

Answer on Rural Zones to John Doe

Plan de sondage
avril 1991

L'unité idéale de sondage aréolaire aurait les caractéristiques suivantes :

- Les unités dans leur ensemble couvriraient le territoire d'étude d'une manière exhaustive et sans chevauchement.
- Les unités seraient bien définies : cartographie détaillée et à jour.
- Les unités seraient d'une taille modeste afin de limiter les travaux du dénombrement des ménages au sein de chaque unité ; mais en même temps suffisamment grandes pour qu'on puisse y trouver toujours le nombre voulu de ménages à enquêter (soit 20 dans le cas de la présente enquête).
- Les unités seraient d'une taille assez homogène. Dans le cas contraire une estimation de la taille de chaque unité devrait être disponible avant le tirage de l'échantillon.
- La liste qui constitue la base de sondage comporterait des données sur chaque unité permettant la stratification de la base en fonction des caractéristiques significatives.

La liste des districts de recensement (DR) constitue une approche assez satisfaisante de cet idéal. Les DR sont assez uniformes de taille avec une moyenne autour de 900 personnes, soit 100 ménages. Le Recensement Général de la Population et de l'Habitat (RGPH) effectué en mai 1988 a donné la taille de chaque DR et des cartes datant de cette époque sont disponibles pour chaque DR. Quant à la stratification, les données sont limitées mais on connaît au moins si chaque DR est urbain ou rural ainsi que sa situation sur la carte du Sénégal.

Pour le statisticien sénégalais concerné par les enquêtes un nouvel élément significatif vient de se présenter grâce aux efforts de la DPS et avec l'assistance technique du Bureau of the Census des Etats-Unis. Il s'agit d'un échantillon maître aréolaire, basé sur les données du RGPH. Cet échantillon a été conçu pour permettre une meilleure intégration des enquêtes menées par les services statistiques du Sénégal, tout en facilitant la tâche des sondeurs appelés à planifier les enquêtes futures. Voici un bref résumé des caractéristiques techniques de cet outil.

L'échantillon maître est constitué de 600 unités primaires appelées secteurs.¹ Les secteurs ont été définis à

¹Dans un échantillon où une seule unité secondaire est tirée dans chaque unité primaire, l'appellation "unité primaire" devient ambiguë. Ici il serait légitime de considérer les communes et les communautés rurales comme étant les unités primaires, en admettant toutefois que certaines de ces unités peuvent être tirées avec certitude (probabilité = 1). Nous le

partir du recensement en regroupant un petit nombre de DR limitrophes, mais un DR supérieur à 200 ménages a été considéré comme étant en lui-même un secteur. L'objet de ce regroupement était de constituer des unités de 200 à 300 ménages, donc en moyenne 2 fois plus grand que le DR, mais d'une taille moins variable.

Les secteurs ont été tirés avec des probabilités proportionnelles au nombre de ménages recensés, mais en appliquant une constante de proportionnalité à peu près 50% plus élevée en milieu urbain qu'en milieu rural. En effet, on a tiré 204 secteurs en milieu rural et 396 en milieu urbain (dont 105 dans la ville de Dakar, 113 dans les autres communes du Cap Vert, et 178 dans les communes ailleurs). L'échantillon en milieu urbain a été stratifié au niveau des secteurs selon une méthode assez complexe qui tient compte de trois indicateurs du niveau de vie des ménages habitant le secteur, basés sur les données du recensement. Quatre strates ont été ainsi créées (bien que dans certaines villes deux ou trois strates ont été groupées ensemble). En milieu rural les secteurs ont été stratifiés selon la communauté rurale.

L'existence de l'EM et l'investissement considérable que représente son établissement constituent des raisons notables pour que cet outil soit employé dans la construction des échantillons destinés aux enquêtes futures proposées par le gouvernement sénégalais. Or nous avons vu que l'unité primaire de l'EM est le secteur, alors que l'unité préférée pour l'ESP serait le DR. Y-a-t-il un conflit entre ces exigences? La réponse est négative. En effet l'utilisation de l'EM n'exclut nullement la construction d'un échantillon de DR, avec un seul DR tiré au sein de chaque secteur de l'EM (ou bien au sein de chacun d'un sous-échantillon de ces secteurs). Si le tirage de ce DR dans le secteur est effectué avec une probabilité conditionnelle qui est proportionnelle à la taille du DR, un tel échantillon est rigoureusement équivalent à un échantillon de DR tiré à un seul degré, le secteur ne servant que d'adresse pour identifier le DR. Les auteurs de l'EM ont en effet envisagé l'adoption d'un échantillon de ce genre par certains des clients de l'EM.

Le principal avantage du DR comme unité finale aréolaire par rapport au secteur est que le DR est moins grand et que le dénombrement des ménages exigera donc moins de travail. Mais malgré cet avantage certain on peut toutefois se demander si on n'aurait pas intérêt à introduire un degré de sondage supplémentaire soit avant, soit après le DR. En effet, l'introduction d'une unité supérieure au DR aurait deux conséquences significatives : on réduirait quelque peu la distance moyenne entre les DR voisins dans l'échantillon et on augmenterait légèrement l'erreur de sondage. Or en

trouvons cependant plus simple de concevoir le secteur dans ce rôle, cette unité étant la dernière qui constitue indiscutablement un degré effectif de sondage dans l'échantillon maître.

sélectionnant 500 DR répartis sur le Sénégal la distance entre les DR échantillons est déjà très réduite et le gain en la réduisant davantage serait négligeable. Il semble donc préférable de maintenir le DR comme unité primaire de sondage. Quant à l'introduction d'une unité inférieure au DR, il s'agirait du découpage de chaque DR en segments dont on tirerait un ou deux au hasard. L'effet principal serait une diminution de la taille de l'opération du dénombrement. Cela présente un avantage certain mais il serait contrebalancé par le travail supplémentaire qu'exige la définition des segments: nouvelles cartes à préparer et reproduire, nouvelle technique à apprendre, nouveau manuel à rédiger ... Nous estimons que la solution la plus simple serait de maintenir le principe du DR comme l'unité unique de sondage aréolaire en admettant toutefois que quelques DR d'une taille exceptionnellement grande pourront être subdivisés en deux ou trois segments avant le tirage, alors que des rares DR excessivement petits pourront être regroupés avec des DR voisins. L'unité de sondage serait alors normalement le DR mais exceptionnellement le segment, ou le groupe de deux DR, et on maintiendrait ainsi rigoureusement le principe d'un seul degré de sondage aréolaire partout.

Enfin quelle base de sondage peut-on proposer pour l'échantillonnage des ménages au sein de chaque DR? L'expérience de multiples tentatives africaines montre que le recensement ne peut jouer ce rôle d'une façon efficace : d'une part les populations africaines sont généralement fort mobiles, d'autre part les adresses font défaut, ce qui rend quasiment impossible la recherche des ménages choisis dans les documents du recensement. La seule solution est celui d'organiser un dénombrement des ménages dans les DR de l'échantillon visant spécifiquement les besoins de l'enquête. Notons à propos que jusqu'ici les travaux d'établissement de l'EM n'ont pas porté sur une mise-à-jour de la liste des ménages dans les zones de l'échantillon.

3 REPRESENTATION DES GROUPES SOCIO-ECONOMIQUES

3.1 Définition des groupes socio-économiques : principes généraux

Le choix des GSE résulte d'une négociation entre analyste et sondeur. Rappelons d'abord qu'il s'agit de trouver une répartition de l'échantillon qui assurera un minimum de 400 ménages pour chaque GSE, ce chiffre étant considéré comme le minimum qui permet à l'analyste d'étudier avec une profondeur suffisante l'impact de la politique de restructuration sur un GSE donné.

Le dialogue sera lancé par le sondeur, qui rappellera la contrainte fondamentale, la taille globale de l'échantillon. Dans notre cas il s'agit d'environ 10.000 ménages. Il ajoutera à titre indicatif que, si on se base sur l'expérience

des autres enquêtes africaines, il y a peu de chances qu'on puisse trouver la place pour plus d'une quinzaine de GSE dans un échantillon de cette taille.

L'analyste demandera peut-être ici pourquoi, si chaque GSE n'exige qu'un échantillon de 400 et si on nous autorise d'aller jusqu'à un total de 10.000, ne peut-on pas admettre 25 GSE? La réponse, donnée par le sondeur, est la suivante. "Si les GSE pouvaient être identifiés dans la base de sondage au moment du tirage de l'échantillon votre raisonnement serait valable. Mais en fait les GSE que vous proposerez seront définis chacun en fonction de deux critères, dont l'un est géographique et l'autre socio-économique. Or au moment du tirage de l'échantillon le sondeur connaîtra la situation géographique de chaque unité mais pas son classement socio-économique. Il devra donc fixer le taux de sondage en fonction du seul critère géographique : chaque strate géographique sera dotée d'un taux unique de sondage qui sera appliqué quelle que soit la classe socio-économique considérée au sein de la strate. Le sondeur cherchera alors la classe la plus rare dans la strate, calculera le taux de sondage qui donnera 400 ménages pour celle-ci et appliquera ce taux à l'ensemble des ménages de la strate. Par conséquent il tirera bien plus de 400 ménages pour les classes autres que la plus rare dans la strate. C'est ainsi que, en exigeant un minimum de 400 ménages pour tous les GSE on peut être amené à tirer une moyenne par GSE beaucoup plus élevée.

3.2 Choix des GSE : Essai préliminaire

Pour se fixer une base de travail l'analyste choisira d'abord un découpage géographique et un classement socio-économique. A ce stade préliminaire on propose un schéma très général.

On sépare d'abord les milieux urbain et rural.

En milieu urbain on distingue la ville capitale, Dakar, des autres villes et parmi ces dernières on peut distinguer les grandes des petites - d'où trois strates en tout.

Toujours en milieu urbain on souhaiterait croiser cette stratification purement géographique avec un classement socio-économique. Les économistes proposent un classement du genre : secteur public, secteur privé formel, secteur privé informel, et autres.

En milieu rural plusieurs modalités existent pour établir une stratification géographique : zones agro-écologiques, régions administratives, un mélange des deux ... On y reviendra.

Comme en milieu urbain on souhaiterait ici encore croiser la stratification géographique avec un classement d'ordre

socio-économique. A cette fin les économistes ont proposé qu'on considère l'activité principal du ménage et notamment la nature de la culture pratiquée, dans le but de distinguer les cultivateurs qui travaillent dans le cadre du secteur exportation de ceux qui se situent dans le secteur de subsistance. C'est ici qu'on s'est heurté à un problème. Différemment de beaucoup d'africains les agriculteurs sénégalais ne semblent pas se spécialiser de cette manière. Après que cette conclusion nous a été affirmée de plusieurs côtés nous avons pu étudier directement les documents d'une enquête effectuée par la SODEVA en 1989 sur les 5 régions centrales du Sénégal². Sur un échantillon de 701 exploitants agricoles dans cette zone, qui représente 60% des ménages ruraux du Sénégal, on a trouvé 7% qui ne cultivaient pas l'arachide et 5% qui ne cultivaient pas le mil. En effet, presque tout le monde cultivaient les deux. Nous avons conclu qu'il serait inutile de chercher à distinguer deux populations de cultivateurs de cette manière. Bien qu'on n'ait pu trouver des données statistique analogues sur des autres régions, il nous semblait évident qu'un classement géographique serait le seul faisable. On pourra chercher à raffiner un tel classement en espérant ainsi de distinguer l'activité des cultivateurs selon la région qu'ils habitent, mais il n'y a actuellement aucun moyen de créer des catégories professionnelles parmi eux et de chiffrer l'importance de ces catégories à l'avance de notre enquête.

Notons que cette conclusion différencie le Sénégal des autres pays africains qui ont entrepris cette enquête. Mais il est certain qu'elle a l'avantage de simplifier notre travail, et même de contourner entièrement le problème décrit à la fin de la section 3.1 plus haut dans le cas du milieu rural. Dans la mesure où nos GSE ruraux deviennent des strates géographiques nous pourrions en effet tirer un échantillon de la taille exacte souhaitée pour chaque GSE en milieu rural.

Nous disposons maintenant des éléments nécessaires pour tenter une définition exacte des GSE.

Pour le milieu urbain nous proposons en premier lieu une stratification géographique ainsi :

- U1 Dakar - commune
- U2 Autres communes de la Région de Dakar
- U3 Autres chefs lieux de régions + Mbour
 + Mbacke + Richard Toll

Les petites communes restantes seront classées avec le milieu rural.

²Diourbel, Kaolack, Thiès, Louga et Fatick.

Toujours en milieu urbain ce classement sera croisé avec un autre socio-économique. On souhaiterait distinguer les secteurs public et privé, et à l'intérieur de celui-ci le secteur formel du secteur informel. Comme les seules données disponibles à cette fin sont celles du recensement le choix des critères sera nécessairement assez simpliste. Nous proposons de classer les ménages en fonction du travail exercer par leur chef, en nous basant sur les variables situation dans la profession (Question P20 du RGPH) et branche d'activité (Question P22). Les catégories seraient les suivantes :

- C1 Salariés du secteur public
- C2 Salariés du secteur privé
- C3 Indépendants et aides familiaux
- C4 Autres (chômeurs et inactifs)

Quand au milieu rural, nous proposons un classement agro-écologique basé sur les zones établies par l'Institut Sénégalais de Recherche Agronomique. L'ISRA a en effet défini 14 zones mais des tailles extrêmement hétérogènes. Nous les avons regroupées et subdivisées dans deux buts : d'atteindre des zones de taille beaucoup moins variable et de nous approcher, là où cela était possible, aux régions administratives. L'ensemble des zones finalement proposées est le suivant :

- R1 Région de Thiès
- R2 Région de Louga
- R3 Région de Diourbel
- R4 Le Sine
- R5 Zone irriguée
- R6 Sud-ouest
- R7 Le Ferlo
- R8 Basse et moyenne Casamance
- R9 Départements de Foundiougne et Nioro

Rappelons qu'il s'agit partout ici du secteur rural proprement dit plus les petites communes. Pour les définitions plus exactes des ces zones, voir l'annexe.

3.3 Choix des GSE : Premier essai

Dans cette section nous examinons la répartition de

l'échantillon dans l'hypothèse que chacune des catégories définies plus haut constituera un GSE, ce qui implique 12 GSE urbains et 9 ruraux.

Commençons avec le cas urbain. Les calculs sont fait dans le tableau 1. La structure de ce tableau est conçue pour permettre au sondeur de voir l'impact précis sur l'échantillon des décisions prises relatives au choix des GSE. Les strates géographiques sont disposées en colonnes : les lignes correspondent aux classes socio-économiques. Pour ce 1er essai, chaque case du tableau représente un GSE potentiel, mais plus tard on pourra regrouper certaines cases afin de constituer les GSE plus convenables. L'indice h désigne la strate et l'indice g la caractéristique socio-économique.

Dans chacune des cases de ce tableau à deux entrées on cherche à inscrire trois quantités :

N_{gh} : le nombre de ménages de la population

n_{gh} : le nombre de ménages de l'échantillon

$$F_{gh} = n_{gh}/N_{gh}$$

On commence avec les N . Ces quantités proviendront du RGPH. Après avoir inscrit les N_{gh} on passe aux n_{gh} . On inscrira 400 pour représenter l'effectif de l'échantillon dans chaque case. Pour ce premier essai chaque GSE est constitué d'une seule case : on verra plus tard comment procéder au cas où un GSE regroupe plusieurs cases.

On passe ensuite aux F_{gh} . Dans chaque case on calcule $F_{gh} = n_{gh}/N_{gh}$ et on inscrit cette quantité en dessous du n_{gh} correspondant.

Enfin on fait le total $N_h = \sum_g N_{gh}$ pour chaque colonne en l'inscrivant en bas. On cherche alors la plus grande fraction de sondage, F_{max} et on l'inscrit en dessous de N_h . C'est le taux de sondage qu'il faudra appliquer à l'ensemble de la strate car il assure l'échantillon voulu pour le plus petit des éléments trouvés dans la colonne.

En multipliant N_h par F_{max} on obtiendra n_h , l'échantillon total par colonne, et en faisant la somme à travers les colonnes on obtiendra la taille globale de l'échantillon urbain. Nous voyons dans la dernière ligne du tableau que ce chiffre se lève à 8.729.

Passons au secteur rural. Ici les GSE sont les strates géographiques et les calculs sont plus simples. On pourra tirer exactement l'échantillon souhaité dans chaque strate. Nous proposons d'augmenter cette quantité de 400 à 500 : en effet le classement appliqué ici est moins raffiné que dans le secteur urbain et on peut prévoir que l'analyste cherchera des catégories plus exactes établies au niveau du ménage même et basées sur les données de l'enquête. Il semble prudent alors

Tableau 1 REPARTITION DE L'ECHANTILLON URBAIN

1er essai

		U1	U2	U3	
		DAKAR COMMUNE	PIKINE RUFISQUE	AUTRE URBAIN	TOTAL GENERAL
Secteur public	N	17.584	12.678	12.440	
	n	400	400	400	
	1000F	22,75	31,55	32,15	
Salariés du secteur privé	N	28.042	19.733	17.505	
	n	400	400	400	
	1000F	14,26	20,27	22,85	
Indépendants	N	24.213	30.234	44.605	
	n	400	400	400	
	1000F	16,52	13,23	8,97	
Autres	N	25.827	30.353	38.023	
	n	400	400	400	
	1000F	15,49	13,18	10,52	
TOTAL	Nh	95.666	92.998	112.573	301.237
	Fmax	22,75	31,55	32,15	
	nh	2.176	2.934	3.619	8.729
NOTE Pour les définitions plus exactes des strates voir l'annexe.					

de se donner une marge de sécurité en augmentant notre échantillon limite à 500. L'analyste aura ainsi un choix un peu plus large.

Le tableau 2 donne les paramètres qui résultent de ces dispositions quant au milieu rural.

Nous voyons que l'échantillon total s'élève à $8.729 + 4.500 = 13.229$. Ce total dépasse de loin la limite de 10.000 qui nous a été imposée. Comme le secteur rural comporte environ 60% des ménages sénégalais il est déjà sous-représenté ici par rapport à l'échantillon visé de 10.000. C'est donc dans le secteur urbain qu'il faudra chercher une réduction de la taille de l'échantillon. Examinons le tableau 1. Nous voyons que ce sont les strates U2 et U3 qui contribuent le plus au total, et surtout la U3. Le problème provient de ce que les taux de sondage, F_{max} , pour ces strates sont excessivement élevés, et ceci provient, à son tour, du fait que les ménages du secteur public ne sont pas très nombreux dans ces strates. Cette observation suggère la possibilité

qu'on regroupe les strates U2 et U3 en ce qui concerne le

Tableau 2 REPARTITION DE L'ECHANTILLON RURAL

ZONE ECOLOGIQUE	N	n	20/F
R1 Thiès	70.041	500	2802
R2 Louga	46.863	500	1875
R3 Diourbel	52.445	500	2098
R4 Le Sine	51.325	500	2053
R5 Zone irriguée	45.658	500	1826
R6 Sud-ouest	52.764	500	2111
R7 Le Ferlo	40.866	500	1635
R8 Casamance	78.565	500	3143
R9 Dép. Foundiougne et Nioro	43.754	500	1750

secteur public. Quel serait l'effet de ce regroupement? La première ligne du tableau sera modifiée de la façon suivante : on devra insérer une colonne correspondant au total des strates U2 et U3, et les $n = 400$ qui ont été jusqu'ici portés dans chacune des colonnes U2 et U3 deviendront une seule entrée $n = 400$ inscrit dans la nouvelle colonne U2+U3. Toujours dans cette case on calculera le taux $F = 400/25.118$ (d'où $1000F = 15.92$). On portera cette même valeur dans les deux colonnes U2 et U3 et on calculera le n qui en résulte. Ces manipulations paraissent dans l'extrait suivant :

			U2	U3	U2+U3
			PIKINE	AUTRE	TOTAL URBAIN
			RUFISQUE	URBAIN	SAUF DAKAR COM
Secteur	N	...	12.678	12.440	25.118
public	n	...	202	198	400
	1000F	...	15.92	15.92	15.92

Si on se réfère alors au tableau 1 on verra que les nouveaux F_{max} pour les colonnes U2 et U3 sont respectivement 20.27 et 22.85. Les nouveaux n_h qui en résultent en bas des colonnes sont 1.885 et 2.572 et l'échantillon total urbain devient $2.176 + 1.885 + 2.572 = 6.633$. Si on y ajoute les 4.500 du secteur rural on a 11.133. Malheureusement ce chiffre dépasse toujours notre limite de 10.000.

La solution évidente est de répéter cette manipulation pour la deuxième catégorie socio-économique, soit les salariés du secteur privé. Eux aussi sont trop peu nombreux en dehors de Dakar et méritent qu'on regroupe à leur égard les strates U2 et U3.

Le tableau 3 effectue ces deux ajustements simultanément et constitue notre 2e essai pour le milieu urbain.

Tableau 3 REPARTITION DE L'ECHANTILLON URBAIN

2ème essai

		U1	U2	U3	U2+U3	
		DAKAR	PIKINE	AUTRE	TOTAL URBAIN	TOTAL
		COMMUNE	RUFISQUE	URBAIN	SAUF DAKAR COM	GENERAL
Secteur	N	17.584	12.678	12.440	25.118	
public	n	400	202	198	400	
	1000F	22.75	15.92	15.92	15.92	
Salariés	N	28.042	19.733	17.505	37.238	
du secteur	n	400	212	188	400	
privé	1000F	14.26	10.74	10.74	10.74	
	N	24.213	30.234	44.605		
Indépendant	n	400	400	400		
	1000F	16.52	13.23	8.97		
	N	25.827	30.353	38.023		
Autres	n	400	400	400		
	1000F	15.49	13.18	10.52		
TOTAL	Nh	95.666	92.998	112.573		301.237
	Fmax	22.75	15.92	15.92		
	nh	2.176	1.480	1.792		5.448

Nous nous trouvons ainsi enfin en dessous de la limite imposée, l'échantillon total étant de 9.948 ménages. Les tableaux 2 et 3 représentent donc la répartition définitive proposée pour l'échantillon. Les GSE retenus sont représentés par les 10 cases du tableau 3 où on voit le chiffre $n = 400$, plus les 9 zones rurales du tableau 2.

Il importe de noter que le fait d'avoir visé certains GSE explicitement n'implique nullement qu'on ne pourra effectuer des analyses sur d'autres catégories. En effet toute catégorie dont l'échantillon dépasse 400 (urbain) ou 500 (rural) peut être considérée a priori comme analysable avec à peu près la même précision que celle des GSE que nous avons visés dans notre proposition finale. Rien n'exclut d'ailleurs la possibilité que l'analyste visant des objectifs plus modeste puisse tirer des résultats utiles à partir des sous-échantillons inférieurs à ces limites.

4. MODALITES DU TIRAGE DE L'ECHANTILLON

4.1 Principes

L'échantillon de DR sera tiré au sein de chaque strate avec probabilité proportionnelle à la taille, c'est-à-dire au nombre de ménages recensés dans chaque DR. Soit M_i ce nombre pour le DR i , et soit M'_i l'effectif de ménages enregistrés lors du dénombrement dans ce DR. Ces deux quantités seront souvent différentes, que ce soit à cause des mouvements réels des populations ou des erreurs dans l'une ou l'autre opération.

On montre facilement que la probabilité pour que le DR i soit inclus dans l'échantillon est donnée par la formule :

$$p_{1i} = a M_i / \Sigma M_i$$

dans une strate donnée, où a = l'effectif des DR tirés dans la strate et la somme ΣM_i s'étend sur la totalité des DR dans la strate (tirés ou non). L'expression p_{1i} représente la probabilité relative au 1er degré de sondage et à l'unité i .

Au 2e degré de sondage on tirera 20 ménages dans chaque DR, à partir des M'_i ménages dénombrés dans ce DR. La probabilité (conditionnelle) pour qu'un ménage soit tiré dans un DR tiré est donc :

$$p_{2i} = 20 / M'_i$$

La probabilité globale pour qu'un ménage paraisse dans l'échantillon est le produit de ces deux probabilités. En symboles :

$$F_i = p_{1i} p_{2i} = (a / \Sigma M_i) (M'_i / M_i)$$

Remarquons que dans le cas où $M_i = M'_i$ cette probabilité est constante pour tous les ménages de la strate. Un tel échantillon est appelé "autopondéré" à l'intérieur des strates. Mais puisqu'en fait M_i et M'_i diffèrent, il faudra pondérer les données lors de l'analyse par les coefficients suivants :

$$W_i = 1 / F_i = (\Sigma M_i / a) (M'_i / M_i)$$

Ces poids devront être calculés pour chaque DR de l'échantillon et inscrits dans le fichier des données de l'enquête.

4.2 Application et lien avec l'échantillon maître

Les paramètres du sondage sont implicites dans les tableau 2 et 3. Dans le cas du secteur rural, on trouve le pas de sondage pour chacune des 9 zones en lisant la colonne

de droite du tableau 2. Quant au secteur urbain un léger ajustement s'impose. La dernière ligne du tableau 3 nous donne l'effectif de ménages visé pour chacune des trois strates, mais comme on se propose de tirer exactement 20 ménages par DR il faudra d'abord arrondir ces chiffres au multiple de 20 le plus près. Le tableau 4 plus bas résume ces résultats en montrant le nombre de DR à tirer par strate. Nous ajoutons dans une dernière colonne le nombre d'unités primaires tirées pour l'échantillon maître dans chacune de nos strates.

Tableau 4 NOMBRE D'UNITES PRIMAIRES DE L'ECHANTILLON SOUHAITE ET DE L'ECHANTILLON MAITRE, SELON LA STRATE

Strate	Echantillon souhaité	Echantillon maître
	Nb. de DR	Nb d'unités primaires
U1 Dakar Commune	109	106
U2 Pikine, Rufisque	74	112
U3 Autre urbain	90	137
R1 Thiès	25	37
R2 Louga	25	23
R3 Diourbel	25	24
R4 Le Sine	25	31
R5 Zone irriguée	25	22
R6 Sud-ouest	25	28
R7 Le Ferlo	25	22
R8 Casamance	25	38
R9 Dép. Foundiougne-Nioro	25	20
TOTAL	498	600

Nous constatons que dans la moitié des strates proposées pour l'échantillon de l'ESP l'échantillon maître ne fournit pas le nombre voulu d'unités primaires. Nous devons conclure que l'échantillon de l'ESP ne peut pas être inclus dans l'EM³.

³Malgré cette insuffisance il serait toutefois possible de tirer le nombre voulu de DR au sein de l'EM si on se permettait de sélectionner parfois plusieurs DR dans un même secteur de l'EM. Il reste vrai que dans la mesure où nous insistons à ce que l'ESP soit basé sur un échantillon de DR tirés à un seul degré il est impossible d'inclure l'échantillon de l'ESP dans l'EM tel que celui-ci est actuellement constitué.

Rien n'empêche d'augmenter la taille de l'EM afin qu'il puisse comporter l'échantillon souhaité pour l'ESP - rien sauf le temps. Compte tenu du calendrier actuel pour l'ESP il nous semble plus prudent de supposer que cette enquête devra se baser sur un échantillon propre à lui.

Cette décision n'exclut pas la possibilité de profiter du travail effectué déjà dans le cadre de l'élaboration de l'EM qui a permis la stratification de la base de sondage et le regroupement des DR limitrophes dans des secteurs. Bien que les caractéristiques des strates de l'EM n'ont pas encore été chiffrées il serait certainement souhaitable d'utiliser celles-ci comme des "strates implicites" dans un sondage systématique.

Nous proposons donc que l'échantillon de DR, de la taille déjà indiquée pour chacune de nos 12 strates explicites, soit tiré par sondage systématique avec probabilités proportionnelles à la taille dans une liste des DR constituée ainsi:

- 1er niveau: ville ou communauté rurale
- 2e niveau: strate niveau-de-vie telle que définie pour les secteurs en vue de l'EM
- 3e niveau: secteur, classé par ordre de contiguïté géographique

L'échantillon de DR sera tiré indépendamment dans chacune des 12 strates avec un pas déterminés de la manière définie plus haut. La méthode pour tirer un échantillon avec probabilité proportionnelle à la taille est bien connue aux statisticiens de la DPS.

Dans chacun des DR tiré un dénombrement des ménages sera organisé. Au 2e degré de sondage on tirera 20 ménages dans chaque DR par tirage systématique à probabilité égale. Le pas de tirage ici sera obtenu en divisant l'effectif de ménages dénombrés ($=M_i$) par 20.

4.3 Sous-échantillon de ménages en vue des mesures anthropométriques

On souhaiterait tirer un sous-échantillon de 20% des enfants de 6 à 59 mois révolus en vue des mesures de poids et de taille. Nous proposons à cette fin un échantillonnage au niveau des ménages.

Dans la liste des 20 ménages tirés dans le DR on en tirera 4 pour ce sous-échantillon. Ce sous-échantillonnage sera effectué par la méthode systématique. On commencera avec un nombre aléatoire compris entre 1 et 5. Si, par exemple, ce nombre est 2, on sélectionnera parmi les 20, le 2e ménage, le

faisant partie du ménage.

4.4 Traitement des DR d'une taille aberrante

Deux petits problèmes se posent habituellement dans le tirage d'un échantillon de ce genre : les DR trop grands et les DR trop petits.

Les DR trop grands ne présentent aucun problème de fond. Le seul inconvénient vient au moment du dénombrement où elles exigent un volume de travail excessif. Au lieu d'engager les équipes dans ce travail les responsables préféreront peut-être l'option qui consiste à subdiviser les grands DR en des segments d'une taille plus convenable. Si cette solution est adoptée on fixera un seuil au-delà duquel les DR seront divisés. L'étude de la distribution des tailles des DR dans le fichier du recensement a donné les résultats suivants relatif aux deux queues de la distribution :

<u>Taille du DR</u> (nb.de ménages)	<u>Nb de DR</u>	
	Urbain	Rural
1 - 9	0	0
10 - 19	3	4
20 - 29	7	13
30 - 39	25	69
..		
300 - 399	31	1
400 - 499	8	0
500 - 599	1	0
600 - 699	0	0
706	1	0

A la lumière de ces résultats nous proposons qu'on adopte le seuil de 400 pour les grands DR. Tout DR de 400 ménages ou plus doit être subdivisé. On divisera l'effectif de ménages recensés par 200 pour déterminer le nombre de segments à créer. Ce travail ne sera entrepris cependant que si le DR est tiré. Pour chacun des DR concerné on enverra une équipe pour diviser le DR dans le nombre de segments prévus, des tailles approximativement égales. Dans certains cas ce travail pourrait être fait entièrement au bureau sur la base des cartes du recensement. Les segments seront alors insérés dans la liste des DR qui constitue la base de sondage en leur affectant une taille estimative ajustée pour que la taille totale du DR ne change pas. Ce procédé permettra de maintenir la série de tirage sans modification et de voir sur lequel des segments le point d'échantillonnage tombe. Le segment en question sera tiré. On maintiendra de cette manière le principe d'un sondage à un seul degré aréolaire.

Quant aux DR trop petits, le seul problème se pose si elles ont moins de 20 ménages, de sorte qu'on ne pourra tirer l'échantillon secondaire souhaité. Comme le nombre de ménages

Quant aux DR trop petits, le seul problème se pose si elles ont moins de 20 ménages, de sorte qu'on ne pourra tirer l'échantillon secondaire souhaité. Comme le nombre de ménages par DR n'est pas connu avant d'aller sur le terrain il serait prudent de fixer le seuil à un niveau un peu plus élevé que 20. A la lumière des résultats présentés plus haut nous proposons le seuil de 30. Les DR inférieurs à cette limite seront en principe regroupés avec un DR voisin. Les DR de cette taille sont très rares dans la base de sondage et seront encore plus rares dans l'échantillon, car celui-ci est tiré d'une manière à favoriser les grands DR. Il est probable qu'on n'aura pas plus d'un seul cas à traiter. Nous proposons que si un tel DR est tiré il soit regroupé avec son voisin précédant dans la liste et qu'aucune autre mesure ne soit prise à cet égard. L'approximation est suffisant pour éviter tout danger de biais appréciable.

5 PARAMETRES DES TRAVAUX DE TERRAIN

Les différentes options ont été étudiées en détail. En voici les conclusions, qui devraient être considérées comme provisoires jusqu'à ce que l'enquête pilote ait été achevée.

- Le dénombrement des ménages sera entrepris au fur et à la mesure des travaux de terrain. Ce sera donc la première tâche de l'équipe dans chaque DR avant d'effectuer les interviews.
- On prévoit 60 minutes pour l'interview au début; 50 minutes une fois que les enquêteurs seront rodés.
- On prévoit une moyenne de 5 interviews par jour et par enquêteur une fois que l'équipe est sur place.
- Pour chaque DR on prévoit 1 jour de déplacement, contact avec les autorités et reconnaissance des limites du DR; 1 jour pour le dénombrement des ménages et le tirage de l'échantillon, et 1 jour pour les interviews.
- Chaque équipe sera composée de 4 enquêteurs, 1 contrôleur, 1 anthropomètre et 1 chauffeur.
- Les enquêteurs travailleront 6 jours par semaine.

Sur la base de ces paramètres on estime que le travail pourra être fait par 14 équipes en 4 mois.