

L'ENQUETE PRIORITAIRE 2 AU BURKINA FASO

Plan de sondage - Janvier 1998

1 GENERALITES

Ce document décrit le plan de sondage proposé en vue de l'Enquête Prioritaire 2. L'enquête initiale, EP1, a eu lieu en 1994. La présente enquête est très similaire dans son contenu: elle suppose la même méthodologie et la même taille d'échantillon avec un questionnaire très peu différent. Par contre certains des objectifs de la nouvelle enquête diffèrent d'une manière significative. Il y eu d'ailleurs entretemps un nouveau recensement et le gouvernement a introduit un nouveau découpage régional pour les besoins de la planification. Ces modifications ont eu une incidence importante sur la structure du plan de sondage de la nouvelle enquête. C'est ainsi qu'on s'est proposé de re-penser le plan de l'enquête en commençant à zéro.

Rappelons à cet effet que la comparaison des résultats de deux enquêtes ne dépend nullement sur le fait qu'elles utilisent des échantillons similaires. A condition que les deux enquêtes fournissent leur estimation (extrapolation) relative à un même domaine, on pourra comparer les deux jeux de résultats. La comparaison se fera au niveau du chiffre extrapolé: sa validité ne dépend pas de ce que les deux chiffres proviennent de deux échantillons de la même taille, ni même de deux plans de sondage analogues.

2 OBJECTIFS ET CONTRAINTES

L'objectif prioritaire de la nouvelle enquête est de fournir les données de base d'ordre économique valables au niveau de chacune des 10 nouvelles régions économiques. Il va de soi que de telles données représenteront la totalité du territoire burkinabé et permettront, après extrapolation éventuelle, de produire des estimations des différents paramètres au niveau national. On doit pouvoir en même temps identifier dans l'échantillon les strates géographiques utilisées pour l'enquête EP1 de façon à ce qu'on pourra comparer les résultats des deux enquêtes en fonction des strates (ou régions) plus étendues, et moins nombreuses, utilisées à l'époque. On donnera au chapitre 3 la correspondance entre les différents classements régionaux concernés.

Un deuxième objectif, partagé avec l'enquête précédente, est de permettre des analyses économiques relatives à un certain nombre de groupes socio-économiques bien définis. Le chapitre 4 sera consacré à la définition des ces groupes.

Au chapitre 5 on décrira la structure générale de l'échantillon proposé ainsi que les bases de sondage disponibles.

Dans le chapitre 6 on étudiera la répartition de l'échantillon permettant d'atteindre les objectifs définis plus haut, en fonction de la régionalisation et des GSE.

Au chapitre 7 on fera explicite la méthodologie précise du tirage de l'échantillon avec les paramètres exactes, ainsi que le support théorique qui la justifie.

Enfin au chapitre 7 on explicitera les formules d'extrapolation à utiliser (coefficients de pondération).

Avant de quitter ce chapitre sur les contraintes ajoutons deux paramètres numériques qui se sont fixées à l'avance lors de la définition du projet:

- 1) Il est convenu que la taille de l'échantillon sera la même que dans l'EP1, soit 8.500 ménages au total.
- 2) Il est également convenu que les paramètres du travail du terrain ne changent pas, ce qui implique qu'on enquêtera 20 ménages dans chaque grappe de l'échantillon.

Il suit de ces deux décisions que le nombre de grappes à sélectionner sera de 425.

3 LES REGIONS DANS L'EP1 ET L'EP2

Dans l'EP1 on avait 7 strates, dont 2 urbaines et 5 rurales. Dans l'EP2 on propose un ensemble de 10 strates, appelées régions économiques (RE), dont chacune contient une partie rurale et une partie urbaine.

Le tableau 1 donne la correspondance entre ces deux systèmes de classement géographique. Les lignes du tableau sont les RE, les colonnes les strates de l'EP1. On a inscrit dans les cellules du tableau le nombre de ménages existant dans les fichiers du recensement. Rappelons que les RE sont constituées des groupes de provinces. Quant aux strates de l'EP1 les strates rurales sont des groupements des zones rurales de provinces.

Dans ce tableau la dichotomie urbain/rural est celle utilisée pour le recensement. Or dans l'EP1 le nombre de villes classées dans la strate "autres villes" était de 16, alors que dans le RGPH il y en a 26. Cela représente un décalage significatif. Par conséquence, des différences éventuelles observées entre les deux enquêtes peuvent s'expliquer, dans ce secteur, non seulement par l'évolution au sein de chaque localité mais aussi par le fait que la population des "autres villes" couvre aujourd'hui un plus grand nombre de villes.

Tableau 1. DONNEES DU R.G.P.H. -96 VENTILEES SELON LE STRATES DE L'E.P.-94
Nombre de ménages recensés

REGIONS ECONOMIQUES	STRATES URBAINES			STRATES RURALES						TOTAL GENERAL
	Ouaga/Bobo	Autres villes	Total urbain	Ouest	Sud,Sud-E	Centre-S	Centre-Nord	Nord	Total rural	
Ouest	57377	14035	71412	156679	0	0	0	0	156679	228091
Nord-O	0	14025	14025	131956	0	0	49619	0	181575	195600
Sahel	0	4937	4937	0	21764	0	0	109396	131160	136097
Est	0	6368	6368	0	130899	0	0	0	130899	137267
Sud-O	0	2908	2908	0	79493	0	0	0	79493	82401
Centre-N	0	8608	8608	0	0	0	129221	0	129221	137829
Centre-O	0	20248	20248	0	33583	79880	0	0	113463	133711
Centre	146453	3254	149707	0	18053	177165	0	0	195218	344925
Nord	0	13665	13665	0	0	0	128443	0	128443	142108
Centre-E	0	14643	14643	0	0	120319	0	0	120319	134962
Total	203830	102691	306521	288635	283792	377364	307283	109396	1366470	1672991

4 GROUPES SOCIO-ECONOMIQUES

L'analyse des résultats de l'EP1 (voir le rapport de l'enquête¹) a mis l'accent sur la ventilation des différentes caractéristiques observées en fonction d'un certain nombre de groupes clefs appelées "les groupes socio-économiques" (GSE). Ces groupes ont été définis au niveau des ménages et selon l'activité économique du chef de ménage. Le rapport distingue les GSE suivants. Nous avons ajouté à titre d'information la répartition en pourcentage des ménages selon l'EP1.

	<u>Pour cent</u>
1. Salariés du secteur public	5,9
2. Salariés du secteur privé	3,9
3. Artisans, commerçants	8,6
4. Autres actifs	1,2
5. Agriculteurs - rentes	8,5
6. Agriculteurs - vivriers	64,4
7. Inactifs	<u>7,5</u>
	100,0

Il a été proposé que l'EP(II) suive le même système. Il s'agit donc d'assurer par le plan de sondage que chacun de ces groupes soit représenté avec un nombre suffisant de ménages dans l'échantillon.

Il convient d'étudier d'un peu plus près les définitions de ces groupes et leur utilité quant au classement aux fins de l'analyse économique.

1. Salariés publics. Essentiellement les fonctionnaires. Les organismes parapublics figurent dans la catégorie suivante.

2. Salariés privés. L'ensemble des autres salariés. A noter qu'on y trouvera les secteurs formels et informels ensemble. La distinction entre ces deux grandes catégories serait en principe intéressante mais les effectifs sont trop faibles à l'heure actuelle pour permettre une analyse de ce genre dans le cadre d'une enquête de cette taille.

3. Artisans, commerçants. On visait l'ensemble des indépendants, à la seule exclusion des agriculteurs.

4. Autres actifs. Il s'agit des 4 groupes: employeurs, aides familiaux, apprentis/stagiaires, et bénévoles. Dans la pratique c'est un petit groupe très hétérogène qui ne semble pas porter un grand intérêt pour l'analyste.

5. Agriculteurs - rente. Il s'agissait au départ de tous ceux qui cultivent une au moins des cultures classées comme étant "de rente". Mais lorsque le projet d'analyse avançait il s'est aperçu que seul le coton pouvait être compté avec confiance dans cette catégorie. C'est ainsi que si on veut interpréter les tableaux du rapport de l'enquête on doit traduire cette catégorie comme: "Agriculteurs - cultivateurs du coton"

¹ Analyse des résultats de l'Enquête Prioritaire sur les Conditions de Vie des Ménages. Direction des Statistiques Générales, Ouagadougou, 1996.

6. Agriculteurs - vivriers. L'ensemble des autres agriculteurs non classés ailleurs. Donc: "Agriculteurs - non cultivateurs du coton".

7. Autres inactifs et chômeurs. Le fait que le système de classement est basé sur l'activité du chef de ménage rend cette catégorie trompeuse. Peu des chômeurs figurent ici. On aura beaucoup de retraités de la fonction public ou des forces armées, plus un certain nombre de femmes solitaires. Comme le GSE 4, ce groupe est hétérogène et d'un intérêt limité pour l'analyste.

5 STRUCTURE DE L'ECHANTILLON ET BASES DE SONDAGE

La structure générale de l'échantillon sera la même que celle de l'EP1. On tirera un échantillon à deux degrés: au 1er degré l'unité de sondage sera la zone de dénombrement (ZD) du recensement et au 2ème le ménage. Le 1er degré sera tiré avec des probabilités proportionnelles à la taille estimée de chaque unité. Au 2ème degré on sélectionnera 20 ménages dans chaque ZD, par le tirage systématique avec probabilité fixe dans la ZD.

Dans une strate où le pas de sondage au 1er degré ne change pas, ce plan donnera un échantillon approximativement autopondéré. Ce n'est qu'une approximation parce que la taille supposée pour chaque ZD en faisant le tirage au 1er degré n'est pas exactement égale au nombre d'unités secondaires (ménages) qui constitue la base de sondage secondaire. Afin de rectifier ce décalage on devra introduire des poids dans l'analyse.

De plus, les contraintes que nous avons déjà citées nous obligent d'introduire certaines variations importantes des taux entre des domaines géographiques. On verra le détail de ces variations au chapitre 6.

Du fait de ces différentes sources de variation dans les taux, on devra appliquer des coefficients de pondération spécifiques à chaque ZD lors de l'extrapolation des données.

Passons alors à la question des bases de sondage. Au niveau du 1er degré on dispose de l'ensemble des ZD du recensement. Ces unités sont d'une taille assez homogène et chacune est accompagnée d'un plan ou croquis qui permet de situer ses limites sur le terrain. Les mesures de taille sont disponibles dans les données du recensement. En principe on pourrait choisir à cette fin entre le nombre de ménages recensés dans chaque ZD et l' effectif de population. Un facteur donne la préférence à la deuxième de ces variables: l'effectif de population. Il s'agit de l'opération mise-à-jour du recensement qui a été exécutée fin 1997 dans les zones non loties des deux grandes villes. Cette opération a permis la révision de l'ensemble des ZD dans ces zones: on a créé les nouvelles ZD avec des nouvelles estimations de leurs populations. Cependant on n'a pas rénuméré les ménages. Il est fort souhaitable que ces nouveaux résultats soient intégrés au recensement avant d'utiliser ce dernier comme base de sondage pour les enquêtes. Comme les effectifs de population sont disponibles pour les nouvelles zones mais que les effectifs de ménages ne le sont pas, nous donnons la préférence à la population comme mesure de taille à utiliser partout.

Quant au second degré de sondage, on ne dispose pas d'une base de sondage valable: aucune liste des ménages n'est suffisamment exacte et à jour pour servir à cette fin. On devra donc, comme dans la quasi-totalité des enquêtes africaines, procéder à un dénombrement des ménages dans chaque ZD de l'échantillon afin d'obtenir une telle liste. Cette opération sera organisée dans chaque ZD juste avant l'enquête proprement dite dans la zone, en utilisant la même équipe d'enquêteurs.

En résumé:

La base de sondage au 1er degré sera les ZD du RGPH-96, avec mise-à-jour des zones non loties dans les deux grandes villes, les mesures de taille étant l'effectif de population signalé par le RGPH ou sa mise-à-jour.

La base de sondage au 2ème degré sera une liste des ménages préparée lors d'une opération de ratissage dans chaque ZD échantillon, effectuée juste avant l'enquête dans la zone.

6 REPARTITION DE L'ECHANTILLON DANS L'ESPACE ET ENTRE GSE

Au chapitre 2 de ce rapport nous avons fixé les deux paramètres de base de l'échantillon:

Ménages à tirer au total	8.500
Ménages à tirer dans chaque ZD	20

Les différentes contraintes sur l'échantillon mentionnées aux chapitres 2, 3 et 4 vont déterminer la répartition de l'échantillon au sein de cette enveloppe. Nous avons notamment cité la nécessité d'avoir un échantillon d'une taille suffisante au sein de chacune des 10 régions économiques et de chacun d'un certain nombre de groupes socio-économiques.

Comment doit-on interpréter "échantillon d'une taille suffisante"? Il est évident que plus l'échantillon est grand plus on peut approfondir les analyses. En toute rigueur scientifique il ne peut y avoir un seuil précis au delà duquel un échantillon est "suffisant". Malgré cela le sondeur est tenu à chercher des contraintes explicites, ne serait-ce que pour des raisons d'ordre financier. Dans cette situation assez mal définie il est raisonnable de consulter l'expérience des analystes face aux fichiers des enquêtes réelles effectuées dans d'autres pays ou précédemment dans le pays concerné. Or l'expérience des nombreuses enquêtes du genre EP et EI (enquête intégrée) dans des pays africains semble montrer qu'il faudrait environ 500 ménages dans chaque domaine où on désire faire des analyses utiles. Faute d'une meilleure règle on adoptera ce chiffre comme si c'était un seuil précis afin de déterminer les paramètres de la présente enquête.

La stratégie que nous suivrons sera donc d'ajuster les taux de sondage afin d'assurer un minimum de 500 ménages dans chacun des domaines ciblés, soit les 10 régions économiques et certains des GSE. Examinons cela de plus près.

Abordons d'abord le problème des régions. Notre première idée a été d'adopter dans chaque région économique (RE) le taux de sondage qui donne un échantillon de 500 ménages exactement dans la RE. Cela impliquerait 10 taux différents, avec 10 coefficients de pondération dans l'analyse pour les redresser - un dispositif assez compliqué. En examinant la situation de plus près cependant on s'est aperçu que la partie rurale des RE ne varient pas beaucoup de taille (nombre de ménages). Dans cette situation une solution plus simple semblait s'offrir: ne pourrait-on pas adopter un taux homogène pour l'ensemble du secteur rural, choisi de façon à ce que l'échantillon dans la RE la plus petite s'élève à 500, tous les autres étant donc supérieurs à ce seuil? Malheureusement, une des RE posait un problème: la Sud-ouest. Celle-ci était bien plus petite que toutes les autres. En lui accordant 500, et en respectant le principe d'un taux unique pour l'ensemble du secteur rural, on serait amené à hausser l'échantillon dans toutes les autres RE jusqu'à un niveau qui ne laisserait plus l'espace pour le secteur urbain.

Si, par contre, on admettait un échantillon de 350 seulement pour la Sud-ouest avec un minimum de 500 pour les neuf autres RE (taux de sondage de 0,004408), on arriverait à un échantillon rural total de 6024, ce qui laisserait 2.476 pour l'urbain.

Avant de prendre une décision définitive à cet égard, étudions la question de la représentation des GSE.

Le tableau 2 (page 8) donne la répartition des 7 GSE entre les strates de l'EP1, selon les données de l'EP1 extrapolées au niveau national.

Si on examine les résultats du tableau 2 au niveau des ensembles urbain et rural on voit que les trois premiers GSE sont fortement concentrés dans l'urbain. Si on appliquait au secteur urbain le taux de sondage qu'on vient de mentionner (0,004408) on n'aurait que 432 ménages salariés du secteur public et 283 du secteur privé. On doit conclure qu'il faudra sur-échantillonner le secteur urbain de l'ordre de 2 fois par rapport au rural.

En ce qui concerne le 4ème GSE, autres actifs, comme nous l'avons vu au chapitre 4 celui-ci est essentiellement un reliquat: il ne peut être considéré comme un groupe cible méritant l'analyse économique.

Les GSE 5 et 6, les agriculteurs, constituent en effet des groupes cibles. Cependant leurs effectifs sont suffisamment nombreux pour qu'on ne sera en aucun cas appelé à ajuster les taux pour assurer leur bonne représentation. Autrement dit, ils seront suffisamment nombreux dans n'importe quel genre d'échantillon d'une taille de plusieurs milliers.

Enfin le GSE 7, inactifs, présente le même genre de problème que le GSE 4 : c'est un reliquat plutôt qu'un groupe à cibler pour analyse.

Ce n'est donc que les 3 premiers GSE qui devront nous préoccuper en ce qui concerne la répartition de l'échantillon et ceux-ci nous amènent à vouloir doubler le taux de sondage urbain par rapport au rural. Nous aboutissons ainsi à un régime très simple avec un taux unique de sondage pour le rural and un autre taux deux fois plus grand pour l'urbain, régime qui satisfait à toutes les contraintes à la seule exception que le rural de la région économique du sud-ouest aura un échantillon nettement inférieur au seuil de 500 souhaité. Après discussion au sein de l'équipe de l'INSD il a été décidé qu'il ne fallait pas faire exception d'une seule région: le seuil de 500 doit être appliqué partout. On introduira donc un troisième taux égal à 1,5 fois le taux rural, soit 1,5 T. Ce taux sera appliqué à une seule région: le secteur rural de la RE 5 (sud-ouest). Le restant du rural sera affecté du taux T et l'ensemble de l'urbain (y compris la partie urbaine de la RE 5) du taux 2T.

Tableau 2. REPARTITION DES MENAGES SELON LA STRATE ET LE GSE
Extrapolation des données de l'EP1

GROUPES	STRATES URBAINES		STRATES RURALES					TOTAL		
	Ouaga/ Bobo	Autres villes	Ouest	Sud,sud-Est	Centre-Nord	Centre-Sud	Nord	Urbain	Rural	Total
Salariés du sect.public	31642	13917	7119	8279	3686	5033	1212	45559	25331	70890
Salariés du sect. privé	35618	3026	2776	972	1502	2153	392	38644	7795	46439
Artisans, commerçants	58982	9861	11612	6163	4015	8265	6788	68843	35843	104686
Autres actifs	4155	817	5811	200	1042	1923	658	4972	9640	14612
Agriculteurs de rentes	161	285	62581	14296	14269	10816	1009	446	102970	103416
Agriculteurs vivriers	17246	25784	114157	126226	199534	236598	59941	43030	736457	779487
Inactifs, chômeurs	26269	6733	13899	8333	15203	13815	5322	33002	56571	89573
TOTAL	17407073	60423	217955	163470	239257	278602	75323	234496	974607	1209103

En recalculant les paramètres exactes on obtient les valeurs suivantes:

	<u>Urbain</u>	<u>Rural sauf S-O</u>	<u>Rural S-O</u>	<u>Total</u>
Ménages recensés	306521	1286977	79493	1672991
Taux de sondage	0.008418932	0.004209466	0.006314199	
Echantillon	2581	5417	502	8500

Il convient de reproduire ce calcul de la taille de l'échantillon au niveau de chaque RE. Ceci s'est fait au tableau 3 plus bas (partie gauche du tableau). Enfin rappelons qu'on doit enquêter 20 ménages au sein de chaque ZD de l'échantillon. Dans la partie droite du tableau 3 on divise par 20 les données de gauche en arrondissant à l'entier, pour donner le nombre de ZD à tirer dans chaque RE du côté urbain et du côté rural.

Tableau 3. REPARTITION DE L'ECHANTILLON, NOMBRE DE MENAGES ET DE ZONES: MILIEUX URBAIN ET RURAL SEPARÉMENT

REGIONS ECONOMIQUES	NOMBRE DE MENAGES			NOMBRE DE ZD		
	U	R	T	U	R	T
Ouest	601	660	1261	30	33	63
Nord-ouest	118	764	882	6	38	44
Sud	42	552	594	2	28	30
Est	54	551	605	3	28	31
Sud-ouest	24	502	526	1	25	26
Centre-nord	72	544	616	4	27	31
Centre-ouest	170	478	648	8	24	32
Centre	1260	822	2082	63	41	104
Nord	115	541	656	6	27	33
Centre-est	123	506	629	6	25	31
TOTAL	2579	5920	8500	129	296	425

Enfin, avec ce régime très simple des taux, comment se passe-t-il avec la représentation des trois premiers GSE?

Le tableau 2 permet d'estimer, à partir de l'EPI, la proportion de ménages appartenant à un GSE donné, pour chacune des strates de l'EP. En appliquant ces proportions à la répartition de l'échantillon observée au tableau 3 nous pouvons estimer le nombre de ménages de chaque GSE qui devraient paraître dans notre échantillon. Les résultats pour les trois premiers GSE figurent au tableau 4. On peut résumer ces prévisions de la manière suivante.

Salariés du secteur publique: Le seuil de 500 ménages est dépassé si on groupe la totalité de secteur urbain (511 ménages). On pourrait étudier donc ce groupe, ou bien la totalité de la population burkinabée, mais on ne pourra pas étudier comme groupe distinct la population des grandes villes ni la population rurale.

Tableau 4. COMPOSITION DE L'ECHANTILLON EN FONCTION DE CERTAINS GSE

GROUPES SOCIO- ECONOMIQUES	STRATES URBAINES		STRATES RURALES	TOTAL
	Ouaga/Bobo	Autres villes		
Salariés du secteur pubc	312	199	164	675
Salariés du secteur privé	276	177	47	500
Artisans, Commerçants	581	46	181	808

Salariés du secteur privé: Le seuil est atteint si on étudie la population totale (500 ménages) mais aucun découpage géographique au sein de ceci ne sera analysable.

Artisans, commerçants: Le seuil est dépassé si on étudie la population des grandes villes (581 ménages). On pourrait également étudier l'ensemble du secteur urbain (6276 cas) ou, bien entendu, l'ensemble de la population nationale (1071 ménages). Mais le secteur "autres villes" n'est pas suffisamment grand, ni le secteur rural tout seul.

7 TIRAGE DE L'ECHANTILLON

La méthodologie s'inspire d'un plan classique de sondage appelé "sondage autopondéré avec probabilité proportionnelle à la taille". Il convient de décrire ce modèle idéal avant de passer à l'application.

On suppose un échantillon tiré à deux degrés et sans stratification. Au 1er degré on tire un échantillon d'unités primaires (UP) en donnant à chaque UP une probabilité de tirage proportionnelle à sa taille M_i , où l'indice i désigne l'UP i .

Au 2ème degré on tire un nombre b_i de ménages sur un ensemble de M_i ménages trouvés dans cette UP.

On a donc au 1er degré:

$$\text{probabilité de sélection } p_{1i} = k M_i$$

et au 2ème degré:

$$\text{probabilité de sélection } p_{2i} = b_i / M_i' .$$

La probabilité globale de sélection pour un ménage est le produit des probabilités aux deux degrés, car pour sélectionner un ménage il faudra d'abord sélectionner son UP et ensuite ce ménage parmi l'ensemble des ménages dans l'UP. On a donc:

$$\text{Probabilité globale } P_i = p_{1i} p_{2i} = k b_i M_i / M_i'$$

Le cas d'intérêt particulier est celui où la mesure de taille de l'UP i est égale au nombre de ménages existant dans cette UP, c'est-à-dire que $M_i = M_i'$. Dans cette situation le sondeur pourra fixer la valeur b_i pour être constant partout, soit b . Autrement dit, il tirera le même nombre de ménages dans chaque UP. La probabilité globale deviendra alors égale partout au constant $k b$. Donc, tous les ménages ont la même probabilité d'être tirés. Cet arrangement présente plusieurs avantages.

- 1) Tirer un nombre fixe de ménages partout simplifie considérablement les travaux de l'enquête.
- 2) Donner à chaque ménage la même probabilité d'être tiré implique un échantillon "autopondéré", c'est-à-dire échantillon qui constitue une image directe de la population sans qu'on ait besoin d'appliquer des pondérations. Cela simplifie l'analyse. En général, d'ailleurs l'échantillonnage sera ainsi d'une efficacité maximale (erreur minimale pour un coût donné).

Malheureusement dans la pratique on ne pourra pas souvent appliquer ce plan parce qu'on ne dispose pas, au moment du tirage primaire, des données sur le nombre de ménages qu'on va trouver dans l'UP. La mobilité des populations africaines, en particulier, empêche que les données du recensement restent valables pendant plus d'un an, même si on suppose qu'elles étaient exactes au départ.

Il est donc quasiment universel dans les enquêtes africaines de conserver les deux paramètres M_i et M_i' distinctes. La quantité M_i' qui vient du dénombrement des ménages, constitue une sorte de mise-à-jour de la M_i . Dans les enquêtes sur les ménages on conservera souvent la constance du b qui présente un avantage notable. L'échantillon reste d'ailleurs approximativement optimal car il s'approche de l'autopondéré.

Passons donc au cas de la présente enquête.

La mesure de taille des ZD sera, nous l'avons vu, la population recensée, soit N_i dans chaque ZD i . L'effectif de ménages recensés, M_i , aurait été une mesure plus précise mais l'incidence d'une erreur de ce genre sur l'efficacité du sondage est négligeable. Qu'on choisisse M ou

N ce facteur figurera dans la formule pour la probabilité, et sera donc annulé lors de la pondération qui précède l'extrapolation: dans l'un ou l'autre cas l'estimation sera sans biais.

Le tirage au 1er degré, effectué avec probabilité proportionnelle à la taille, se fera par la méthode systématique, c'est-à-dire par tirage à un intervalle fixe à partir d'un départ aléatoire. Avant d'entamer l'échantillonnage on inscrira la taille à côté de chaque ZD dans la liste. Ensuite on fera le cumul des tailles dans une colonne à côté. Enfin on tirera à un intervalle fixe dans la colonne des cumuls. L'intervalle, ou le "pas", de tirage est obtenu en divisant le total des tailles (ce qui égale le dernier cumul) par le nombre d'UP qu'on désire tirer. C'est l'inverse du k qui figure dans la première formule citée plus haut.

Nous avons donc pour la probabilité globale de sondage d'un ménage au sein d'une strate donnée dans l'enquête actuelle:

$$P_i = (a N_i / SN_i) \cdot (b / M_i')$$

où a = le nombre de ZD à tirer

N_i = la population recensée dans la ZD i

SN_i = la population recensée au total dans la strate

b = le nombre de ménages enquêtés dans chaque ZD

M_i' = le nombre de ménages dénombrés dans la ZD i lors du dénombrement des ménages pour l'enquête.

Cette formule doit être appliquée séparément pour chaque strate. Les quantités a proviennent du côté droit du tableau 3, colonnes U et R. Il y a 20 chiffres dans ces deux colonnes donc 20 strates en tout². Les N viennent de la base de sondage. Le b est égale à 20 en principe mais on va tirer 24 ménages dans chaque ZD afin de disposer de 4 ménages de réserve. Enfin les M_i' viendront du dénombrement des ménages effectué avant l'enquête dans chaque ZD.

Pour le tirage au 1er degré on devra d'abord classer le fichier dans les 20 strates. Au sein de chaque strate on classera les ZD par ordre des départements ou communes. On fera le cumul de la population des ZD, en revenant au cumul zéro au début de chaque strate. En divisant le dernier cumul de la strate par la quantité a on obtiendra le pas de tirage. On commencera le tirage avec un nombre aléatoire inférieur ou égale au pas.

² Pour les besoins du calcul des erreurs de sondage chaque strate doit contenir un minimum de 2 unités sélectionnées. Lorsqu'on veut faire des tels calculs on devra alors regrouper la strate urbain du Sud-ouest avec la strate urbaine du Sud.

Enfin pour tirer les ménages, on obtiendra d'abord comme base de sondage la liste des ménages dénombrés ("énumérés"). On numérote ces ménages. On divise le nombre de ménages dénombrés par 24 pour obtenir le "pas", à un décimal près. On fait le tirage en appliquant un tirage systématique à intervalle fixe. Une fois l'échantillon de 24 tiré, on fera un sous-échantillonnage pour distinguer les 4 réserves, à intervalle fixe de 6 en comptant uniquement parmi les 24 ménages tirés. Par exemple, si on numérote de 01 à 24 les ménages tirés, on doit affecter aux réserves un des 6 échantillons suivants:

A:	1	7	13	19
B:	2	8	14	20
C:	3	9	15	21
D:	4	10	16	22
E:	5	11	17	23
F:	6	12	18	24

Le contrôleur arrangera à ce qu'on utilise ces séries en rotation successivement en passant d'une ZD à la suivante. Il doit placer l'indication R à côté de chaque ménage affecté aux réserves.

8 EXTRAPOLATION

La formule d'extrapolation n'est que l'inverse de la probabilité globale de sondage. On doit donc multiplier (pondérer) les données de la ZD i par le poids

$$W_i = [SN_i / a b_i] \cdot [M_i' / N_i]$$

A noter que nous admettons b_i dans la formule plutôt que 20. Normalement b_i sera égale à 20 mais au cas où il diffère de cette norme sans qu'on ait ajusté l'erreur, quelle qu'en soit la raison, on l'ajustera en appliquant la formule comme elle figure ici, avec le b_i réel tel qu'il a été.