ces sources. Manifestement cette phase du travail ne relève pas seulement du statisticien.

Pour récapituler, la confrontation de l'échantillon à une source statistique peut connaître trois types de distorsion : un décalage temporel, une frontière démographique dissemblable et un champ géographique légèrement distinct[1]. Nous avons réussi à éviter un décalage temporel au prix d'une petite hétérogénéité sur les deux autres plans. La difficulté est aussi reportée aux étapes ultérieures. On ne pourra affecter une pondération aux "Axes et vallées" qu'au prix d'une substitution d'hypothèses aux statistiques manquantes ; l'"oubli" de la stratification dans le redressement obligera à une autre gymnastique lors des calculs de variance. Il est manifestement difficile de maintenir une cohérence statistique complète tout au long de l'enquête.

III - LE REDRESSEMENT DE L'ECHANTILLON

Tout tableau relatif à l'ensemble de l'échantillon PDP nécessite l'emploi d'une pondération. Destinée à compenser l'inégalité volontaire des taux de sondage entre strates, cette pondération n'est autre que l'inverse des taux de sondage. Le redressement consiste à calculer une autre pondération qui per-

1 - Si un concept a été mis en oeuvre d'une façon différente dans les deux sources, la comparaison en sera également perturbée.

met d'ajuster l'échantillon sur les distributions connues de la population tout entière. Si les écarts observés sont faibles, on peut très bien éviter le redressement. En fait, nous l'avons trouvé opportun pour cette enquête (voir chapitre II) et réalisé en collaboration avec Arnaud BRINGE.

L'opération de redressement est souvent conçue comme une mécanique appliquée systématiquement à toutes les enquêtes. C'est oublier les quelques principes théoriques sous-jacents aux algorithmes mis en oeuvre. Avant de présenter les résultats, nous allons nous arrêter sur les procédures utilisées et leurs hypothèses implicites.

Nous n'avons pas eu recours à la technique de la post-stratification qui ne bénéficie pas de «l'usage universel» du redressement : la post-stratification ne fournit l'estimation optimale que d'une variable prédéfinie. Pour une autre variable, il faudra réaliser une autre post-stratification à partir d'un critère d'ajustement différent. Dans cette enquête aucune variable d'analyse ne s'impose comme prépondérante ; de là vient notre choix d'une technique plus neutre de redressement.

III. 1 - Les hypothèses

Tout redressement d'un échantillon repose sur une substitution de certains enquêtés à des personnes défaillantes. Elle suppose que certains répondants sont représentatifs de non-répondants.

Cette hypothèse était explicite à l'époque du traitement mécanographique ; les cartes perforées relatives à certains enquêtés étaient dupliquées pour rétablir un échantillon représentatif. Actuellement encore, dans une enquête sur adresses, on dénombre les échecs de la collecte (665 dans cette enquête selon le tableau 3) ; les informations absentes sont compensées par celles fournies par les répondants, sans qu'on puisse dire qu'une personne réponde pour 2 ou 3 autres. (Les pondérations de redressement sont plus souvent de l'ordre de 1.2 que 2 ou 3).

Dans une enquête sur quotas, l'enquêteur recrute ses enquêtés sans qu'on puisse dénombrer les refus. La catégorie des non-répondants s'évanouit au risque de faire oublier que la difformité de l'échantillon tient à

Trois types de pondération

A - La pondération a priori ou probabiliste

Il s'agit de l'inverse des taux de sondage. Elle fournit l'estimation correcte (la somme dilatée de Horvitz-Thomson) à partir d'un sondage à probabilité inégale. C'est donc une simple généralisation de la loi mathématique des grands nombres. Dans les sondages auto-pondérés, ou à taux uniforme, elle vaut 1 et on n'en parle pas.

B - Les pondérations a posteriori ou de post-stratification

Les distributions d'un échantillon parfaitement réalisé (taux d'échec nul) diffèrent de celles de l'univers sondé par le simple fait de l'échantillonnage.

En connaissance des distributions exactes sur l'univers, on n'a aucune raison de s'appuyer sur la distribution imparfaite de l'échantillon : on se «recalle» sur l'univers par simple règle de trois. Ce facteur correctif est la pondération de post-stratification. Par construction, elle assure dans l'enquête un écart nul avec l'univers pour le critère de post-stratification ; elle réduit de plus les fluctuations des autres variables en raison directe de leur corrélation avec le critère de post-stratification.

Elle relève donc de la seule théorie mathématique de l'estimation.

C - Les pondérations de redressement

Quand des biais de sélection sont intervenus lors de la réalisation imparfaite d'un échantillon, on recalle certaines distributions de l'échantillon sur celles de l'univers.

La post-stratification et les redressements selon un critère unique reposent sur le même algorithme mais n'ont pas du tout le même statut théorique. Les estimations issues d'une post-stratification pure sont vraies dans la limite d'un intervalle de confiance et d'un seuil de signification. Cette mesure de l'incertitude ne peut être conférée au terme d'un redressement qu'en posant un nouveau modèle probabiliste aux hypothèses plus lourdes et moins vérifiables que la bonne exécution d'un tirage ou de la collecte. C'est ce qu'on appelle un modèle de comportement ou de surpopulation. Il consiste à réinterpréter les biais significatifs d'un échantillon comme le résultat d'un autre tirage probabiliste mettant en oeuvre des probabilités inégales de répondre (ou encore plus inégale). Par exemple, un commerçant aurait six chances sur dix de répondre, pendant qu'un employé de bureau en aurait 9 sur 10 [9].

***

Ces distinctions sont le plus souvent omises dans les enquêtes par quotas : on peut toujours y réinterpréter les biais en terme de probabilité différentielle d'avoir été contactée : la probabilité a priori d'inclusion d'un individu spécifié dans l'échantillon (sur quotas) ne peut être mesurée et n'a probablement aucun sens[1].

---

1 - Ceci ne met nullement en cause la qualité empirique des enquêtes sur quotas ni l'intérêt de leur associer des modèles théoriques.

leurs désistements, comme à l'inégale probabilité de contact (et alors le redressement opéré serait le simple rétablissement de la pondération classique).

D'un point de vue théorique, un redressement se fonde sur une hypothèse d'homogénéité entre répondants et non répondants. L'apparition même du biais nie cette équivalence au niveau de l'échantillon complet. Dans notre cas, les répondants comprennent une moindre proportion de célibataires que les non-répondants. Cependant, les répondants célibataires pourraient constituer un pseudo-tirage représentatif de l'ensemble des célibataires et à ce niveau l'hypothèse d'homogénéité serait valide -ou le serait davantage-.

Le statisticien HANSEN a déduit de ce raisonnement une méthode de redressement théoriquement infaillible : il suffit d'enquêter un échantillon représentatif de non-répondants, en opérant une phase de résorption. C'est bien ce qui a été tenté auprès des personnes qui avaient refusé, mais avec un succès partiel.

On pourrait alors faire l'hypothèse que les refus résorbés et ceux réaffirmés relèvent d'une population homogène, de même que les migrants d'Ile de France retrouvés ou non. En pondérant chaque questionnaire résorbé par l'inverse du taux de succès de la résorption, on obtiendrait une pondération correcte sous cette dernière hypothèse. La prudence s'impose : les distributions de l'enquête PDP pondérées selon cette méthode ne se rapprochent nullement de celles de l'enquête Emploi. La résorption des refus serait donc intervenue avec le même effet de sélection que les premiers refus !

III. 2 - La méthode

Nous avons appliqué la méthode classique du redressement RAS, fondée sur un calcul par itérations. On réalise successivement un ajustement proportionnel sur chaque critère dont on veut assurer la distribution. Le principe en est donc très simple.

Supposons en premier exemple que dans une population comprenant moitié d'hommes, on ait tiré un échantillon de 1000 personnes où ne figurent que 40 % d'hommes (à la suite d'un tirage à probabilités inégales entre hommes et femmes). La théorie des sondages nous commande de pondérer

chaque observation par un coefficient inversement proportionnel aux probabilités de tirage. Les 400 hommes munis d'une pondération POND(1) de 5/4 (50/40) équilibrent alors les 600 femmes recevant le poids 5/6 (50/60) (voir exemple joint).

Cette méthode conforme à la règle statistique suppose que tous les hommes ont été tirés avec la même probabilité, et par ailleurs toutes les femmes, c'est-à-dire qu'il n'y a pas eu de sélection due à des refus.

Supposons maintenant en second exemple que le précédent échantillon était réparti par moitié entre grandes et petites villes, conformément à la réalité.

La pondération calculée précédemment redresse la répartition par sexe, mais perturbe de 2 % celle par type de ville. Une seconde itération conduit à multiplier la pondération antérieure POND(1) par 50/48 pour un enquêté d'une grande ville et par 50/52 pour celui d'une petite ville.

Si la pondération par sexe en est affectée, on réalise une troisième itération. En fait, le calcul converge rapidement vers des pourcentages stables comme l'indique l'écart minime (0,6 %) avec la répartition souhaitée par sexe. Ce principe mécanique est applicable à un grand nombre de critères qu'on redresse successivement avant de revenir au premier pour une nouvelle itération.

Comme plusieurs tableaux croisés peuvent avoir les mêmes marges, le tableau ne converge pas nécessairement vers le tableau "vrai", mais vers le tableau "le plus proche" satisfaisant la contrainte des marges. La proximité entre tableaux peut recevoir plusieurs définitions mathématiques et donc d'autres algorithmes conduisent à d'autres pondérations optimales selon d'autres points de vue [10,11].

Si l'échantillon est très déformé, il est même improbable que la convergence conduise à ce tableau vrai. En particulier, lorsqu'une case de tableau est nulle, les règles de trois sont impuissantes à lui fournir un effectif non nul; c'est par déformation abusive des cases voisines qu'on compensera sa contribution nulle aux marges. C'est une limite des redressements : ils ne peuvent inventer l'information; il faut donc éviter de le leur demander. Pour cette raison, M. DEROO et A.M. DUSSAIX [12] conseillent de limiter le nombre

de critères et donc de limiter les cases nulles et H. JACQUART [13] recommande que les critères pris isolément ne comprennent pas de catégorie de trop faible fréquence. Nous suivons sur ce point H. JACQUART, plus que les précédents auteurs : pour le redressement de l'enquête 3B [14], nous avions hésité à maintenir trois critères bien corrélés aux quatre critères principaux ; finalement nous les avions retenus : leur représentation en était très améliorée pour un accroissement très faible de la dispersion de la pondération, donc pour une faible «trituration» additionnelle des données. La vérité empirique est difficile à établir.

1ER EXEMPLE DE REDRESSEMENT

	Echantillon		Population	
SEXE	Masculin	Féminin	Masculin	Féminin
	40 %	60 %	50 %	50 %
POND	5/4	5/6		

2EME EXEMPLE DE REDRESSEMENT

Echantillon

	Masculin	Féminin	Ensemble
Grandes villes	15 %	35 %	50 %
Petites villes	25 %	25 %	50 %
Ensemble	40 %	60 %	100 %

Population

Masculin	Féminin	Ensemble
25 %	25 %	50 %
25 %	25 %	50 %
50 %	50 %	100 %

Première itération M = 5/4 F=5/6

Grandes villes	18,75 %	29,2 %	48 %
Petites villes	31,25 %	20,8 %	52 %
Ensemble	50 %	50 %	100 %

POND (1)
= 1,25 M.GV
= 0,833 F.GV
= 1,25 M.PV
= 0,833 F.PV

Seconde itération GV = 50/48 (1,042) PV = 50/52 (0,96)

Grandes villes	19,54 %	30,42 %	50 %
Petites villes	30,06 %	20,01 %	50 %
Ensemble	49,60 %	50,40 %	100 %

POND (2)
= 1,30 M.GV
= 0,87 F.GV
= 1,20 M.PV
= 0,80 F.PV

Troisième itération

III.3 - Les résultats

Nous avons eu recours à la routine REDRE du logiciel SPAD élaboré par L. LEBART et A. MORINEAU. Nous l'avons appliquée à sept critères croisés avec le sexe que l'on trouvera au tableau joint. De la sorte, les redressements des échantillons masculin et féminin sont indépendants.

Le premier critère regroupe les strates d'égal taux de sondage. Ainsi le programme commence par rétablir la pondération issue du sondage à probabilité inégale, puis déforme celle-ci au cours de quinze itérations.

Les distributions recherchées et obtenues sont très proches. Les trois écarts les plus forts, de l'ordre de 7 pour mille, concernent les femmes de proche banlieue, "sans catégorie sociale attribuable" ou au foyer. Le dernier écart exprime un classement contradictoire selon ces sources. Dans l'enquête Emploi, toutes les enquêtées inactives non retraitées étaient classées à la fois au foyer et "sans catégorie sociale attribuable". Dans l'enquête PDP, 2,6 % d'actives à la profession mal déclarée font diverger ces deux groupes qui se voient affecter la même fréquence 17.2 théorique alors que l'un n'est qu'un sous-ensemble de l'autre. (Les non-déclarations entraînent des difficultés [13], compliquées ici par l'emboitement des critères). Les écarts par rapport à la distribution recherchée sont en général le témoin d'un problème logique dans les données plus grave en général que celui-ci (par exemple, la substitution pour une des sources du classement par âge croissant à celui par année de naissance croissante !). Ils doivent donc être soigneusement contrôlés.

Le recours à 15 itérations (en fait 15 x 7 calculs de pondération) est certainement superflu [13]; nous n'avons pas observé la convergence du

TABLEAU 16
RESULTATS DU REDRESSEMENT APRES 15 ITERATIONS

	Pourcentages demandés	Pourcentages obtenus
VARIABLE 1 : SEXE STRATE		
Hommes * Paris	11,70	12,21
Hommes * Proche banlieue	29,40	29,01
Hommes * Grande banlieue	8,50	8,56
Femmes * Paris	12,60	13,15
Femmes * Proche banlieue	29,20	28,52
Femmes * Grande banlieue	8,60	8,55
VARIABLE 2 : SEXE AGE		
Hommes * 1931-1935	28,00	28,26
Hommes * 1926-1930	21,60	21,52
Femmes * 1931-1935	26,70	26,43
Femmes * 1926-1930	23,70	23,79
VARIABLE 3 : SEXE STATUT MATRIMONIAL		
H * Célibataires	5,50	5,18
H * Mariés	40,40	40,92
H * Veufs divorcés	3,70	3,63
F * Célibataires	5,00	4,67
F * Mariées	34,50	34,86
F * Divorcées	5,50	5,23
F * Veuves	5,40	5,41
VARIABLES 4 : SEXE PROPRIETAIRES		
Hommes * Propriétaires	23,90	23,91
Hommes * Locataires (1)	25,70	25,87
Femmes * Propriétaires	25,60	25,37
Femmes * Locataires (1)	24,80	24,85
(1) ou logés gratuitement		
VARIABLE 5 : SEXE NATIONALITE		
Hommes * Français	40,80	41,16
Hommes * Etranger	8,80	8,62
Femmes * Français	45,60	45,59
Femmes * Etranger	4,80	4,63
VARIABLE 6 : SEXE CSP MIXTE (ACTIFS ET RETRAITES)		
H * Indépendants	5,00	5,04
H * Cadres supérieurs	12,80	12,16
H * Catégories intermédiaires	10,20	10,04
H * Employés	5,60	5,47
H * Ouvriers	16,00	15,68
F * Indépendants	2,30	2,34
F * Cadres supérieurs	3,70	3,78
F * Catégories intermédiaires	7,00	7,24
F * Employées	15,90	16,04
F * Ouvriers	4,30	4,30
F * Sans CS attribuable	17,20	17,90
VARIABLE 7 : SEXE TYPE D'ACTIVITE		
H * Actifs	45,20	45,11
H * Inactifs	4,30	4,67
F * Inactives	30,00	30,37
F * Foyer	17,20	16,57
F * Retraitées	3,30	3,28

calcul, pour la réduire au minimum utile. Il n'est pas sain de torturer les données, car on ne maîtrise pas les compensations qui s'établissent entre cases du tableau croisé dans un tel calcul d'ajustement des marges. Ces procédures seraient dangereuses sur de trop mauvais échantillons dont elles pourraient ne corriger que l'apparence : les distributions marginales connues.

Par ailleurs, chaque itération accroît la dispersion de la pondération et par le fait même la variance des estimations[1] . On ne gagne rien à réduire des biais minimes au prix d'une fluctuation élevée. Ce coût des estimations redressées s'apprécie à partir du coefficient de variation de la pondération (rapport de son écart-type à sa moyenne). Nul pour un tirage uniforme (de pondération constante), ce coefficient s'établit à 0,26 à la suite du triple taux de sondage et à 0,32 après le redressement[2] . Ceci signifie en première approximation un accroissement de 23 % des intervalles de confiance des estimations. Le chapitre 4 nous permettra de confronter cet effet à celui du plan de sondage.

IV - L'INCERTITUDE D'ECHANTILLONNAGE

Les calculs de variance [15] ont été établis par François MARCHAND à partir du logiciel SESUDAAN de RTI (Research Triangle Institute, North Carolina 27709). Ces procédures sont écrites en langage SAS mais sans être intégrées au système. Elles jouissent d'une facilité d'usage inconnue des logiciels antérieurs. Par contre, elles ne permettent pas de décrire les plans de sondage les plus complexes sans quelques hypothèses parfois trop simplificatrices.

---

1 Sauf si le critère de redressement est (fortement) corrélé avec la variable analysée (sondage de Neyman et post-stratification heureuse).

2 . Dans le même temps, l'amplitude des poids extrêmes (rapportée à un même poids moyen de valeur 1) passe de 1,25 à 2,81 pour les valeurs élevées et de 0,65 à 0,32 pour les valeurs faibles. Ces poids extrêmes rares et relativement modérés ne sont pas de nature à perturber les estimations des sous-groupes. On les aurait probablement réduits en diminuant le nombre d'itérations. Parmi les 29 enquêtés dont le poids atteint la valeur 2 figurent 14 étrangers (dont 10 femmes); ils ne comptent pourtant que pour 9 % de l'échantillon, mais nous savions cette sous-population très mal représentée dans l'échantillon.

IV.1 - La méthode

L'enquête Peuplement et Dépeuplement de Paris semblait l'exercice d'école parfaitement à la mesure du logiciel. Quelques nuances s'imposent. Nous pouvons néanmoins résumer ainsi la situation :

- Le plan de sondage est parfaitement connu de l'utilisateur pour avoir été conçu à l'INED même. Trop de chercheurs reçoivent un fichier sans indication de la stratification, de l'arborescence des grappes et de la distinction entre unités retenues d'office et celles tirées au hasard -sans même parler des pondérations inconnues-. Ils sont donc dans l'impossibilité de réaliser un calcul d'erreur précis.

- Tous les ingrédients classiques du sondeur figurent dans le plan de sondage : stratification, tirage à deux degrés et taux de sondage inégaux entre les strates.

Rappelons l'effet de ces différents outils : le découpage de la population en sous-populations (les strates) donnant lieu à autant de sous-échantillons indépendants améliore nécessairement la précision des estimations.

Le tirage en deux temps (d'une commune puis des enquêtés de la commune) permet de réduire les déplacements d'enquêteurs et donc le coût à taille d'échantillon fixé, mais il réduit la précision des estimations toujours à taille d'échantillon constante. L'économie réalisée sur les déplacements d'enquêteurs (grâce à ce mode de tirage) accroît la précision des estimations à budget constant car il permet d'enquêter un échantillon de dimension bien supérieure.

La précision des estimations est très sensible au nombre d'unités primaires (les communes) qu'on veillera donc à garder en nombre suffisant, et bien distribuées grâce à la stratification géographique.

Certaines communes sont tirées d'office et forment ensemble une strate unique. Leurs enquêtés sont donc tirés à un seul degré selon une procédure plus précise. Aussi avons-nous recouru à ce procédé pour les communes de plus grande taille ainsi que pour l'ensemble des arrondissements parisiens.

Le sondage à probabilités inégales était destiné à permettre une exploitation isolée pour la capitale, mais le gain relatif à cet objectif se paie par une précision sensiblement moindre des estimations générales.

Nous n'avons pas contraint le redressement à se caler sur l'effectif exact de chaque strate (d'ailleurs inconnu pour l'année 1986 de l'enquête), mais seulement sur ceux des trois regroupements de strates. Nous négligeons cet aléa pour bénéficier pleinement de l'effet de stratification des unités primaires.

Nous négligeons enfin l'effet bénéfique du tirage systématique (le fait de tirer un individu tous les 200, ou une commune tous les 9000 individus). Ce dernier répartit régulièrement les enquêtés, mieux que ne le ferait le hasard. Il réalise une quasi-stratification efficace, mais difficile à intégrer

Précisons enfin la notion d'**effet de plan de sondage** DEFF : c'est le rapport de la variance des deux estimations issues pour l'une de notre échantillon "complexe" et pour l'autre d'un tirage simple (à un seul degré, un seul taux de sondage, "avec remise"). C'est l'indicateur du gain ou de la perte relative de précision due à un échantillon de structure complexe. Attention, c'est la racine carrée du DEFF qui décrit l'incidence du plan de sondage sur les intervalles de confiance des estimations.

IV.2 - Les principaux résultats

Nous allons d'abord présenter les résultats relatifs à 6 variables dont quatre qualitatives (lieu de naissance, nationalité, type de logement à 25 et à 50 ans) et deux quantitatives (le nombre d'enfants à l'enquête et le nombre de logements personnels occupés avant 25 ans). Nous fournirons les estimations pour l'ensemble de l'échantillon et par strate, mais pas par sous-ensemble transversal aux strates (les "domaines"). Dans ce dernier cas, la précision est généralement moindre que pour une strate ou un ensemble de strates. Si le domaine est régulièrement réparti entre les unités primaires, les estimations sur ce domaine pourront être précises. C'est le cas, par exemple, des deux sous-populations masculine et féminine. Si au contraire l'appartenance au domaine est très inégalement probable selon les unités primaires d'une strate, alors les estimations sur ce domaine seront très fragiles. Ainsi, nous devrons éviter toute estimation sur un département de banlieue, représenté par un nombre d'unités primaires strictement

Dans le cas d'une proportion, cinq estimations sont fournies : sa fréquence dans l'échantillon (ou la strate), et l'écart-type de cette estimation, la taille de l'échantillon concerné, la taille de la population correspondant à ce même échantillon et l'effet de plan de sondage pour cette estimation.

40,3 % des enquêtés sont nés en Ile de France. Notre échantillon représente 1046175 individus en 1986 . L'écart-type de 0,013 (1,3 %) doublé de chaque côté de l'estimation permet d'affirmer que la proportion vraie de ces franciliens nés en Ile de France se situe entre 37,7 % et 42,9 % (avec une certitude de 95 chances pour cent).

Les Parisiens de chacune des deux strates ouvrière et bourgeoise sont nés en Ile de France dans une proportion d'à peine 36 % ±6 % (les effectifs faibles des strates accroissent l'intervalle de confiance), tandis que ceux des axes et vallées (strate 10) paraissent plus souvent originaires de la région (45% ±11,5 %). Nous voyons que les intervalles de confiance se recoupent. L'écart ne peut être affirmé significatif. Par contre, le regroupement très homogène des strates les plus centrales (1 et 2) et les plus extérieures (9 et 10) pourrait permettre de conclure positivement.

L'effet de plan de sondage est relativement élevé (1,36 pour l'ensemble de l'échantillon), beaucoup plus faible pour les strates 1 à 5 (de 1.03 à 1.33), inexplicablement fort pour la strate 6, inverse pour la strate 7 (0,75) et élevé pour les strates 8 à 10 (de 1,41 à 2). Ces chiffres appellent des explications :

- Il s'ensuit un accroissement de 17 % de l'intervalle de confiance par rapport au tirage simple, principalement sous l'effet de la pondération. Le tirage à plusieurs degrés semble de peu d'effet sur cette variable. Prenons à témoin la marge faible des strates 3 à 5 où le taux de sondage est constant. C'est encore vrai des strates 1 et 2 tirées à un seul degré : on s'y attend à un coefficient égal à 1 (tirage simple) ou même légèrement inférieur à 1 (au bénéfice du tirage systématique). Mais le redressement a introduit une pondération inégale dont on mesure l'effet modéré sur la variance (+ 8 % et + 3 %).

- Le niveau élevé du DEFF dans la strate 6 (banlieue extérieure bourgeoise) appellerait une étude plus précise.

Tableau 17

ESTIMATION DE LA PRECISION DE PROPORTIONS ET DE MOYENNES

LIEU DE NAISSANCE DES ENQUÊTÉS

		Ensemble	Strate 1	Strate 2	Strate 3	Strate 4	Strate 5	Strate 6	Strate 7	Strate 8	Strate 9	Strate 10
lieu de naissance												
Ile de France	PVAL	0.402302	0.357210	0.362364	0.396886	0.473026	0.362785	0.409175	0.389501	0.437285	0.472650	0.453502
	STDERR	0.012363	0.028939	0.027197	0.032974	0.030666	0.037391	0.044431	0.050124	0.071582	0.073078	0.057736
	SAMSIZE	1986.	294.	323.	245.	275.	221.	251.	71.	96.	66.	144.
	POPEST	1046175.	127253.	126328.	163711.	165906.	141111.	141715.	38841.	40956.	32675.	67674.
	DEFF	1.3660	1.0750	1.0333	1.1230	1.0375	1.3365	2.0496	0.7502	1.9990	1.4141	1.9368
Province	PVAL	0.346396	0.349028	0.348655	0.293932	0.303656	0.382135	0.402954	0.339090	0.420682	0.319022	0.321719
	STDERR	0.011433	0.028403	0.027576	0.028564	0.027960	0.032011	0.036531	0.027109	0.081240	0.075280	0.047215
	SAMSIZE	1986.	294.	323.	245.	275.	221.	251.	71.	96.	66.	144.
	POPEST	1046175.	127253.	126328.	163711.	165906.	141111.	141715.	38841.	40956.	32675.	67674.
	DEFF	1.1459	1.0439	1.0816	0.9632	1.0167	0.9591	1.3923	0.2176	2.5998	1.7217	1.4711
Dom-Tom	PVAL	0.012731	0.024959	0.012075	0.018018	0.001981	0.031473	0.006702	0.0	0.0	0.0	0.0
	STDERR	0.002373	0.011394	0.006053	0.009596	0.002002	0.011799	0.006565	0.0	0.0	0.0	0.0
	SAMSIZE	1986.	294.	323.	245.	275.	221.	251.	71.	96.	66.	144.
	POPEST	1046175.	127253.	126328.	163711.	165906.	141111.	141715.	38841.	40956.	32675.	67674.
	DEFF	1.2993	1.5634	0.9936	1.2751	0.5575	1.0093	1.6249	1.0000	1.0000	1.0000	1.0000
Etranger	PVAL	0.234783	0.268903	0.276906	0.292105	0.217189	0.218696	0.181169	0.211409	0.142033	0.208328	0.224779
	STDERR	0.011150	0.028995	0.027553	0.034645	0.031385	0.026180	0.029045	0.066328	0.050058	0.055907	0.033343
	SAMSIZE	1986.	294.	323.	245.	275.	221.	251.	71.	96.	66.	144.
	POPEST	1046175.	127253.	126328.	163711.	165906.	141111.	141715.	38841.	40956.	32675.	67674.
	DEFF	1.3742	1.2459	1.2247	1.4221	1.5933	0.8865	1.4274	1.7623	1.9741	1.2508	0.9465

NATIONALITE

		Ensemble	Strate 1	Strate 2	Strate 3	Strate 4	Strate 5	Strate 6	Strate 7	Strate 8	Strate 9	Strate 10
Français de naissance	PVAL	0.804022	0.757663	0.776663	0.723187	0.822713	0.806622	0.872496	0.858833	0.900838	0.823122	0.806924
	STDERR	0.011292	0.028246	0.025716	0.039544	0.029796	0.030180	0.023162	0.073028	0.030125	0.050442	0.037662
	SAMSIZE	1986.	295.	322.	245.	275.	221.	251.	71.	96.	66.	144.
	POPEST	1046259.	127771.	125907.	163711.	165906.	141111.	141715.	38841.	40956.	32675.	67674.
	DEFF	1.5072	1.2818	1.3145	1.8192	1.5739	1.2905	1.2104	3.1231	0.9753	1.1534	1.2971
Naturalisé	PVAL	0.062357	0.061656	0.051028	0.074195	0.047720	0.077435	0.076047	0.056951	0.038621	0.064208	0.043547
	STDERR	0.005550	0.013677	0.012362	0.015219	0.011327	0.017635	0.019474	0.037523	0.020203	0.030422	0.018250
	SAMSIZE	1986.	295.	322.	245.	275.	221.	251.	71.	96.	66.	144.
	POPEST	1046259.	127771.	125907.	163711.	165906.	141111.	141715.	38841.	40956.	32675.	67674.
	DEFF	1.0844	0.9538	1.1000	0.8262	0.7765	0.9676	1.3547	1.8544	1.0554	1.0166	1.0333
Etranger	PVAL	0.133522	0.180702	0.152309	0.202623	0.129568	0.115943	0.051459	0.084317	0.060542	0.112670	0.144529
	STDERR	0.010068	0.026622	0.023900	0.037079	0.027481	0.024535	0.012324	0.065577	0.021175	0.043500	0.032917
	SAMSIZE	1986.	295.	322.	245.	275.	221.	251.	71.	96.	66.	144.
	POPEST	1046258.	127771.	125907.	163711.	165906.	141111.	141715.	38841.	40956.	32675.	67674.
	DEFF	1.7389	1.4132	1.4245	2.0849	1.8415	1.2979	0.7810	3.9545	0.7568	1.2492	1.2519

TYPE DE LOGEMENT A 25 ANS

		Ensemble	Strate 1	Strate 2	Strate 3	Strate 4	Strate 5	Strate 6	Strate 7	Strate 8	Strate 9	Strate 10
Individuel	PVAL	0.607336	0.734828	0.763022	0.534443	0.638591	0.544681	0.641745	0.541158	0.596088	0.434724	0.393922
	STDERR	0.015139	0.029070	0.027328	0.050452	0.029274	0.050902	0.032165	0.069641	0.049778	0.086194	0.064865
	SAMSIZE	1818.	267.	291.	226.	252.	207.	230.	69.	80.	61.	135.
	POPEST	958352.	115075.	112515.	149607.	151899.	132665.	130334.	37504.	34605.	30427.	63731.
	DEFF	1.7481	1.1579	1.2019	2.3120	1.6842	2.1627	1.0351	1.3477	0.8233	1.8442	2.3791
Collectif	PVAL	0.372459	0.239131	0.234220	0.432424	0.336721	0.425710	0.353891	0.443794	0.403912	0.546375	0.586703
	STDERR	0.014482	0.027926	0.027226	0.045067	0.035285	0.052176	0.032249	0.060992	0.049778	0.094435	0.064729
	SAMSIZE	1818.	267.	291.	226.	252.	207.	230.	69.	80.	61.	135.
	POPEST	958352.	115075.	112515.	149607.	151899.	132665.	130334.	37504.	34605.	30427.	63731.
	DEFF	1.5312	1.1444	1.2026	1.8702	1.4048	2.3050	1.0460	1.0399	0.8233	2.1949	2.3327

TYPE DE LOGEMENT A 50 ANS

		Ensemble	Strate 1	Strate 2	Strate 3	Strate 4	Strate 5	Strate 6	Strate 7	Strate 8	Strate 9	Strate 10
Individuel	PVAL	0.687027	0.962472	0.958229	0.722844	0.749761	0.607530	0.596114	0.323626	0.283060	0.360802	0.387734
	STDERR	0.015350	0.011123	0.010983	0.035643	0.034693	0.061359	0.040604	0.105474	0.143730	0.073478	0.062256
	SAMSIZE	1984.	295.	323.	244.	274.	221.	250.	71.	96.	66.	144.
	POPEST	1044058.	127771.	126328.	163109.	164517.	141111.	141077.	38841.	40956.	32675.	67674.
	DEFF	2.1742	1.0135	0.9734	1.5473	1.7375	3.4896	1.7120	3.6085	9.7725	1.5451	2.3510
Collectif	PVAL	0.309506	0.032710	0.035712	0.274805	0.245779	0.392470	0.399747	0.662522	0.716940	0.639198	0.612266
	STDERR	0.015285	0.010088	0.010156	0.035979	0.034332	0.061359	0.041708	0.094907	0.143730	0.073478	0.062256
	SAMSIZE	1984.	295.	323.	244.	274.	221.	250.	71.	96.	66.	144.
	POPEST	1044058.	127771.	126328.	163109.	164517.	141111.	141077.	38841.	40956.	32675.	67674.
	DEFF	2.1688	0.9439	0.9674	1.5849	1.7422	3.4896	1.8124	2.8603	9.7725	1.5451	2.3510

VARIABLES QUANTITATIVES

		Ensemble	Strate 1	Strate 2	Strate 3	Strate 4	Strate 5	Strate 6	Strate 7	Strate 8	Strate 9	Strate 10
Nombre d'enfants	MEAN	2.26	1.93	1.77	2.60	1.97	2.44	2.45	2.51	2.19	1.98	3.09
	STDERR	0.05	0.12	0.10	0.17	0.11	0.13	0.12	0.08	0.14	0.16	0.20
	SAMSIZE	1936.	295.	322.	245.	275.	221.	251.	71.	96.	66.	144.
	POPEST	1046263.	127771.	125907.	163711.	165306.	141111.	141715.	38841.	40956.	32675.	67674.
	DEFF	1.3398	1.2732	1.1435	1.5063	1.2037	1.3437	1.1575	0.2385	1.0634	1.2784	0.9378
Nombre de logements à 25 ans	MEAN	1.31	1.32	1.28	1.30	1.27	1.29	1.36	1.27	1.42	1.14	1.36
	STDERR	0.02	0.04	0.03	0.05	0.05	0.05	0.05	0.05	0.12	0.05	0.05
	SAMSIZE	1318.	267.	291.	226.	252.	207.	230.	69.	80.	61.	135.
	POPEST	958352.	115075.	112515.	149607.	151899.	132665.	130334.	37504.	34605.	30427.	63731.
	DEFF	1.2473	0.9773	1.0007	1.4263	1.5873	0.9710	1.2106	0.7035	1.9323	0.9693	0.5690

- Les effectifs très faibles rendent les estimations très aléatoires pour les trois ou quatre dernières strates, y compris celle du coefficient DEFF décrivant l'effet de plan de sondage.

L'effet de plan de sondage est beaucoup plus faible pour la proportion de personnes nées en province dont la variation intercommunale paraît faible. L'effectif des originaires des DOM-TOM est dérisoire.

La proportion de personnes nées à l'étranger (23 % ± 2 %) paraît assez homogène entre strates ; (n'oublions pas qu'elle comprend des Français de naissance et des naturalisés). L'effet de plan de sondage paraît du même ordre que pour les individus nés en Ile de France. Il est cependant élevé dans les strates 1 et 2 sous l'effet de la pondération, très sensible au biais de représentation des étrangers.

Si 23 % des enquêtés sont nés à l'étranger, 80,4 % d'entre eux sont pourtant Français de naissance. Les autres se répartissent en 13,4% d'étrangers et 6,2 % de naturalisés. L'écart-type de ces estimations est plus faible que pour le lieu de naissance, comme pour toute fréquence s'approchant des valeurs extrêmes 0 ou 1. Cependant, l'effet de plan de sondage est nettement plus fort (DEFF = 1.74 et non plus 1.37) : la répartition spatiale des étrangers est nettement plus polarisée que celle des lieux de naissance ; selon ce critère, la ségrégation est plus forte. On voit au contraire que les naturalisés sont saupoudrés d'une façon plus homogène entre communes (DEFF=1,08) ; mais leurs fluctuations par strate ne sont pas interprétables, faute d'effectifs suffisants. De plus, la pondération des étrangers a été fortement relevée par le redressement ; cela se paie sous forme de fluctuation lors de la mesure des proportions d'étrangers ou de Français.

Soyons plus brefs pour l'habitat individuel ou collectif à 25 et 50 ans. La proportion en collectif croît de 61 à 69 %. Avec des intervalles de confiance de 3 %, l'écart est significatif (au seuil de 5 %).

L'hétérogénéité entre strates est considérable. Le résultat est rassurant ! 96 % de nos enquêtés **aujourd'hui** franciliens et parisiens (strates 1 et 2) vivaient à 50 ans en habitat individuel. Cette proportion fluctue de 59,6 % à 28 % entre les franges de l'agglomération (strates 6 et 7), les villes nouvelles et les axes et vallées. Les intervalles de confiance sont

larges de (8 % à 20 %), mais en regroupant ces quatre strates assez homogènes entre elles, on totalise un effectif de 345 enquêtés susceptible d'une meilleure précision.

Comme attendu, l'effet de plan de sondage est très élevé : l'habitat à 50 ans est très corrélé à l'actuel. Il est donc très dépendant de la commune de résidence et donc du tirage des unités primaires[1]. Dans les strates de la capitale (1 et 2), l'échantillon est meilleur qu'à l'issue d'un tirage simple : le tirage systématique, très efficace, réalise un échantillonnage régulier des quartiers et donc des types d'habitat.

Le nombre d'enfants s'établit significativement au dessus du fameux seuil de reproduction (2,1 enfants) avec 2,3 ± 0,1. Il s'agit de générations fécondes. Le malthusianisme parisien n'est pas perceptible -si ce n'est au centre-ville, significativement au dessous de ce seuil (1,83 ± 0,24 et 1,77 ± 0,20) ! - L'accroissement des descendances vers la périphérie est manifeste et culmine dans les axes et vallées (3,1±0,4).

Le nombre de logements personnels jusqu'à 25 ans est très peu variable selon les strates et peu sensible à l'effet de plan de sondage. Les strates 1 et 2 réalisent les conditions du tirage simple (DEFF = 0,98 et 1,00).

Résumons l'apport de ce chapitre :

. Grâce à un logiciel spécialisé en mathématiques des sondages, nous avons calculé des intervalles de confiance plus rigoureux que ceux issus du logiciel statistique SAS qui, d'une façon générale, sous-estime la taille de ceux-ci.

. On a mesuré l'effet de plan de sondage. Ainsi, sait-on dans quelle proportion la variance des estimations a été accrue par l'arsenal technique des sondages. C'est la contrepartie statistique du renoncement à l'échantillon simple, beaucoup plus coûteux.

. Le rapport d'exploitation de l'enquête, élaboré dès le redressement, fournit des intervalles de confiance qui tiennent compte de l'effet le plus souvent négatif de la probabilité inégale, mais pas du cumul de la

---

[1] D'une certaine homogénéité interne, mais hétérogènes entre elles, elles réalisent la condition même de l'effet de grappe.

stratification (effet positif faible) et du tirage à deux degrés (effet négatif fort). Prenons l'exemple du nombre de logements occupés avant un âge donné (Tableau 18). Le DEFF réel dépasse celui estimé par SAS d'environ 25 %, quoique de façon assez inégale selon l'âge.

TABLEAU 18
PRECISION DU NOMBRE DE LOGEMENTS ATTEINTS A UN AGE DONNE

AGE	25	30	35	40	45	50	55*	60**
DEFF réel	1,25	1,25	1,35	1,46	1,52	1,53	1,30	1,19
DEFF de SAS	1,25	0,86	1,34	1,12	1,51	0,98	0,96	0,83

(* ** : effectifs faibles ou très faibles)

Le tableau 19 permet la même comparaison pour d'autres variables. Le DEFF de SAS est inférieur à 1 pour le nombre de logements atteints à 30 ans, le nombre d'enfants à l'enquête (50 à 59 ans), le type d'habitat à 45 ans (individuel ou collectif), la surreprésentation de Paris est donc statistiquement efficace (sondage de Neyman); la stratification joue dans le même sens, mais l'effet de grappe des unités primaires est très défavorable et porte la valeur du DEFF réel à 1.89. Cette variance plus que doublée correspond à un accroissement de 42 % de l'intervalle de confiance des estimations par rapport à celles erronées de SAS. C'est un écart extrême, comme en témoignent les autres exemples du tableau 19 choisis parmi les situations défavorables.

TABLEAU 19

	Nombre d'enfants	Nombre de pièces à 50 ans	Habitat individuel à 45 ans	Originaire de l'Ile de France
DEFF réel	1,34	1,39	1,89	1,37
DEFF de SAS	0,93	1,11	0,92	1,00

Les résultats de ce chapitre sont assez cohérents. Ils révèlent le prix des choix opérés et enlèvent quelques illusions sur la précision de ces petits échantillons.

V - LES ERREURS DE MESURE

On ne peut clore ce bilan méthodologique de l'enquête Peuplement et Dépeuplement de Paris sans attirer l'attention sur les erreurs de mesure qui peuvent dépasser l'amplitude des biais et de l'incertitude d'échantillonnage. Par exemple, le recensement de la population lui-même fournit une mesure du chômage national de moins bonne qualité que l'enquête Emploi qui ne porte "que" sur 70 000 ménages. Le questionnaire auto-administré du recensement, insuffisamment précis, sous-enregistre en effet le chômage au sens du BIT sans qu'intervienne de compensation des erreurs.

L'enquête PDP repose sur la mémoire des enquêtés pour de nombreuses informations rétrospectives (dates d'emménagement, surface et nombre de pièces des logements antérieurs, année de départ des enfants, amis...). Or l'enquêté peut ignorer la date d'entrée dans un logement. Il peut également en donner une date erronée. Cet oubli sera patent tandis que l'erreur de date n'apparaîtra que dans certains cas favorables ; par exemple, la date fournie peut s'avérer contradictoire avec d'autres informations. Le plus souvent, les erreurs de mesure ne s'apprécient que partiellement et par une évaluation globale. Nous en avons vu le cas au chapitre 2, lorsqu'on a dissocié les erreurs de sondage et de mesure relatives au statut d'occupation du logement.

Certaines erreurs de datation peuvent se compenser (elles sont alors aléatoires), mais les statisticiens ont repéré deux tendances : l'attirance pour les chiffres ronds (les nombres terminés par 0 ou 5) et l'effet de télescopage des dates (c'est à dire la sous-estimation des durées longues). A l'expérience, un bon questionnaire vise à réduire ces risques ; par exemple, on ancre le calendrier sur quelques dates certaines pour l'enquête, puis les dates plus floues sont définies par référence aux dates connues.

Les erreurs de mesure sont moins bien connues que les erreurs d'échantillonnage ; elles constituent le maillon fragile des enquêtes (comme

des recensements) car il n'en existe pas de théorie générale ; c'est une question d'expérimentation propre à chaque domaine d'études. Celle-ci nécessite de pouvoir confronter plusieurs sources concernant les mêmes individus; ainsi l'INED et l'Université Catholique de Louvain mènent en Belgique une recherche sur les erreurs dans les enquêtes rétrospectives. Cette expérience fournira une information précieuse pour la présente enquête. Les résultats en seront disponibles en 1989[1].

Le registre belge de population fournit en effet l'occasion de ce test : cette source administrative relève l'ensemble des résidences occupées au cours de sa vie par l'individu, ainsi que la liste des entrées et sorties des membres du ménage et celle des changements d'état civil. Toutes ces informations sont datées sur le registre. La confrontation de nos biographies familiales et migratoires à ce registre va permettre d'en apprécier la qualité. Dans ce but, l'INED et l'ULC ont réalisé ensemble une enquête auprès de 450 couples de Wallonie sur la base du questionnaire 3B de l'INED. A l'issue de leur interview simultanée dans deux pièces différentes, les conjoints confrontent leurs récits des événements communs, tranchent et interprètent leurs divergences. L'interrogation des registres, indépendante des interviews, permet alors de tester la fiabilité des enquêtes et des

Nos erreurs de mesure n'en seront que partiellement connues : nous n'apprendrons rien sur la mémorisation des caractéristiques du logement, comme son nombre de pièces ou sa surface. N'oublions pas l'existence de ces incertitudes du seul fait qu'on les perçoit mal.

CONCLUSION

Sans doute, un bilan aussi détaillé de la qualité de l'échantillon est-il inhabituel. Une fois une méthodologie retenue, le choix opéré s'impose comme évident tandis que les solutions écartées disparaissent dans l'oubli. Cette note devrait permettre de mieux comprendre la démarche suivie dans la conception d'un échantillon.

---

[1] Deux communications en seront faites aux congrès de l'Institut International de Statistique (séance 19, Paris, Août-septembre 1989) et de l'union des démographes (séance 1.7, New-Delhi, Septembre 1989).

Les tableaux récapitulatifs de la collecte, les confrontations de structure d'échantillon avant les procédures de redressement et les estimations de variance permettent de lever pour le lecteur le mystère dont on croit enrobée la production des données d'enquêtes. Cette présentation sans fard de l'enquête conduit à une appréciation plus juste de sa précision et de celle des enquêtes en général. Puisse ce document avoir qualité

Nous voulons terminer en rappelant un fait d'évidence : quel que soit le soin de l'analyse, la qualité de l'enquête se joue sur le terrain dans la qualification et la motivation du réseau d'enquêteurs. Toute notre gratitude va donc au service d'enquêtes et aux enquêteurs de la Direction Régionale de Paris de l'INSEE qui ont assumé la difficile tâche de la collecte.

BIBLIOGRAPHIE

[1] G. POURCHER **"Le peuplement de Paris"**. Cahier de l'INED n°43, 1964.

[2] C. BONVALET **"Les parisiens dans leur maturité : origines, parcours, intégration"**. Population n°2, 1987.

[3] C. BONVALET, M. LEFEBVRE **"Le dépeuplement de Paris, 1968-1975"**. Population n°6, 1983.

[4] C. BONVALET **"Histoire résidentielle d'une génération de parisiens nés entre 1926 et 1935"**. Rapport à la CNAF comprenant la première version de ce texte. Juin 1988.

[5] D. COURGEAU **"Migrants et migrations"** Population n°2, 1973.

[6] B. RIANDEY **"Le suivi des échantillons dans les enquêtes démographiques : un bilan"**. Population n°4-5, 1988.

[7] A. FOUCHER **"Annuaire statistique de la région d'Ile de France"**. IAURIF, 1980.

[8] L. LEBART, N. TABARD **"Application des méthodes d'analyse des données à la préparation des enquêtes auprès des ménages (morphologie sociale des communes de la région parisienne)"**. Journal de la Société Statistique de Paris, 1971.

[9] M. GLAUDE cité chapitre 8 dans J.J. DROESBEKE, B. FICHET, Ph. TASSI **"Les sondages"**. Economica, ASU, 1987.

[10] J.C. DEVILLE Séminaire de l'INSEE du 16.06.1988 et compte rendu par F. GUGLIELMETTI **"Quelques développements récents dans la théorie des sondages"**. Le courrier des statistiques n°47, Juillet 1988.

[11] Y. LEMEL **"Une généralisation de la méthode du quotient pour le redressement des enquêtes par sondage"** Annales de l'INSEE n°22-23, 1976.

[12] M. DEROO, A.M. DUSSAIX **"Pratique et analyse des enquêtes par sondage"** PUF, 1980.

[13] H. JACQUART **"Qui ? Quoi ? Comment ? Ou la pratique des sondages"**. Eyrolles, 1988.

[14] B. RIANDEY **"Le redressement et la précision de l'enquête Triple biographie de l'INED"** Note 1985.

[15] F. MARCHAND **"Fluctuation d'échantillonnage dans l'enquête Peuplement et Dépeuplement de Paris ; traitement par SESUDAAN"**. Note septembre 1988.

CARTE 1

LES SECTEURS DE LA REGION D'ILE-DE-FRANCE