

# ENQUETE PRIORITAIRE A MADAGASCAR

## Plan de sondage proposé

Rapport du consultant Chris Scott, Novembre 1996

### 1. CADRE

L'Enquête Prioritaire auprès des ménages (EP) à Madagascar fait partie d'une série d'enquêtes statistiques exécutées par la Direction des Statistiques des Ménages (DSM) au sein de l'Institut National de la Statistique (INSTAT). Ces enquêtes s'insèrent dans le programme du Projet d'Appui Institutionnel à la Gestion Publique (PAIGEP) à Madagascar, parrainé par la Banque Mondiale. Elles s'inspirent d'ailleurs des plans d'enquête proposés par le projet Dimensions Sociales de l'Ajustement au sein de la Banque Mondiale.

Ce programme d'enquêtes est constitué de deux éléments. Le premier, appelé à Madagascar "Enquête Permanente auprès des Ménages (EPM)" (ou "enquête intégrale" dans certains autres pays) vise un échantillon assez modeste, de 4 à 5 mille ménages, mais avec un questionnaire très détaillé exigeant plusieurs visites à chaque ménage, les travaux de terrain étant étendus sur une année entière. Cette enquête a été effectuée à Madagascar en 1993. Le deuxième type d'enquête est "l'enquête prioritaire". Celle-ci est constituée d'une enquête de suivie : elle serait répétée en principe chaque année en visant un échantillon de ménages environ 2 fois plus grand que l'EPM. L'enquête durerait de 3 à 4 mois sur le terrain moyennant un questionnaire beaucoup plus simple avec une seule visite de l'enquêteur à chaque ménage. La première enquête de ce genre à Madagascar est proposée pour le deuxième trimestre de 1997. On se propose de la répéter tous les 2 ans.

Le présent rapport a pour objectif de proposer un plan de sondage en vue de l'Enquête Prioritaire de 1997.

### 2. FACTEURS DETERMINANT DU PLAN DE SONDRAGE

Trois types de contraintes déterminent le choix du plan de sondage de l'EP : celles qui traduisent les objectifs spécifiques de l'enquête, celles qui résultent des moyens logistiques adoptés pour son exécution, et enfin celles qui résultent du cadre institutionnel dans lequel le projet se situe.

En ce qui concerne les objectifs, il y en a un qui est déterminant quant au plan de sondage : la taille ainsi que la répartition de l'échantillon doit permettre une analyse adéquate de l'impact de la politique de restructuration sur chacun d'une vingtaine de groupes socio-économiques bien définis. Or l'expérience de multiples enquêtes africaines indique qu'un échantillon d'environ 400 ménages constitue un minimum pour une sous-population ainsi visée. En élaborant le plan de sondage la première étape consistera donc à spécifier ces sous-populations visées pour l'analyse. Celles-ci seront appelées *groupes socio-économiques* ou GSE. On déterminera par la suite les taux de sondage par strate de façon à assurer l'échantillon minimal requis pour chaque GSE.

Quant aux moyens logistiques, les principales contraintes sont les suivantes. L'enquête serait effectuée par 14 équipes, chacune constituée d'un contrôleur et de 3 enquêteurs. Au moins 8 des

équipes seraient munies d'une véhicule tout-terrain. Pour plus de détail, voir le dernier chapitre de ce rapport. Un dénombrement des ménages dans chaque aire de sondage sélectionnée sera nécessaire. Celui-ci sera organisé pour précéder immédiatement les entrevues dans chaque aire. Ce sera donc l'équipe des enquêteurs qui devront assumer cette tâche dès leur arrivée sur le lieu de l'enquête. A partir des listes de ménages ainsi dressées le contrôleur de l'équipe tirera un échantillon de 25 ménages à enquêter. Le questionnaire proposé étant assez long on peut supposer que l'enquêteur moyen traitera de 3 ménages par jour. On verra plus tard que ces dispositions impliquent une durée totale de l'ordre de 4 mois pour l'enquête de 8.500 ménages.

Enfin l'influence du cadre institutionnel dans lequel se situe le projet se fait senti dans l'obligation de construire l'échantillon sur la base de *l'échantillon maître*. Cet outil a été élaboré par la Direction des Statistiques Sociales et Démographiques au sein de l'INSTAT sur la base des résultats du recensement national de 1993. Il a été conçu dans l'hypothèse que toute enquête nationale auprès des ménages, pour les 10 années à venir, tirerait son échantillon à partir de l'échantillon maître. Or la Direction des Statistiques des Ménages, responsable de la présente enquête, et la Direction des Statistiques Sociales et Démographiques, responsable de l'échantillon maître, dépendent toutes les deux de l'INSTAT. L'obligation de profiter de cet outil pour la constitution de l'échantillon de la présente enquête est donc quasiment absolue. Nous donnons dans le chapitre suivant un bref résumé de l'échantillon maître tel qu'il a été constitué pour Madagascar et tel qu'il existe jusqu'ici, car les travaux de sa constitution ne sont pas encore complets.

### **3. L'ECHANTILLON MAITRE**

L'échantillon maître couvre l'ensemble du territoire national à l'exception des îles de Nosy-Bé et Sainte-Marie.

L'échantillon est basé sur le recensement de 1993, l'unité de base étant la zone de dénombrement (ZD) dont la taille est connue en termes du nombre de ménages dénombrés lors de ce recensement.

Les unités primaires (UPS) sont des ZD ou des groupes de ZD. On a groupé certaines ZD avec des ZD voisines afin de constituer des UPS d'une taille plus homogène, mais dans des tels cas l'utilisateur est libre à choisir la ZD comme unité primaire. Les UPS sont tirés avec probabilités proportionnelles à la taille ("PPT") et une ZD a été tirée avec PPT dans chaque UPS. Dans ces circonstances l'utilisateur peut demander un échantillon d'UPS (unités très homogènes) ou un échantillon de ZD (unités moins homogènes de taille, mais en moyenne plus petites). La dernière option réduira la taille de l'opération de dénombrement, parce que les ZD sont moins grandes que les UPS, et nous avons choisi cette option pour la présente enquête. Nous aurons donc un échantillon à deux degrés : ZD au 1<sup>er</sup> degré, fournies par l'échantillon maître, et ménages au 2<sup>e</sup> degré, tirés sur le terrain par le contrôleur de chaque équipe.

L'échantillon maître est stratifié en 3 grandes strates explicites : grands centres urbains (GCU), centres secondaires urbains (CUS), et rural. Le taux de sondage au niveau aréolaire dans le secteur urbain (GCU + CUS) est environ 2,7 fois plus élevé que dans le secteur rural. En plus il y a des légères variations des taux entre faritany. Voir le tableau 4 au chapitre 6 pour les plus amples détails.

Dans la première grande strate on a également appliqué une stratification au 1<sup>er</sup> degré de sondage selon des indicateurs basés sur des données socio-économiques extraites du recensement au

niveau des ZD. A part cette stratification explicite, une stratification implicite a été appliquée partout par moyen du tirage systématique sur une liste ordonnée géographiquement.

Faute d'un financement suffisant, le travail de construction de l'échantillon maître s'est arrêté avant d'être complété. Dans de nombreux cas (environ 8% dans l'échantillon maître) le croquis de la ZD a disparu : dans d'autres cas le croquis n'a pas été mis à jour lors des travaux du recensement de 1993. Jusqu'ici aucune démarche n'a été tentée en vue de réparer ces lacunes. Le consultant Megill a recommandé que les divers travaux de mise à jour nécessaires soient pris en charge par la première enquête cliente de l'échantillon maître. On proposait initialement d'ailleurs de scinder des ZD trop grandes en des segments (aires d'enquête, ou AE) et on a même déterminé dans ce but le nombre d'AE à créer dans chaque ZD. Mais ici encore, le manque de financement n'a pas permis la moindre démarche sur le terrain. Autant que ces lacunes subsistent, on pourrait se demander si l'existence de l'échantillon maître représente un avantage pour les organisateurs des enquêtes.

Le problème des multiples utilisateurs de l'échantillon maître sera traité au chapitre 8 plus bas.

En ce qui concerne la construction de l'échantillon pour l'EP, la principale contrainte imposée par l'échantillon maître est celle de la taille totale de l'échantillon. Il paraîtra plus loin qu'on a dû modifier le plan de sondage en milieu urbain en fonction de cette contrainte.

## **4. REPRESENTATION DES GROUPES SOCIO-ECONOMIQUES**

### **4.1. Définition des GSE : principes généraux**

Le choix des GSE résulte d'une négociation entre analyste, sondeur et financier. 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 permette à l'analyste d'étudier avec une profondeur suffisante l'impact de la politique de restructuration sur un GSE donné. La contrainte fondamentale étant le financement, et de ce fait la taille de l'échantillon globale (dans notre cas entre 8000 et 9000 ménages), l'analyste souhaiterait étudier, dans cette enveloppe, le nombre maximum de GSE qu'on puisse inclure.

De prime abord on pourrait imaginer que ce nombre serait obtenu en divisant 8500 par 400, soit 21. Mais ce serait faux pour deux raisons. En premier lieu il y aura peut-être des GSE qui chevauchent entre eux. Cela pourrait permettre de satisfaire les conditions imposées pour un nombre plus grand de GSE simultanément. Mais il y a un autre facteur qui joue dans le sens inverse. Le sondeur ne sera pas en mesure d'identifier les GSE avec précision au moment du sondage. On verra plus loin que les GSE qui intéressent les analystes seront définis en fonction d'un double critère, en partie géographique et en partie socio-économique. On voudrait étudier, par exemple, les ménages *résidant dans la capitale* et en même temps *rémunérés par le secteur publique* : ou bien les ménages *habitant le faritany de Mahajanga* et qui *cultivent un produit d'exportation*. 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. C'est ainsi que chaque strate géographique doit être dotée d'un taux unique de sondage qu'on devra appliquer 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. Il en résulte qu'il tirera bien plus de 400 ménages pour les classes autres que la plus rare dans la strate. Bref, en exigeant un minimum de 400 ménages pour tous les GSE, on obligera le sondeur à tirer beaucoup plus que cela pour certains GSE.

## 4.2- Choix des GSE

Pour se fixer une base de travail l'analyste choisira d'abord un découpage géographique et ensuite un classement socio-économique. A la suite des discussions entre les intéressés il a été proposé que la division primordiale devrait être entre les grandes villes et le restant du pays. On regroupera ainsi les Centres Urbains Secondaires (CUS) avec le rural et au sein de ce grand secteur on distinguera les faritany (provinces). Dans le secteur des grandes villes (GCU) on distinguera la capitale des autres. Cela nous amène au classement suivant.

### Découpage géographique

GRANDS CENTRES URBAINS : Ville de Antananarivo  
Autres GCU

CUS + RURAL : Faritany de..... Antananarivo  
Fianarantsoa  
Toamasina  
Mahajanga  
Toliary  
Antsiranana

En ce qui concerne le classement socio-économique on propose une base de classification totalement différente entre le secteur des GCU et celui du CUS + Rural.

Dans le secteur urbain l'intérêt principale est de distinguer entre les secteurs publique et privé, et au sein du privé de distinguer entre les secteurs formel et informel. Nous proposons donc le découpage suivant pour les GCU.

### Découpage socio-économique : secteur des GCU

Classement en fonction de l'activité principale du chef du ménage :

- (1) Secteur public
- (2) Secteur privé formel
- (3) Secteur informel
- (4) Aides familiaux etc.

La dernière classe consiste essentiellement du groupe " autres ", ce qui implique tous ceux qui ne sont pas rémunérés. Comme la grande majorité sont des aides familiaux nous les avons caractérisés ainsi pour rappeler au lecteur la nature essentielle du groupe.

Il convient d'ailleurs de rappeler qu'un nombre important des ménages du secteur informel (3) sont des agriculteurs, même en milieu GCU.

Enfin il importe de noter que les catégories sont *exclusives* : il n'y pas de chevauchement entre elles.

### Découpage économique : secteur CUS + Rural

La grande majorité des ménages de ce secteur ont leur activité principale dans l'agriculture. Si on cherche un classement basé sur des catégories exclusives on trouvera très peu d'activité dans des

autres domaines. Nous avons identifiés environ 10% seulement de ménages ruraux dont l'activité principale n'est ni agricole ni pastorale.

On sera donc obligé de construire la principale classification en fonction des activités agro-pastorale en admettant des catégories en chevauchement mutuel.

Les statistiques fiables manquent sur la distribution des superficies cultivées par ménage. On cherchera plutôt un classement en fonction de la nature de l'activité.

D'après l'Enquête Permanente auprès des Ménages, environ 90% des ménages cultivent un ou plusieurs produits vivriers. Parmi ceux-ci la grande majorité s'occupe également de l'élevage. La culture des produits d'exportation ou des cultures industrielles est moins fréquente mais suffisamment répandue pour qu'elle soit ciblée dans l'échantillon. Nous proposons donc pour un premier essai le classement suivant :

- (1) Cultivateurs de cultures vivrières
- (2) Cultivateurs de cultures d'exportation
- (3) Cultivateurs de cultures industrielles
- (4) Eleveurs
- (5) Autres (sans activités agricoles ou pastorales)

Toutes ces catégories se chevauchent sauf la dernière, qui est exclusive.

On aimerait, si possible, croiser chacune de ces catégories par faritany dans un tableau à deux entrées. Un examen rapide des effectifs concernés révèle que cela ne présentera aucun problème quant aux catégories (1) et (5). En ce qui concerne les autres catégories, leurs effectifs en milliers de ménages par faritany sont estimés comme suit :

	<b>Antananarivo</b>	<b>Finanrantsoa</b>	<b>Toamasina</b>	<b>Mahajanga</b>	<b>Toliary</b>	<b>Antsiranana</b>	<b>TOT CUS+Rural</b>
Cultures d'export.	4	152	195	24	3	127	505
Cultures industr	161	143	243	100	66	77	790
Autres	67	33	16	41	40	16	213

On verra plus loin qu'un effectif inférieur à 200.000, si on voulait le cibler comme un GSE tout seul, nous obligerait d'augmenter la taille de l'échantillon par un facteur excessif de manière à ce qu'on déborderait l'enveloppe totale qui nous est imposée. Cela suggère la possibilité qu'on réunisse les deux catégories dans une seule qui serait "cultivateurs d'une ou plusieurs cultures de rente". Celle-ci ne serait pas la somme des deux premières lignes du tableau car il y a chevauchement entre ces deux catégories. Les données estimées sont les suivantes :

Cultures de rente	161	244	286	103	69	149	1001
-------------------	-----	-----	-----	-----	----	-----	------

Nous voyons que le problème n'est pas entièrement résolu : il restent des cases excessivement petites (inférieur à 200.000). Il a été donc proposé de regrouper dans un seul GSE les faritany de Antananarivo et Mahajanga pour cette variable, ainsi que les faritany de Toliary et Antsiranana. On arrive ainsi à distinguer seulement 4 GSE pour cette variable.

Enfin, en ce qui concerne la catégorie “ Autres ”, les effectifs sont si faibles que la seule solution est de cibler cette catégorie sur l’ensemble de Madagascar, sans aucune ventilation géographique.

#### **4.3- Estimation de la taille des GSE**

Afin de mettre au point le plan de l’échantillonnage il faudra estimer d’abord la taille de chaque GSE. Cela se fera ici en consultant les données du recensement de 1993 et celles de l’EPM. Ces sources se concordent suffisamment dans le secteur CUS+Rural, mais dans le secteur des GCU il y a des incohérences considérables. Notre stratégie pour les réconcilier sera de prendre le recensement comme cadre au niveau des principales catégories géographiques et d’y affecter les distributions en pourcentages trouvées pour les différentes sous-populations socio-économiques. Cela nous donnera les valeurs estimées des N (ou  $N_{gh}$ ) dans les tableaux 1 et 2 qui suivent.

### **5. REPARTITION DE L’ECHANTILLON**

Nous pouvons maintenant passer à des calculs détaillés sur la répartition de l’échantillon. Ceux-ci figurent dans les tableaux 1 (urbain) et 2 (rural). Rappelons que pour les besoins de la présente enquête les CUS sont regroupés avec le secteur rural et l’expression “ urbain ” se rapporte aux grandes villes, appelées GCU.

Les tableaux sont structurés afin de 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. Chaque case de chaque tableau représente un GSE potentiel mais on pourra regrouper les cases, soit horizontalement soit verticalement, pour constituer les GSE plus convenables. Un premier regroupement de ce genre a été déjà décrit plus haut : verticalement en regroupant deux catégories de cultivateurs et horizontalement en regroupant des faritany pour une seule variable. On a sorti cet exemple de sa bonne place dans l’exposition afin de simplifier la présentation. Le tableau 2 suppose ces regroupements déjà effectués.

L’indice **h** désigne la strate géographique et l’indice **g** la caractéristique socio-économique.

Dans chacune des cases de chaque tableau 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 des sources citées plus haut : recensement et EPM.

Après avoir inscrit les  $N_{gh}$  on passe aux  $n_{gh}$ . On inscrira 400 pour le total de chaque GSE. Certains GSE sont constitués de plusieurs cases et dans ces cas on inscrira les 400 dans la colonne du total correspondant à cet ensemble. Il a fallu prévoir les colonnes ou lignes de totaux nécessaires à cette fin.

On passera ensuite aux  $F_{gh}$ . Pour chacune des cases où on voit  $n = 400$  on calculera  $F_{gh} = n_{gh} / N_{gh}$  et on inscrira cette quantité en dessous du  $n$  correspondant. Là où les 400 figurent dans une colonne ou une ligne *total* (c'est-à-dire là où le GSE s'étend sur plusieurs cases) on recopiera la valeur  $F_{gh}$  calculée déjà sur ce total, dans chacune des cases constituant le GSE. Par exemple, dans la ligne des 'Autres' au tableau 2 le taux  $F_{gh} = 1,875$ , calculé initialement sur le total des faritany, doit être porté dans chacune des colonnes des 6 faritany. En multipliant ces valeurs de  $F$  par les  $N$  correspondant dans la case on complètera les  $n$  pour ces cases ( $n = 126$  ;  $n = 77$  ;  $n = 52$  ; etc. à la ligne "Autres" du tableau 2).

**TABLEAU 1**

**REPARTITION DE L'ECHANTILLON**

**URBAIN**

(1)		(2)	(3)	(4)
Catégories Socio-économiques <sup>1</sup>		CAPITALE	AUTRES GCU	TOTAL GCU
Secteur	N	29.090	26.020	55.110
Public	n	400	400	
	F	13,75	15,37	
Privé	N	36.540	36.750	73.290
Formel	n	400	400	
	F	10,95	10,88	
Informel (y compris agriculture)	N	44.150	40.270	84.420
	n	400	400	
	F	9,06	9,93	
Aides Familiales etc	N	37.140	26.400	63.540
	n	400	400	
	F	10,77	15,15	

TOTAL	N	146.920	129.440	276.360
	Fmax	13,75	15,37	
	n	2.020	1.989	4.009
	Arrondi : n' =	2.100	2.000	4.100
	=	25 X 84	25 X 80	

N = nombre de ménages existants

n = nombre de ménages à sélectionner

$F = 1000n/N$

**TABLEAU 2**

**REPARTITION DE L'ECHANTILLON : RURAL + CUS**

g	h	(1) ANTANANARIVO	(4) MAHAJANGA	TOTAL (1) + (4)	(2) FIANARANTSOA	(3) TOAMASIN A	(5) TOLIARY	(7) ANTSIRANANA	TOTAL (5) + (7)	TOTAL RURAL et CUS
	<b>GSE<sup>2</sup></b>									
Cultivateurs	N	474	261		421	363	289	194		2.002
Cultures	n	400	400		400	400	400	400		
Vivrières	F									

<sup>1</sup> Catégories socio-économiques basées sur activité principale du chef du ménage.

<sup>2</sup> Les 3 premières catégories se chevauchent, alors que la quatrième ne chevauche pas avec les autres. Les données N : ENSEMBLE ne sont donc pas égales à la somme des N sur les 4 catégories.

		0,844	1,533		0,950	1,102	1,384	2,062		
Cultivateurs	N	161	103	264	244	286	69	149	218	1.011
Cultures de	n	244	156	400	400	400	127	273	400	
Rente	F	1,515	1,515	1,515	1,639	1,399	1,835	1,835	1,835	
				5					5	
Eleveurs	N	424	229		342	343	234	156		1.728
	n	400	400		400	400	400	400		
	F	0,943	1,747		1,170	1,166	1,709	2,564		
Autres <sup>3</sup>	N	67	41		33	16	40	16		213
	n	126	77		52	30	75	30		400
	F	1,875	1,875		1,875	1,875	1,875	1,875		1,875

#### ENSEMBLE

N	541	305	455	380	330
213					
Fmax	1,875	1,875	1,875	1,875	1,875
2,564					
n	1.014	572	853	713	619
	4.317				546
Arrondi : n' =	1.025	575	850	725	625
	4.350				550
=	25 X 41	25 X 23	25 X 34	25 X 29	25 X 25
25 X 22	25 X 174				

N = nombre de ménages existants en milliers

n = nombre de ménages à sélectionner, toujours 400 pour les GSE ciblés

F = n/N

Enfin on fera le total  $N_h = S_g N_{gh}$  pour chaque colonne en l'inscrivant en dessous du tableau. On cherchera alors la plus grande fraction de sondage,  $F_{max_h}$  dans chaque colonne h. C'est le taux de sondage qu'il faudra appliquer à l'ensemble de la strate h car il assure l'échantillon voulu pour le plus petit des éléments trouvés dans la colonne. En multipliant la somme  $N_h$  par  $F_{max_h}$  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.

En étudiant les totaux en dessous des tableaux et à droite on verra que l'échantillon d'ensemble s'élève à 4009 ménages urbains plus 4350 ruraux, soit un total de 8359. Quelques petits mises au point reste à faire pour que ces totaux soit conformes aux exigences des travaux de terrain (25 ménages sélectionnés par ZD tirée). On arrive enfin à un total de 8.450 ménages, ce qui concorde très bien avec la taille prévue pour l'enquête.

Nous avons eu de la chance. Si cela en avait abouti autrement on aurait dû procéder à une modification du choix des GSE.

Sur la base de ces calculs on peut donner au tableau 3 la liste des 25 GSE que nous avons ciblés avec succès, ainsi que le nombre de ménages que nous pouvons espérer trouver pour chaque GSE dans notre échantillon.

Enfin il importe de noter qu'en visant certains GSE explicitement on n'exclue pas les analyses sur d'autres catégories. En effet toute catégorie dont l'échantillon dépasse 400 peut être considérée a priori comme analysable avec à peu près la même précision que celles des GSE ciblés. D'ailleurs

<sup>3</sup> La catégorie " Autres " est constituée des ménages n'ayant aucune activité agricole ou pastorale.



d'autres analyses peuvent être envisagées sur des échantillons inférieurs à 400 dans des cas éventuels où une précision moins rigoureuse peut être admise.

### **TABLEAU 3**

#### **LISTE DES GSE CIBLES**

Dans cette liste les ménages de l'échantillon sont classés en fonction du lieu de résidence et l'activité principale de leur chef .

<b>URBAIN</b>		Nombre estimé de ménages dans l'échantillon
1. Secteur public	Capitale	416
2.	Autres grands centres urbains	402
3. Secteur privé formel	Capitale	522
4.	Autres GCU	568
5. Secteur informel (y compris l'agriculture)	Capitale	631
6.	Autres GCU	622
7. Aides familiales etc	Capitale	531
8.	Autres GCU	408
<b>ECHANTILLON TOTAL URBAIN</b>		<b>4100</b>

#### **RURAL et CENTRES URBAINS SECONDAIRES**

Cultivateurs de cultures vivrières, par faritany :

9.	Antananarivo	889
10.	Fianarantsoa	789
11.	Toamasina	681
12.	Mahajanga	489
13.	Toliary	542
14.	Antsiranana	497

Cultivateurs de cultures de rente :

15.	Antananarivo + Mahajanga	495
16.	Fianarantsoa	456
17.	Toamasina	536
18.	Toliary + Antsiranana	511

Eleveurs, par faritany :

19.	Antananarivo	795
20.	Fianarantsoa	641
21.	Toamasina	643
22.	Mahajanga	429
23.	Toliary	439
24.	Antsiranana	400
25. Résidents du secteur rural + CUS sans activité agricole ou pastorale	Madagascar	400

**ECHANTILLON TOTAL RURAL + CUS 4350**

## 6. MODALITES DU TIRAGE DE L'ECHANTILLON

### 6.1- Principes

Nous donnons d'abord un bref aperçu de la théorie qui est supposée par notre plan de sondage. Pour l'instant on mettra de côté le fait que l'échantillon devra être sélectionné au sein d'un échantillon maître.

L'échantillon de ZD sera tiré au sein de chaque strate avec des probabilités proportionnelles à la taille de chaque ZD, cette taille étant mesurée en fonction du nombre de ménages recensés dans la ZD. Soit  $M_i$  ce nombre pour la ZD  $i$  et soit  $M_i'$  l'effectif de ménages enregistrés lors du dénombrement qui précède l'enquête dans la ZD. 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.

Le tirage au 1<sup>er</sup> degré se fera en sélectionnant des ZD à un intervalle fixe, égal à  $sM_i/a$ , dans la colonne des tailles  $M_i$  cumulées, où la somme  $s$  se rapporte à la population totale des ZD et  $a$  est l'effectif de ZD tirées dans la strate. Il est évident que la probabilité pour une ZD donnée d'être tirée est égale au rapport entre la taille de la ZD et le pas de sondage. Mathématiquement, on aura :

$$p_{1i} = a M_i / sM_i$$

où on entend par  $p_{1i}$  la probabilité au 1<sup>er</sup> degré pour l'unité  $i$  dans une strate donnée.

Au 2<sup>ème</sup> degré de sondage on tirera un nombre fixe de  $b$  ménages dans chaque ZD à partir des  $M_i'$  ménages dénombrés dans cette ZD. La probabilité pour qu'un ménage donné soit tiré dans une ZD tirée donnée est donc :

$$p_{2i} = b / M_i'$$

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

$$F_i = p_{1i} p_{2i} = b (a / sM_i) . (M_i / M_i')$$

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. Dans la pratique cependant, puisque ces deux quantités diffèrent, il faudra pondérer les données lors de l'analyse. D'ailleurs dans la pratique il y se passera parfois qu'on ne réussit pas à compléter les 25 ménages échantillon dans certaines ZD : la quantité  $b$  deviendra ainsi variable et il convient de la désigner par  $b_i$ . Les coefficients de pondération qui s'imposent en conséquence seront :

$$W_i = 1 / F_i = (sM_i / a) . (M_i' / M_i) / b_i$$

Voir le chapitre 7 pour les détails plus amples sur les pondérations.

## 6.2- Application

Les paramètres de sondage sont implicites dans les résultats qui figurent en bas des tableaux 1 et 2. La ligne **n** nous donne le nombre de ménages souhaités dans l'échantillon. Avant de mettre au point la répartition de l'échantillon on s'est proposé de fixer le nombre de ménages à enquêter à 20 par grappe (**b = 20**). Or en divisant les valeurs de **n** en dessous du tableau 1 par 20 on obtiendrait 101 grappes (ZD) pour la capitale et encore 100 pour les autres GCU. C'est ici qu'on se heurte à la contrainte qu'impose l'échantillon maître (EM). Le tableau 4 montre les effectifs de ZD tirées dans l'EM selon les strates. On constatera que l'EM ne fournit pas le nombre voulu d'unités primaires. Le maximum qu'on puisse trouver de ZD de l'EM dans les 2 strates urbaines citées est respectivement 84 et 80.

Plutôt que de sortir du cadre de l'EM il nous a paru plus convenable d'augmenter le paramètre **b** de 20 à 25. En ajustant légèrement les chiffres relatifs à la taille de l'échantillon voulue on trouve que la valeur **b = 25** nous permet d'atteindre nos objectifs pour ces deux strates en supposant qu'on prenne pour l'EP *la totalité des ZD de l'EM*. Ces résultats figurent à la dernière ligne en dessous du tableau 1. La taille de l'échantillon urbain, ainsi ajustée, devient 4100 ménages pour les deux strates ensemble.

Passons au cas du secteur CUS+RURAL. Il convient d'adopter la même valeur **b = 25** pour ce secteur. La dernière ligne en dessous du tableau 2 montre le nombre de ZD à tirer par faritany. En rapprochant ces données avec celles du tableau 4 on verra que l'EM est largement suffisant pour nos besoins dans ce secteur. On souhaiterait tirer 174 ZD sur un ensemble de 596 disponibles dans l'EM, environ 30%. Les modalités précises de ce sous-échantillonnage sont à étudier soigneusement.

Les ZD de l'EM sont choisies avec PPT. L'échantillon de ZD visé pour l'EP est également basé sur le principe PPT. Par conséquent, le sous-échantillonnage des ZD qui doit nous amener de l'EM à l'EP devrait être fait avec *probabilité constante* au sein de chaque strate. Désignons par  $p_{sh}$  le taux de sous-échantillonnage requis dans la strate **h**. Si  $a_h$  est le nombre de ZD dans l'EM dans cette strate et que  $a_h'$  est le nombre de ZD souhaité pour l'EP dans la strate, nous savons que ce taux sera le rapport du nombre à tirer pour l'EP sur le nombre existant dans l'EM, d'où

$$p_{sh} = a_h' / a_h$$

Les  $a_h$  figurent au tableau 4. Les  $a_h'$  cependant ne peuvent être lus directement dans le tableau 2 car la stratification de l'EM diffère de celle de l'EP. L'EM en particulier distingue les ZD du secteur CUS, en les affectant un taux de sondage bien plus élevé que les ZD du secteur rural. On doit donc d'abord ventiler les effectifs  $a_h'$  en fonction des deux groupes CUS et Rural. Cela s'est fait dans le tableau 4.

Dans ce tableau nous partons des effectifs totaux des ZD requis dans l'échantillon de l'EP secteur CUS + Rural, chiffres qui viennent du tableau 2 dernière ligne. Ces nombres sont ensuite ventilés entre les deux strates CUS et Rural en proportion avec les effectifs de ménages existant dans la population, selon les chiffres de la colonne (2) du tableau 4.

Ensuite en comparant les effectifs du sous-échantillon avec ceux de l'EM, par strate, on obtient les coefficients de pondération  $W_{sh} = 1 / p_{sh}$  à appliquer pour tenir compte du sous-échantillonnage.

**TABLEAU 4**

Paramètres de sondage appliqués pour tirer l'échantillon maître par tirage systématique avec probabilités proportionnelles à la taille.

(1)	(2)	(3)	(4)	(5)
<b>Strate H</b>	<b>Mesure de taille Somme des <math>M_{hi}</math> dans la base de sondage</b>	<b>Nombre d'unités tirées <math>n_h</math></b>	<b>Intervalle de tirage <math>I_h</math></b>	<b>Nombre aléatoire pour débiter le tirage <math>R_h</math></b>
<b>GCU</b>				
(111) Antananarivo	145576	84	1733.05	655.18
(112) Antsirabe	25150	16	1571.88	401.85
(211) Fianarantsoa	20434	12	1702.83	1627.17
(311) Toamasina	29016	16	1813.50	808.41
(411) Mahajanaga	24224	16	1514.00	1092.29
(511) Toliary	15573	8	1946.63	354.08
(711) Antsiranana	15052	12	1254.33	923.71
	275025	164		
<b>AUTRE URBAIN</b>				
(110) Antananarivo	48663	28	1737.96	1388.75
(210) Fianarantsoa	60645	36	1684.58	1487.93
(310) Toamasina	60410	36	1678.06	684.09
(410) Mahajanaga	37545	20	1877.25	1799.14
(510) Toliary	66344	40	1658.60	1324.59
(710) Antsiranana	24354	16	1522.13	1258.84
	297961	176		
<b>RURAL</b>				
(120) Antananarivo	485768	104	4670.85	3732.32
(220) Fianarantsoa	405367	88	4606.44	4068.70
(320) Toamasina	324318	72	4504.42	1836.31
(420) Mahajanaga	228722	48	4765.04	4566.77
(520) Toliary	306913	68	4765.04	3604.50
(720) Antsiranana	179174	40	4479.35	3704.54
	1930262	420		

**Source** : Rapport du Consultant David Megill, Juillet 1996.

**TABLEAU 5****SOUS-ECHANTILLONAGE DE L'ECHANTILLON MAITRE  
SECTEUR CUS + RURAL**

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Faritany	Nombre de ZD à tirer			Nombre de ZD dans l'EM		Pondération à prévoir	
	TOTAL	CUS	RURAL	CUS	RURAL	CUS	RURAL
Antananarivo	41	4	37	28	104	7	2,811
Fianarantsoa	34	4	30	36	88	9	2,933
Toamasina	29	5	24	36	72	7,2	3
Mahajanga	23	3	20	20	48	6,667	2,4
Toliary	25	4	21	40	68	10	3,238
Antsiranana	22	3	19	16	40	5,333	2,105
---	---	---	---	---	---		
	174	23	151	176	420		

**6.3- Précisions sur les modalités de tirage**

Le tirage de l'échantillon est donc effectué en 3 étapes.

- 1) Tirage d'un ensemble appelé "Echantillon Maître". Il s'agit d'un échantillon de ZD tirées avec PPT par la Direction des Statistiques Sociales et Démographiques.
- 2) Tirage d'un sous-échantillon de l'échantillon maître. Il s'agit d'un échantillon de ZD tirées parmi les ZD de l'EM avec probabilité constante au sein de chaque strate mais variable entre strates. Dans le secteur des GCU il n'y a pas de sous-échantillonnage : taux = 1.
- 3) Tirage de 25 ménages au sein de chaque ZD. Tirage systématique (à probabilité égale) à partir de la liste des ménages établie lors du dénombrement.

En ce qui concerne les modalités, pour l'étape (1) tout a été fait déjà par le service concerné. Pour l'étape (2) le tirage reste à faire seulement dans les secteurs CSU et Rural. Les ZD seront tirés par sondage systématique. L'intervalle à utiliser dans chaque strate est égal au rapport porté dans la colonne (8) du tableau 5. On commencera avec un nombre aléatoire entre 1 et la valeur de l'intervalle.

Quant à l'étape (3) on espère enquêter 25 ménages partout mais on en tirera 30 pour disposer d'une réserve pour le remplacement éventuel des ménages défaillants. Le tirage des 30 se fera par sondage systématique en calculant le pas par la formule  $M_i' / 30$ , la valeur de  $M_i'$  étant fournie par le comptage des ménages dénombrés. Après le dénombrement et le tirage le contrôleur notera R (= réserve) à côté de chaque 6<sup>ème</sup> ménage tiré. Les autres seront alors renumérotés de 1 à 25 et distribuer aux enquêteurs pour interview. Seul le contrôleur sera habilité à autoriser l'utilisation d'un ménage de remplacement.

Le pas de sondage pour le tirage des ménages sera souvent un chiffre petit, en dessous de 5 par exemple. Dans cette situation les erreurs d'arrondi peuvent être important si on suit la méthode du

tirage systématique classique. Une méthode simple pour éviter ce problème est d'appliquer le sondage systématique dit "circulaire". Dans cette méthode le nombre aléatoire pour débiter le tirage est choisi entre 1 et  $M_i'$ , au lieu de entre 1 et le pas. On procède comme dans la méthode classique sauf que, lorsqu'on arrive à la fin de la liste, on continue sans interruption en sautant au début de la liste, l'intervalle étant maintenu comme si la liste était circulaire. On s'arrête dès qu'on a tiré l'effectif souhaité de ménages. Il peut arriver exceptionnellement que cette méthode nous amène à re-sélectionner un ménage tiré vers le début de l'opération. Dans un tel cas le sondeur se permet de déplacer la sélection à l'unité suivante.

#### 6.4- Précisions sur la pondération

Le coefficient de pondération qu'on doit appliquer aux données de chaque ZD est constitué de 3 facteurs, correspondant aux 3 étapes du sondage.

Le 1<sup>er</sup> facteur est associé avec le tirage de la ZD dans l'EM. Le coefficient est égal à

$$W_{E,hi} = S_{Mhi} / (a_{hi} M_{hi})$$

et s'applique à la ZD  $i$  de la strate  $h$ , la somme  $S$  s'étendant sur l'ensemble des unités  $i$  de la strate, qu'elles soient sélectionnées ou non.

Le 2<sup>ème</sup> facteur est associé avec le sous-échantillonnage. Sa valeur  $W_{S,h}$  se trouve dans la colonne (8) du tableau 5. Pour le secteur des GCU on aura  $W_S = 1$ .

Enfin le 3<sup>ème</sup> facteur représente le tirage des ménages, et englobe l'effet de la non réponse. Si  $b_{hi}$  est le nombre de ménages retenus dans le fichier de l'enquête (normalement 25, mais peut-être pas toujours), on aura

$$W_{M,hi} = M_{hi}' / b_{hi}$$

La pondération globale se fait en multipliant ensemble ces trois coefficients :

$$W_{hi} = W_{E,hi} \cdot W_{S,h} \cdot W_{LM,hi}$$

Les données qui constituent cette expression devront être retenues dans le fichier de l'enquête, en association avec chaque ZD.

### 7. PARAMETRES DES TRAVAUX DE TERRAIN : RECAPITULATION

Les différentes options ont été discutées en détail avec le personnel technique affecté à l'EP. En voici les conclusions, qui devraient être considérées comme provisoires jusqu'à ce que l'enquête pilote ait été achevée.

- 1) Le dénombrement se fera normalement au fur et à la mesure des visites pour les interviews. Exceptionnellement, on devra faire le dénombrement plus tôt, là où une visite préliminaire de reconnaissance paraît nécessaire.

- 2) On prévoit un maximum de 2 heures par interview.
- 3) On prévoit une moyenne de 3 interviews par jour travaillé et par enquêteur.
- 4) On prévoit 2 jours pour le déplacement d'une équipe entre ZD rurales ou entre villes, mais un temps zéro pour le déplacement entre ZD urbaines à l'intérieur d'une même ville.
- (5) On prévoit 14 équipes chacune constituée d'un contrôleur et de 3 enquêteurs. Environ 8 pourront être munis de voitures.
- (6) Les équipes travailleront 6 jours par semaine.

## **8. LE PROBLEME DES MULTIPLES UTILISATEURS DE L'ECHANTILLON**

Il semble probable que trois organisations seront utilisateurs de l'échantillon maître pendant une période assez restreinte. Cela présente des problèmes et des avantages.

### Avantages

Le dénombrement des ménages dans des ZD pourront servir à plusieurs enquêtes, si on organise le chevauchement des échantillons.

La mise au point des cartes et documents effectuée par une organisation servira aux besoins des autres.

### Inconvénients

Si plusieurs enquêtes cherchent simultanément des renseignements d'un même ménage cela risque de créer des résistances auprès des enquêtés.

On pourrait même s'attendre à une réticence de la part des autorités locales s'il se considèrent "sur-enquêtées".

### Solutions

En général il paraîtrait souhaitable de répéter les mêmes ZD pour les enquêtes différentes, mais pas les mêmes ménages. Il conviendrait donc que les organisations concernées communiquent leurs projets entr'elles et se consultent en ce qui concerne les unités tirées. En particulier chacun doit essayer d'éviter de tirer les ménages déjà enquêtés récemment. Une gestion rigoureusement paraît un élément indispensable pour bien résoudre les problèmes potentiels.