

ENQUETE PERMANENTE AUPRES DES MENAGES

Plan de sondage proposé

JN 2741

1. Objectifs et contraintes

L'objectif principal du plan de sondage est de fournir un échantillon de 4.500 ménages qui serait rigoureusement probabiliste, statistiquement efficace et représentatif de l'ensemble de Madagascar.

Un des objectifs secondaires est de pouvoir présenter des résultats, avec un degré de précision utile, au niveau des milieux urbain et rural séparément. Or la précision d'une estimation dépend, autres choses égales, de la taille de l'échantillon. Comme le secteur urbain ne constitue que 20 à 25% de la population un échantillon de 4.500 tiré à probabilité égale sur l'ensemble de Madagascar n'aurait que 1.000 ménages pour représenter le secteur urbain. Avec un échantillon aussi faible les possibilités pour les analyses relatives à ce secteur seraient très restreintes. Cette situation nous a amené à vouloir doubler le taux de sondage urbain relatif au rural. Une telle stratégie, couramment appliquée dans des enquêtes de ce genre, se justifie également par deux autres arguments : 1° les activités économiques des ménages sont plus variables en milieu urbain, et 2e la collecte des données y coûte moins cher. Il est donc statistiquement plus efficace de sonder avec une intensité plus élevée en milieu urbain.

Un autre objectif secondaire est de pouvoir donner des résultats utilisables au niveau des 6 faritany (provinces). Si on voulait accorder la priorité absolue à cet objectif on affecterait un échantillon de la même taille (donc de 750 ménages) à chaque faritany. Comme la population par faritany varie par un facteur de 3,3 à 1, un tel échantillon exigerait une variation de cet ordre dans les taux de sondage par faritany. Or des telles variations des taux de sondage (qui seraient, bien entendu, compensées lors de l'analyse par des pondérations inverses) sans rapport avec les variances et les coûts, diminueraient d'une façon appréciable l'efficacité des estimations au niveau national. Choisir ce type d'échantillon serait répondre à un objectif secondaire au détriment de l'objectif primaire. Cette solution est à notre avis à écarter, au moins en ce qui concerne l'échantillon de la première année de l'enquête.

La taille totale de 4.500 ménages a été fixée en fonction de la disponibilité des crédits et en tenant compte des contraintes d'ordre pratique. Une enquête plus grande serait difficile à gérer et contrôler.

Ce total correspond au maximum qu'on a jusqu'ici admis pour les enquêtes parrainées par la Banque Mondiale (programmes LSMS et DSA). La question nous a été posée à plusieurs reprises : pourra-t-on obtenir des résultats valables au niveau des faritany avec un tel échantillon? Il n'est pas possible de répondre avec un simple oui ou non. Tout dépend de deux facteurs : la variance des variables concernées telle qu'elle existe dans la population et la complexité du tableau visé. Le premier facteur n'est connu en général qu'avec une précision très approximative, mais sera facile à estimer à partir des résultats de la première année de l'enquête. Quant au deuxième facteur, on pourra dire dès maintenant que certains tableaux très simples, par exemple à une seule entrée, pourront être obtenus au niveau de chaque faritany. Par contre lorsqu'on considère les tableaux de plus en plus complexes et à multiples entrées, les cases vides, ou insuffisamment peuplées paraîtront de plus en plus fréquemment, et cela surtout dans les faritany les moins grandes. En somme, il n'y a jamais un échantillon "suffisant". Plus l'échantillon est grand, plus on pourra désagréger les données mais on ne fera jamais tout-ce qu'on aimerait faire. Mais d'autre part, plus l'échantillon total devient grand plus on augmente les chances d'une perte de contrôle sur la précision des opérations : autrement dit, il y a une relation inverse entre la précision du sondage et la qualité des données collectées sur le terrain.

## 2.1 Méthodologie de la collecte

Le plan de sondage devra être adapté aux exigences des procédures envisagées pour la collecte des données. Ces procédures détermineront la fréquence et la périodicité des visites à chaque ménage, le nombre de ménages qui doivent être enquêtés dans chaque localité et éventuellement des différences entre les milieux urbain et rural quant aux paramètres de l'échantillonnage.

Une méthodologie de collecte adaptée aux spécificités malgaches a été proposée dans un rapport précédant du présent consultant.<sup>†</sup> Il convient de préciser ici ses éléments principaux en tant qu'ils influent sur le plan de sondage.

Le travail de terrain sera effectué par des équipes mobiles. Chaque équipe sera constituée de deux enquêteurs principaux (peut-être trois dans certaines équipes urbaines), d'un contrôleur et d'un chauffeur. L'équipe circulera dans une région limitée autour de son siège, qui se trouve en général dans une des capitales des faritany et qui comporte un bureau, un micro-processeur et un agent de la saisie des données. Les questionnaires seront passés au moins deux fois par mois à l'agent de saisie.

La possibilité d'ajouter un enquêteur de réserve à chaque équipe sera examinée plus loin (chapitre 5).

Le questionnaire est divisé en une dizaine de modules qui, dans leur ensemble, peuvent exiger plus de 5 heures d'interrogatoire pour un ménage

---

<sup>†</sup> Enquête Permanente auprès des ménages. Programmation du travail des équipes de terrain : 1er projet. C. Scott.

donné. Pour cette raison on propose de répartir le travail entre 5 interviews étalées sur une période totale de 31 jours en milieu urbain ou environ 20 jours en milieu rural. Ces dispositions présentent d'ailleurs deux autres avantages :

- 1° En étalant sur 4 visites l'interrogatoire sur les dépenses du ménage on pourra couvrir une période assez longue par ménage, tout en limitant (cas rural: ) la période de mémoire pour une visite donnée ou (cas urbain: ) les périodes sur lesquelles le ménage est censé remplir son journal de transactions sans surveillance.
- 2e En fixant la date de la dernière visite à une dizaine de jours après les autres on laissera un temps suffisant pour que l'agent de saisie puisse produire, et que les enquêteurs puissent ramasser, la liste des erreurs trouvées par le micro-processeur relatives aux 3 ou 4 premières interviews. Les enquêteurs pourront corriger ces erreurs lors de leur dernière visite au ménage.

Le travail se divise en des cycles de 33 jours. En milieu urbain : un enquêteur s'occupera de 24 ménages par cycle, répartis dans 2 grappes : les enquêteurs d'une même équipe travailleront dans des grappes différentes mais si possible dans la même ville. En milieu rural les 2 enquêteurs travaillent dans la même grappe, en enquêtant 8 ménages chacun par grappe : après une dizaine de jours ils seront transférés à une autre grappe (en passant si possible par leur siège) où ils doivent enquêter pendant encore 10 jours. Enfin ils retourneront à la première grappe pour 2 jours, puis à la deuxième grappe pour encore 2 jours, ces courtes visites permettant le passage des derniers modules ainsi que la correction des erreurs. En somme, les 2 enquêteurs compléteront 2 grappes (32 ménages) par cycle en milieu rural, alors qu'en milieu urbain le même nombre d'enquêteurs compléteront 4 grappes (48 ménages) par cycle.

Ces différents paramètres devront être pris en compte dans l'élaboration du plan de sondage.

### 3. Base de sondage aréolaire

On cherchera d'abord une base de sondage aréolaire. Dans le cas idéal on trouverait un ensemble d'unités aréolaires, bien cartographiées, d'une population par unité qui soit à la fois connue, modeste (moyenne autour de 1.000 habitants) et assez constante (coefficient de variation inférieur à 40%).

Dans la plupart des pays les districts de recensement correspondent suffisamment bien à ces conditions. Mais à Madagascar le dernier recensement remonte à 1975, ce qui serait excessivement vieux pour nos besoins. Par contre une nouvelle cartographie en vue du recensement prochain (1990?) sera lancée dans l'immédiat mais on ne peut guère espérer disposer de ses résultats avant la mi-1990.

En l'absence des districts de recensement nous avons étudié le découpage administratif comme source possible d'unités aréolaires de sondage. L'hierarchie des unités administratives à Madagascar est la suivante :

6	faritany	(provinces)
110	fivondrona	(sous-préfectures)
1252	firaisan	(cantons)
11393	fokontany	(villages, localités)

Les fokontany sont normalement bien connus à la population bien qu'il n'existe pas de cartes sur lesquelles ils sont localisés. Un recensement administratif est entrepris en principe chaque année : celui de 1987 est disponible au EDE (moyennant un certain nombre de rectificatifs et d'estimations pour des données manquantes). Le Service Informatique de la EDE a pu produire un fichier complet jusqu'au niveau du fokontany avec la population administrative de chaque unité.

A première vue cet ensemble répond suffisamment bien à nos besoins. Certes les populations seront souvent très imprécises mais on pourra adopter une méthode d'échantillonnage qui ne suppose pas leur exactitude (voir chapitre 6). ...

Le tableau 1 donne les principaux paramètres des fokontany en fonction de la population administrative. Le classement urbain/rural est celui du code officiel géographique.

TABLEAU 1 Taille des fokontany  
(populations administratives, 1987)

Milieu	Nombre de fokontany	Population moyenne	Ecart-type	Coefficient de variation
Urbain	841	2.959	2.560	87%
Rural *	10.515	952	891	94%

\* Nous avons exclu le cas aberrant trouvé dans l'échantillon avec données erronées - voir tableau 2, note 2.

On voit tout de suite deux problèmes : 1° Les fokontany sont d'une taille excessivement variable et 2° La taille moyenne de ceux du milieu urbain est bien plus grand qu'on l'aurait souhaité.

Le 1er problème peut être surmonté en appliquant la méthode du sondage avec probabilité proportionnelle à la taille (voir chapitre 6). Le seul inconvénient en est qu'on sera amené à inclure dans l'échantillon un plus grand nombre de grands fokontany. Ainsi les deux problèmes cités auraient la même conséquence nocive pour l'enquête : un nombre excessif de grands fokontany dans l'échantillon. Ce phénomène est nocif parce qu'il implique une augmentation excessive du travail de dénombrement des ménages.

L'opération du dénombrement des ménages sera étudiée dans le chapitre 5. Pour l'instant nous remarquerons seulement que, si on devait dénombrer tous les ménages dans chaque fokontany dans un échantillon tiré avec probabilité proportionnelle à la taille on serait obligé de dénombrer autour de 250.000 ménages, ce qui occuperait une trentaine d'agents à plein temps pendant bien plus de 6 mois. C'est évidemment une solution à éviter à tout prix.

Comme il n'existe pas d'unité aréolaire reconnue qui soit plus petite que le fokontany la seule solution est d'en créer une. Nous proposons donc de diviser les fokontany tirés dans des segments d'une taille plus convenable et d'en tirer un dans chaque fokontany. Cela implique, bien entendu, une opération supplémentaire de terrain : délimitation de segments.

Le premier problème est de fixer une norme quant à la taille d'un segment. Plus les segments sont petits, plus on aura réduit le travail du dénombrement des ménages, mais en revanche plus on aura du travail pour délimiter les segments. Le choix d'un optimum pourrait être étudié dans les semaines qui viennent en fonction de l'organisation des deux opérations, dénombrement des ménages et délimitation des segments. On n'oubliera pas notamment qu'en fixant le seuil plus haut on augmentera le nombre de fokontany qui échapperont entièrement la nécessité d'une subdivision en segments. Vraisemblablement la norme se situerait entre 800 et 1200 personnes, mais rien ne nous oblige de fixer la même norme partout. Le tableau 2 (à compléter en ce qui concerne le milieu urbain) doit aider dans l'établissement des normes.

Il s'agit ensuite de délimiter dans chaque fokontany le nombre de segments impliqué par la norme. Par exemple, si on fixe la norme à 1000, tout fokontany dont la population administrative ne dépasse pas 1500 ne doit pas être subdivisé : ceux qui se situent entre 1500 et 2500 seront divisés en 2 segments, ceux entre 2500 et 3500 en 3 segments, etc. En consultant le tableau 2 on voit, par exemple, qu'en supposant une norme de 1000 on aurait 69 fokontany ruraux à subdiviser sur le total de 166. Comme le seul but de la segmentation est de réduire le travail du dénombrement des ménages on ne sera pas obligé de respecter la norme avec rigueur. Dans un endroit où la délimitation de segments se révèle particulièrement difficile rien nous empêche de créer moins de segments et de laisser plus de travail aux dénombreurs (et, à un degré moindre, aux enquêteurs, qui devront marcher plus si les segments sont plus grands).

La subdivision en segments sera faite sur le terrain. Une petite équipe (2 agents plus 1 contrôleur?) se rendra dans le fokontany pour dessiner un croquis et tracer là-dessus les limites des segments convenables. On utilisera des limites naturelles autant que possible. Il n'est pas obligatoire que les segments soient d'une taille strictement égale mais on évitera des très grandes différences de taille (un segment plus de 3 fois plus grand qu'un autre, par exemple). A la fin du travail dans le fokontany  $i$  l'équipe devra estimer très approximativement le pourcentage  $S_{ij}$  que représente chaque segment  $j$  dans la population du fokontany  $i$ . Ces chiffres peuvent être basés sur le nombre de logements. Ils n'ont pas besoin d'être très précis à condition qu'ils s'ajoutent à 100% dans le fokontany.

A Antananarivo, et peut-être dans certaines autres grandes villes, il semble qu'il existe déjà un découpage des fokontany en secteurs (ou quartiers). Ces zones seraient connues aux autorités du fokontany.

TABLEAU 2 Taille des fokontany tirés  
dans l'échantillon  
(Population administrative, 1987)

Classe- taille	F R E Q U E N C E	
	Urbain	Rural
1- 99		-
100-199		1
200-		4
300-		7
400-		4
500-		11
600-		12
700- ...		12
800-		12
900-		10
1000-		4
1100-		8
1200-		5
1300-		4
1400-		3
1500-1999		18
2000-2999		24
3000-		9
4000-		8
5000-		5
6000-		3
7000-		-
8000-		1
9000-9999		-
32964		1
	154	166

NOTES

1. Données à fournir plus tard pour l'échantillon urbain.
2. Le cas exceptionnel en bas du tableau représente une zone qui n'avait pas été divisée en fokontany au moment de la constitution de la liste. Son classement correcte reste à déterminer.

Normalement on devrait pouvoir demander à ces autorités la population de chaque secteur. On pourrait alors se baser sur ces unités sans avoir des segments à créer, à condition que les secteurs ne sont pas trop grands. Au cas où ils sont trop variables de taille on pourrait regrouper certains uns. C'est un problème à étudier dans les semaines qui viennent.

En milieu urbain, et surtout à Antananarivo au cas où les secteurs ne se révèlent pas utiles, les cartes ou photographies aériennes peuvent faciliter le travail du découpage en segments. Nous nous sommes adressés à l'Institut de Géodésie et Cartographie (ITM) où la Division des Photographies Aériennes (M. Rakotozafy Robert, Adj; Chef) nous a donné des renseignements extrêmement utiles. Une couverture aérienne à 1 : 15.000 des centres des villes avait été commandée par La Poste (PTT) et effectuée en 1985-86. On devra pouvoir obtenir des photos à un prix modeste moyennant l'autorisation du service qui les a commandées initialement. A Antananarivo même, on pourrait peut-être compléter la couverture avec des photos commandées par la Direction des Routes. Rappelons, cependant, que dans la mesure où les secteurs peuvent servir de segments les photos ne seront pas nécessaires.

Enfin on devrait étudier la possibilité d'une aide de la part du projet Cartographie Recensement qui est sur le point d'être lancé, quant au travail de la délimitation des segments en milieu rural.

Passons maintenant à l'autre bout de l'échelle pour examiner la question des fokontany excessivement petits.

La limite inférieure pour notre unité finale aréolaire est imposée par l'exigence qu'il faudra y tirer un nombre donné de ménages, soit 16 en milieu rural ou 12 en milieu urbain. Seize ménages correspondraient normalement à 80 personnes mais dans des cas exceptionnels le chiffre pourrait peut-être monter jusqu'au double de cela. Etant donné la sur-estimation probable de certaines des estimations en milieu rural on devrait peut-être tracer la limite à 200 personnes. Donc, au cas où on tire un fokontany dont la population administrative est inférieure à 200 on devra le regrouper avec un fokontany voisinant.

Un seul fokontany a été effectivement tiré qui soit en dessous de cette limite : il possède 139 habitants.

En principe le regroupement des unités trop petites avec des unités voisines devrait être effectué avant le tirage. Comme il faut des efforts considérables pour déterminer quels sont les fokontany limitrophes on s'est permis de négliger cette condition dans l'unique cas de l'unité mentionnée. Pour traiter de cette unité on cherchera donc un fokontany limitrophe et dans le même firaïsan, on regroupera les deux fokontany et on les traitera comme une seule unité tirée par la méthode normale, en supposant une taille égale à la somme des tailles des deux fokontany.

#### 4. Classement urbain/rural

Le code géographique, basé sur le recensement de 1975 et encore utilisé pour la présentation des résultats du recensement administratif, distingue 48 villes considérées comme urbaines en 1975. Nous avons étudié la possibilité de mettre à jour ce classement. Les difficultés en seraient pourtant considérables car la ville n'a aucune place dans l'hierarchie des unités

administratives sur laquelle nous nous basons. D'ailleurs il semble probable que la mise-à-jour, même si on pouvait l'effectuer, ne changerait pas beaucoup la situation par rapport à l'ancien classement.

Il a été donc décidé de maintenir le classement géographique tel qu'il est. Celui-ci classe 20% de la population comme urbaine selon le recensement administratif. Le vrai pourcentage est probablement plus près à 23%. La différence vient non pas de l'erreur de classement mais de la sur-estimation du recensement administratif en milieu rural.

Le classement que nous évoquons ici concerne, et détermine, le taux de sondage à appliquer pour un fokontany donné (en principe 2 fois plus élevé pour les villes) ainsi que le coefficient de pondération à lui appliquer lors des analyses (en principe 2 fois plus petit pour les villes). Rien ne nous oblige, cependant, de respecter le même classement en ce qui concerne la méthodologie de la collecte (voir le chapitre 2, où on a distingué les méthodes à utiliser en milieu urbain et rural), ni en ce qui concerne l'analyse des résultats, où l'analyste est parfaitement libre à classer les milieux selon toute définition qui lui convienne, dans les limites des données qui lui sont offertes. Il conviendra donc d'établir trois variables urbain:rural, qu'on pourra désigner ainsi :

U/R SONDA : concerne le taux de sondage  
U/R OP : concerne les procédures opérationnelles de collecte  
U/R REEL : concerne les caractéristiques d'intérêt analytique

En ce qui concerne le calcul du coefficient de pondération pour une zone donnée on devra tenir compte à la fois de U/R SONDA et de U/R OP. Le problème sera examiné au chapitre 7.

Les trois variables devraient figurer dans le fichier pour chaque unité finale aréolaire (fokontany ou segment). Les éléments pour la construction de la 3ème variable pourraient être obtenus au cours de l'enquête au niveau communautaire.

##### 5. Base de sondage pour les ménages

Aucune liste valable des ménages<sup>n</sup> étant disponible à Madagascar nous serons obligés d'établir notre propre liste dans chaque segment de l'échantillon. Ces listes constitueront la base de sondage au 3e degré.

Avant d'étudier les modalités de cette opération il convient d'ajouter quelques précisions sur la méthode qui est proposée pour lier les échantillons à travers des années. L'EPAM est un projet permanent : en principe l'enquête ne s'arrêtera jamais. A la fin de l'année on propose de garder la moitié des zones<sup>n</sup> tirées et de remplacer les autres par un nouvel échantillon qui

---

\*\* A partir d'ici nous adopterons le terme "zone" pour désigner l'ultime unité aréolaire de sondage, le plus souvent un segment mais parfois un fokontany entier et parfois peut-être un secteur urbain.

serait tiré, en général, suivant les mêmes principes.<sup>++</sup> Dans les zones restant inchangées on retournera aux mêmes ménages-échantillon autant que possible, mais à la fin de la 2<sup>e</sup> année la zone disparaîtra de l'échantillon. Dans les zones nouvellement tirées on doit effectuer un dénombrement des ménages.

Ces dispositions ont pour conséquences que :

- 1° Pour la 1<sup>ère</sup> année on devra dénombrer les ménages dans toutes les zones de l'échantillon
- 2° Pour les années futures on devra normalement dénombrer les ménages seulement dans la moitié des zones
- 3° Pour les années exceptionnelles on pourra néanmoins être appelé à dénombrer toutes les zones.

Comment va-t-on organiser le dénombrement? Trois méthodes ont été proposées.

- A. On pourrait organiser un dénombrement complet de l'échantillon préalablement à l'enquête. (Côte d'Ivoire)
- B. On pourrait organiser deux ou trois petites équipes de dénombreurs qui feraient le dénombrement en 3 ou 4 vagues, de 2 ou 3 mois chacune, espacées dans l'année. Durant chaque vague ils se déplaceraient sur l'ensemble du territoire en s'occupant des zones qui sont destinées pour des visites prochaines des enquêteurs. (Mauritanie)
- C. On pourrait affecter à chaque équipe un agent spécialisé qui fera le dénombrement juste en avance de l'arrivée de l'équipe dans la zone. (Méthode proposée pour le Sénégal). La moitié de son temps il restera avec l'équipe d'enquêteurs, où il pourra entreprendre des tâches diverses (anthropométries? relevés des prix? questionnaire communauté?). Au cas où la zone à dénombrer est particulièrement grande l'agent dénombreur pourrait être épaulé par les autres membres de l'équipe, qui seraient transportés dans la nouvelle zone, un peu en avance du programme normal.

Chaque méthode a ses avantages et inconvénients. La méthode A est la plus simple mais pose la question de la méthode pour les années futures. Comment va-t-on faire le dénombrement les années après la première lorsque les voitures seront en emploi permanent avec les équipes d'enquêteurs? Malgré cette objection on pourrait peut-être envisager d'adopter la méthode A pour la première année et une autre méthode pour les années à suivre. L'autre inconvénient de la méthode A est que les listes seront plus vieilles au moment de l'enquête. On aura beaucoup de ménages à remplacer et cela entraînera des biais.

La méthode B s'adapte particulièrement à une population d'une forte mobilité, ce qui explique son adoption en Mauritanie. Elle est plus chère et moins pratique que les autres.

La méthode C assure une liste aussi à jour que possible. On profite d'ailleurs de la présence de l'équipe, avec sa voiture, dans une zone proche. Les déplacements sont ainsi réduits, avec les coûts. Le principal inconvénient est que l'agent de dénombrement travaillera sans supervision directe, son contrôleur étant avec l'équipe principale. Au

---

<sup>++</sup> Voir page 10

Sénégal on a essayé de combler ce problème en envoyant le contrôleur accompagner le dénombreur pendant le premier jour (ou les 2 premiers jours) de sa visite dans la zone à dénombrer avant de rentrer avec le chauffeur dans la zone où se trouvent les enquêteurs.

On devra choisir entre ces options pour le Madagascar. Il s'agit d'examiner de près l'organisation du travail avec les distances et les moyens de déplacement.

## 6. Structure de l'échantillon

Selon le schéma d'activités des équipes de terrain, lorsqu'une équipe se déplace d'une zone à une autre elle passera toujours par son siège local. Par conséquent on n'aurait aucun intérêt à regrouper les zones échantillons par l'introduction d'un degré de sondage antérieur au fokontany. Comme chaque degré supplémentaire de sondage doit normalement augmenter l'erreur de l'échantillonnage on fixera le fokontany comme unité primaire.

L'unité secondaire sera le segment. Pour les fokontany plus petits que le seuil le segment est identique au fokontany : dans de tels cas on peut dire que la probabilité de sondage au 2e degré est de 1.

L'unité tertiaire, et finale, sera le ménage.

En règle générale un échantillon est efficace dans la mesure où il n'y a pas une grande <sup>+++</sup> variabilité entre les probabilités globales relatives aux unités finales <sup>+++</sup>. Or la probabilité globale est le produit des probabilités conditionnelles aux 3 degrés. Ecrivons :

$$P = P_1 P_2 P_3$$

Au 3ème degré on va tirer un nombre fixe de ménages dans chaque zone : 12 dans l'urbain ou 16 dans le rural - mettons c dans le cas général. Ce nombre représentera un taux élevé si la zone est petite et un taux faible si la zone est grande. Voilà donc déjà une source de variation. On pourrait l'éliminer si on pouvait créer des segments d'une taille constante. Malheureusement il est encore plus important que les segments

---

++ (NOTE REPORTEES DE PAGE 9). Cette règle générale sera cependant écartée dans des circonstances anormales. Par exemple, si on souhaite une année introduire une nouvelle base de sondage (après le recensement?) ou si on désire introduire un échantillon qui vise particulièrement certains groupes cibles.

+++ Cette proposition suppose que la variation en question n'est pas en corrélation avec la variance de la variable telle qu'elle existe dans la population. Dans le cas contraire des variations dans les probabilités peuvent être avantageuses.

soient établis avec des limites bien définies et réperables, d'où une certaine variation dans leur taille est inévitable. Une solution plus commode est de tirer les segments avec probabilités proportionnelles à leur taille. Avec cette méthode, là où  $p_3$  est élevée (petit segment)  $p_2$  sera faible, et là où  $p_3$  est faible (grand segment)  $p_2$  sera élevée. On obtiendra ainsi  $p_2 p_3$  constante. Le même principe s'applique entre le 1er et le 2ème degré de sondage.

En termes mathématiques, soit  $M_{ij}$  le nombre de ménages dans le segment  $j$  du fokontany  $i$ , et soit  $M_i$  le nombre total de ménages dans ce fokontany.

Tirons les fokontany (1er degré) avec des probabilités proportionnelles à leur taille  $M_i$ . Nous aurons :

$$P_{1i} = k_1 M_i \quad \text{-----} \quad (1)$$

On montre facilement que la constante  $k_1$  est égale à  $a / \sum M_i$ , où  $a =$  le nombre de fokontany tirés et la somme  $\sum$  s'étend sur l'ensemble des unités existantes.

Ensuite tirons 1 segment dans chaque fokontany, encore une fois avec probabilité proportionnelle à la taille. On aura :

$$P_{2ij} = k_2 M_{ij}$$

où  $k_2 = 1 / \sum_j M_{ij}$  = formule identique à celle utilisée pour déterminer  $k_1$

mais en remplaçant  $a$  par 1 car on ne choisit qu'une unité segment dans chaque unité fokontany. Or la somme  $\sum_j M_{ij}$  est égale à  $M_i$ . Donc :

$$P_{2ij} = \frac{1}{M_i} M_{ij} \quad \text{-----} \quad (2)$$

Enfin au 3ème degré nous tirerons  $c$  ménages sur les  $M_{ij}$  existants, d'où

$$P_{3ij} = \frac{c}{M_{ij}} \quad \text{-----} \quad (3)$$

La probabilité globale d'être tiré pour un ménage dans la zone  $ij$  est donc :

$$\begin{aligned} P_{ij} &= P_{1i} P_{2ij} P_{3ij} \\ &= k_1 M_i \frac{M_{ij}}{M_i} \frac{c}{M_{ij}} \\ &= k_1 c = \text{constante.} \end{aligned}$$

En principe ce système assure donc un échantillon d'une haute efficacité.

Dans la pratique on ne peut atteindre qu'une approximation assez grossière de cette idéale car les quantités  $M_i$  et  $M_{ij}$  ne sont pas connues avant le tirage. Pour s'approcher de  $M_i$ , on ne dispose que de la population administrative  $N_i$  de 1987. Quant à  $M_{ij}$ , on aura les estimations  $E_{ij}$  (voir chapitre 3) du pourcentage de la population du

fokontany  $i$  qui est estimé résider dans le segment  $ij$ . On peut écrire l'équation (2) sous la forme :

$$P_{ij} = S_{ij} / 100$$

Enfin au 3ème degré on peut maintenir l'équation (3) si on suppose que  $H_{ij}$  est le nombre de ménages effectivement dénombrés.

En somme on aura dans la pratique :

$$P_{ij} = \frac{a H_i}{\sum H_i} \cdot \frac{S_{ij}}{100} \cdot \frac{c}{H_{ij}} \quad (4)$$

On devra calculer la quantité  $a$  séparément pour les cas urbain et rural (définition U/R SOND) ainsi que le paramètre  $c$  (définition U/R OP). Nous traiterons de ces calculs au chapitre suivant.

La formule (4) ne donnera pas des  $P_{ij}$  constantes. Cependant les valeurs des  $P_{ij}$  seront certainement beaucoup moins variables que si on n'avait pas fait intervenir les tailles estimatives dans la méthode de sondage. Notre procédure proposée augmentera beaucoup l'efficacité de l'échantillonnage tout en la laissant assez loin de l'efficacité maximale théorique.

### 7. Paramètres de l'échantillonnage et de l'estimation

Le tableau 3 donne la répartition de la population de la base de sondage entre les faritany (provinces) et les milieux urbain et rural (U/R SOND).

TABLEAU 3 Répartition de la population dans la base de sondage, selon le faritany et le milieu urbain/rural

Faritany	Urbain	Rural
Antananarivo	1 063 987	2 417 248
Fianarantsoa	356 669	2 458 670
Taomasina	299 448	1 602 242
Mahajanga	268 761	1 203 281
Toliary	295 115	1 553 940
Antsiranana	204 945	795 376
TOTAL	2 488 925	10 030 757

NOTE : Les îles de Nosy Be et Ste Marie (population totale 53.000) ont été exclues de la base de sondage.

Nous avons estimé la taille des ménages à partir des enquêtes de 1978 et 1980 auprès des ménages, et dans l'hypothèse que le secteur urbain représente 20% de la population, avec les résultats suivants :

Taille moyenne des ménages. Urbain 4,36  
 Rural 5,16  
 Ensemble 5,00

Introduisons ici quelques symboles :

	<u>Urbain</u>	<u>Rural</u>
Population total, base de sondage	$N_u$	$N_r$
Taille estimée du ménage	$H_u$	$H_r$
Nombre de ménages correspondant	$M_u = N_u / H_u$	$M_r = N_r / H_r$
Facteur de sur-estimation supposé	1	$1 + d$
Population réelle impliquée	$N_u$	$N_r / (1+d)$
Effectif de ménages échantillons visé	$m_u$	$m_r$
Taux global de sondage impliqué	$F_u = m_u / M_u$	$F_r = m_r (1+d) / M_r$

Nous estimons le facteur d de la façon suivante. On suppose d'abord que la population totale réelle est actuellement de 11.000.000. On soustrait 50.000 pour les files éliminées (voir tableau 3, note). On suppose que le total urbain dans le tableau 3 est exact. On obtient :

$$N_u + N_r / (1+d) = 10.950.000,$$

d'où  $d = 0,1855$ .

Comme on propose de sonder avec un taux 2 fois plus élevé en milieu urbain qu'en milieu rural, on aura  $F_u = 2F_r$ .

Enfin on nous impose la taille total de l'échantillon, d'où

$$m_u + m_r = 4.500$$

En éliminant progressivement  $m_r$ ,  $F_u$ ,  $F_r$ ,  $M_u$  et  $M_r$  on obtient

$$m_u = F_u M_u = 4500 / \left[ 1 + \frac{1}{2(1+d)} \cdot \frac{N_r}{N_u} \cdot \frac{H_u}{H_r} \right]$$

Enfin, en substituant les valeurs numériques citées plus haut on obtient :

$$m_u = 1847$$

d'où  $m_r = 2653$

Si on suppose (hypothèse A) l'identité des variables U/R SOND et U/R OP on sait qu'on doit tirer exactement 12 ménages par zone en milieu urbain et 16 par zone en milieu rural. On devra donc arrondir  $m_u$  au multiple le plus proche de 12, et  $m_r$  au multiple le plus proche de 16. On obtient les valeurs ajustées :

$$m_u = 1848$$

$$m_r = 2656$$

ce qui implique un échantillon total de 4.504.

On doit donc tirer 154 zones en milieu urbain (1848/12) et 166 en milieu rural (2656/16). Ce sont les quantités  $a_u$  et  $a_r$  qui correspondent au paramètre  $a$  à la formule (4).

Ces quantités correspondent à 77 cycles-enquêteurs en milieu urbain et 166 cycles-enquêteurs en milieu rural, soit 243 au total. Comme chaque enquêteur sera actif pendant 10 cycles on suppose 25 enquêteurs, qui pourraient être répartis dans 11 équipes de 2 et une équipe de 3. (Nous ne parlons pas ici de l'équipe supplémentaire de réserve.)

On peut maintenant fixer le pas de tirage (voir chapitre 8) ainsi :

$$R_u = \frac{N_u}{154} = 16.162$$

$$R_r = \frac{N_r}{166} = 60.426$$

Comme le tirage de l'échantillon doit être effectué avant que les travaux de planification de l'enquête peuvent avancer, nous proposons que l'hypothèse A soit acceptée à ce stade. Si plus tard on veut modifier U/R OP pour certaines zones (par exemple, si des zones classées urbaines ne contiennent pas suffisamment de ménages alphabètes pour qu'on utilise la méthode des journaux de transactions) on pourra modifier la variable U/R OP pour ces zones mais sans changer l'échantillon. On verra un léger changement en ce qui concerne la taille de l'échantillon mais il n'y aura aucune incohérence dans cette procédure, à condition qu'on introduise la modification nécessaire dans le coefficient de pondération. Etudions maintenant ce problème.

Toutes les données de l'enquête devront être pondérées, en vue de leur tabulation et analyse, pour tenir compte de leurs probabilités inégales de sondage. Le coefficient de pondération à appliquer est l'inverse de la probabilité globale donnée par la formule (4). Il est donc :

$$\text{Coefft. pond.} = W_{ij} = \frac{\sum N_i}{aN_i} \cdot \frac{100}{S_{ij}} \cdot \frac{H_{ij}}{c} \quad (5)$$

où  $a$  = le nombre de fokontany tirés :  $a_u = 154$  } U/R SOND  
 $a_r = 166$  }

$N_i$  = la population administrative du fokontany  $i$  telle qu'elle figure dans la base de sondage

$\sum N_i$  = la somme des  $N_i$  sur la population des fokontany :

$$\sum N_i \text{ urb} = 2.488.925 )$$

$$\sum N_i \text{ rur} = 10.030.757 ) \quad \text{U/R SOND}$$

$S_{ij}$  = le pourcentage estimé de la population du fokontany  $i$  qui réside dans le segment  $ij$ .

$c$  = le nombre de ménages enquêtés dans le segment  $ij$ .

$$\text{Normalement } c_u = 12 )$$

$$c_r = 16 ) \quad \text{U/R OP}$$

mais en cas de non réponse on utilisera l'effectif réellement enquêté dans le segment plutôt que le nombre visé.

Le poids  $W_{ij}$  de chaque segment peut être calculé dès qu'on disposera des résultats du dénombrement et qu'on saura le nombre de ménages effectivement enquêtés dans le segment,  $c_{ij}$ .

Attention : Bien que les programmes standards (packages) prévoient normalement les pondérations éventuelles, leur utilisation doit être écartée en ce qui concerne certains tests de signification (notamment chi-2, signification des corrélations, ...)

## 8. Stratification

La stratification a pour but de réduire l'erreur de sondage en assurant que le taux prévu de sondage est effectivement réalisé dans chaque sous-ensemble de la population qu'on veut désigner. Ces sous-ensembles doivent être spécifiés à l'avance (des strates) et l'échantillon sera alors tiré indépendamment dans chaque strate. Avec un tirage aléatoire non stratifié la représentativité dans ce sens est laissée au hasard.

Dans notre cas actuel les milieux urbain et rural (U/R SOND) constituent les deux grandes strates explicites au 1er degré de sondage.

Supposons une région à l'intérieur de laquelle le taux de sondage visé est constant. On voit facilement que si on classe les unités dans la base de sondage selon les strates et qu'on tire un échantillon à un intervalle fixe (tirage dit systématique) on atteindra le même objectif que celui de la stratification, aux arrondis près. Si on connaît peu de choses sur la distribution des caractéristiques des unités cette procédure est convenable et efficace. Dans le même ordre d'idées, on peut classer les unités aréolaires au sein de chaque strate explicite selon leur contiguïté géographique. En général les zones voisines se ressemblent, de manière à ce qu'une progression géographique correspond très souvent à une tendance systématique d'ordre agro-économique et/ou socio-économique. En tirant un échantillon systématique d'une liste ainsi classée on a forte chance de profiter de l'existence des telles tendances même si on ne connaît pas leur nature exacte. Un effet analogue à la stratification sera ainsi produit.

Une telle stratégie est souvent utilisée dans des enquêtes. Dans le cas actuel nous l'avons appliquée de la façon suivante :

### Milieu urbain

- On a classé les villes selon leur taille
- On se propose de classer les fokontany à l'intérieur de chaque ville dans deux, trois ou quatre classes socio-économiques
- On ordonnera les fokontany selon ces catégories
- On tirera un échantillon systématique proportionnellement aux tailles des unités

### Milieu rural

- On a classé les firaïsan selon une progression continue basée sur la contiguïté. (L'ordre des fokontany à l'intérieur des firaïsan n'a pas d'intérêt car on ne tirera jamais deux fokontany d'un même firaïsan.)
- On a tiré un échantillon systématique proportionnellement aux tailles des unités ainsi ordonnées.

Ces mesures devraient normalement avoir pour effet une légère diminution de l'erreur de sondage.

## 9. Modalités de tirage à chaque degré

### 1er degré

Après avoir classé les unités primaires on doit cumuler les tailles.

On calculera ensuite la série de tirage. Le pas de tirage, R, est déjà connu (page 14). La série débute avec un nombre aléatoire A choisi entre 1 et R. Ensuite on ajoute R à chaque terme. On obtient la série :

$$A : \quad A + R : \quad A + 2R : \quad A + 3R : \quad \dots$$

Pour chaque terme de cette série on cherchera le premier cumul qui égale ou dépasse ce terme. L'unité qui correspond à ce terme est celui qui est tirée.

### 2ème degré

La procédure suivante devrait être appliquée à l'intérieur de chacun des fokontany i qui a été divisé en des segments.

- Ecrire la liste des segments, numérotés  $j = 1, 2, \dots$
- Inscrire l'estimation  $S_{ij}$  en face de chaque segment
- Vérifier que la somme sur j est égale à 100% dans le fokontany.
- Cumuler les  $S_{ij}$  sur j.
- Trouver un nombre aléatoire A entre 01 et 100 (chercher parmi les nombres à 2 chiffres : compter 00 comme 100).
- Le segment tiré est celui dont le cumul est le premier qui égale ou dépasse A.
- Renouveler le choix de A pour chaque fokontany.

### 3ème degré

On tirera un surplus de ménages afin de fournir une liste de réserve (voir chapitre 11). Soit  $c'$  le nombre total à tirer par segment - chiffre à fixer à l'avance.

On dispose de la liste des ménages fournie par l'agent dénombreur relative au segment ij.

- Numérotar les ménages dans le segment i de 1 à  $M_{ij}$ .
- Calculer le pas de tirage R en divisant  $M_{ij}$  par  $c'$ , à une décimale près.
- Chercher un nombre aléatoire A entre 1 et R.
- Tirer les ménages dont le numéro correspond à la partie entière de chaque terme de la série :

$$A : \quad A + R : \quad A + 2R : \quad A + 3R : \quad \dots$$

Exemple. Nous supposons  $c^i = 20$ ,  $M_{1j} = 68$

On calcule  $R = \frac{68}{20} = 3,4$

Chercher un nombre aléatoire entre 01 et 34. Supposons qu'on choisit 21. Alors  $A = 2,1$

La série est la suivante :

2,1 : 5,5 : 8,9 : 12,3 : 15,7 : ...

Les ménages tirés sont ceux numérotés :

2            5            8            12            15            ...

#### 10. Affectation des fokontany aux équipes et aux cycles

L'allocation des fokontany de l'échantillon parmi les équipes exige un examen détaillé de la situation exacte de chaque fokontany sur la carte de Madagascar ainsi que des moyens de transport utilisables. On appliquera les principes suivants :

- Dans la mesure du possible les équipes devraient être spécialisées dans la méthode urbaine ou rurale.
- On cherchera à minimiser les déplacements, en n'oubliant pas que l'équipe devra passer à son siège entre chaque paire de fokontany visités consécutivement.

L'affectation des fokontany aux cycles suivra à peu près les mêmes principes mais on devra faire intervenir un facteur simultanément :

- On doit éviter l'introduction des biais qui résulteraient des variations systématiques du contenu de l'échantillon liées aux saisons. Notamment, on doit veiller à ce que le rapport urbain/rural soit approximativement le même dans chaque cycle. Même principe en ce qui concerne les rapports nord/sud et est/ouest.

On déterminera ainsi un itinéraire précis pour chaque équipe en indiquant les dates exactes de son arrivée et départ dans chaque fokontany.

#### 11. Non réponse

Les non réponses sont constituées des ménages dénombrés mais absents au moment de l'enquête, des ménages qui sont partis définitivement, des ménages incapables (maladies, décès ...), des refus, et de quelques autres rares catégories. Les deux premières catégories seront presque certainement les plus fréquentes.

Il est évident que l'importance du problème dépendra surtout de la méthode choisie pour effectuer le dénombrement (voir chapitre 5). Avec la méthode A on peut prévoir plus de 10% de non réponses ; avec C on n'aura peut-être que 2 ou 3%.

Le schéma adopté pour le travail des enquêteurs exige un nombre exact de ménages à enquêter. On devra donc prévoir une méthode de remplacement pour les ménages défaillants. Il importe cependant d'éviter de l'enquêteur le choix de remplacement. Ainsi nous proposons l'établissement d'une

liste de ménages remplaçants (liste de réserve) dans chaque fokontany. Un enquêteur qui considère qu'il ne peut faire une interview programmée doit le signaler à son contrôleur qui seul est habilité à autoriser un remplacement. Le ménage remplaçant doit être toujours le premier qui est disponible sur la liste de réserve.

Une stratégie a été appliquée dans presque toutes les enquêtes dans des pays en voie de développement qui contribue largement à une réduction de la non réponse. Au cas où un ménage a quitté le logement dans lequel il a été dénombré, et qu'un autre ménage est venu résider dans le même logement, on enquêtera automatiquement ce nouveau ménage sans même parler de "remplacement". Cette solution trouve sa justification dans le raisonnement suivant : le dénombrement est en réalité un dénombrement de logements et pas de menages. Il est vrai que le nom du ménage figure sur la liste du dénombreur, mais celui-ci est cité à titre d'adresse, afin d'aider l'équipe à repérer le logement, car il n'y a pas d'adresses dans le sens exact. L'unité de sondage est donc le logement, et l'enquêteur est appelé à enquêter le ménage qui réside dans cette unité à la date de son passage. Vu de cette manière l'échantillonnage est tout aussi cohérent, et beaucoup plus stable, qui si on adoptait le modèle qui désigne le ménage même comme l'unité de sondage.

Compte tenu de cette importante définition, et notamment du fait que des tels cas ne doivent pas être classés comme des remplacements, les organisateurs devront fixer à l'avance le paramètre c' (voir chapitre 9) qui indique le nombre total de ménages à tirer par fokontany, y compris la liste de réserve.

Lorsqu'on a tiré les c' ménages on tirera un sous-échantillon systématique au nombre de c'-c ménages parmi ceux-ci, qui constituera la liste de réserve.

## 12. Problèmes non résolus

Le plus grand problème qui reste à résoudre est celui de la dispersion des équipes et de la distance (en km. ou en heures) entre les zones échantillons et le siège de l'équipe correspondante.

Plusieurs mesures peuvent être envisagées pour combler ce problème. En premier lieu il est certain qu'il faudra prévoir plus de 6 lieux-sièges (le nombre initialement prévu). En supposant un siège pour chaque équipe rurale et un pour deux équipes à Antananarivo on pourrait aussi envisager de déplacer certains sièges à certaines périodes de l'année afin de les rapprocher à leurs équipes. Dans certains cas on pourrait peut-être demander aux enquêteurs de passer directement d'une zone à une autre sans qu'ils passent par le siège : le chauffeur transporterait alors les questionnaires au siège tout seul, un ou deux jours après. Enfin dans des cas exceptionnellement difficiles on pourrait emprunter des jours prévus pour les congés afin d'augmenter le temps disponible pour le transfert de l'équipe entre zones. Toutes ces stratégies sont à étudier.

Enfin l'autre problème important restant à résoudre est celui de l'organisation du dénombrement. Celui-ci a déjà été examiné au chapitre 5. Il n'y a pas de solution idéale : il s'agit de trouver la moins mauvaise.