

Le 19 Mars 2001

## **ENQUETE PERMANENTE AUPRES DES MENAGES A MADAGASCAR**

### **PLAN DE SONDAGE PROPOSE**

Mamadou Thiam

#### **1. Introduction**

L'Enquête Permanente auprès des Ménages de 2001 (EPM-2001) fait suite à celles réalisées en 1993, 1997 et 1999 à Madagascar. Cette série d'enquêtes s'inscrit dans le cadre du projet Dimension Sociale de l'Ajustement parrainé par la Banque Mondiale. Comme pour les enquêtes précédentes, l'EPM-2001 a pour principal objectif de collecter des données socio-économiques telles le revenu et les dépenses des ménages, l'emploi, l'éducation, l'agriculture et la santé. Les résultats de l'enquête seront présentés pour chacune des 6 régions du pays, pour certains groupes socio-économiques cibles (GSE) et pour les milieux de résidence à savoir la Capitale, les Autres Grands Centres Urbains (AGCU), les Centres Urbains Secondaires (CUS), et le milieu rural. Comme dans les autres enquêtes, les 2 îles de Sainte Marie et Nosy Be ne sont pas couvertes par la présente enquête.

Le présent rapport décrit la méthodologie de l'échantillonnage développée en collaboration avec la Direction des Statistiques des Ménages (DSM) de l'Institut National de la Statistique (INSTAT).

#### **2. Base de sondage**

Le dernier recensement général de la population et de l'habitat à Madagascar date de 1993. Cependant, en 1996 la Direction de la Démographie et des Statistiques Sociales (DDSS) de l'INSTAT a constitué un échantillon maître<sup>1</sup> sur la base de ce recensement. L'échantillon maître est un échantillon aréolaire, stratifié et tiré à 2 degrés. L'unité primaire de sondage (UPS) correspond soit à la zone de dénombrement telle que définie au recensement de 1993, soit à un groupe de zones de dénombrement. Trois grandes strates ont été constituées: les grands centres urbains, les centres urbains secondaires et le milieu rural. La première strate a été stratifiée davantage par classes socio-économiques. Dans toutes les strates, les UPS ont été également classées par ordre géographique.

Au premier degré, les UPS ont été tirées avec une probabilité proportionnelle au nombre de ménages recensés en 1993. Les UPS constituées de plusieurs zones de dénombrement ont été découpées en segments dont un seul a été retenu dans l'échantillon maître au deuxième degré. Comme pour l'EPM-1999, l'échantillon maître servira également de base de sondage pour la présente enquête.

Le tableau 1 donne la répartition des 760 UPS de l'échantillon maître.

---

<sup>1</sup> Recommandations préliminaires pour la conception de l'échantillon maître pour le programme d'enquêtes auprès des ménages à Madagascar, David J. Megill, 1996, U.S Bureau of the Census

**Tableau 1:** Distribution des UPS de l'échantillon maître

Région	Capitale	AGCU	CUS	Rural	Total
Antananarivo	84	16	28	104	232
Fianarantsoa		12	36	88	136
Toamasina		16	36	72	124
Mahajanga		16	20	48	84
Toliara		8	40	68	116
Antsiranana		12	16	40	68
Madagascar	84	80	176	420	760

### 3. Détermination de la taille de l'échantillon

La taille de l'échantillon dépend des indicateurs à mesurer, du niveau de précision désiré pour les estimations, du nombre des domaines d'études, et des contraintes budgétaires. Compte tenu de l'expérience des enquêtes précédentes et du budget pour l'EPM-2001, la DSM a signalé que la taille de l'échantillon pour l'EPM-2001 ne pourrait pas dépasser 6000 ménages. Pour déterminer, la taille de l'échantillon, un découpage géographique basé sur le milieu de résidence et un regroupement des GSE à partir des résultats de l'EPM-1999 ont été utilisés.

Dans la capitale et les AGCU, 2 groupes de GSE ont été formés: les salariés (regroupant les cadres salariés, employés, ouvriers, et main d'œuvre sans qualification) et les autres (non salariés). Selon les résultats de l'EPM-1999, les salariés représentent respectivement 46,6% et 44% des ménages dans la capitale et les AGCU.

En milieu rural, l'EPM-1999 a estimé à 79,7% la proportion des ménages dont l'activité principale du chef se trouve dans le secteur primaire (agriculture, élevage, pêche et chasse).

Dans les CUS, 3 grands groupes de GSE se dégagent de l'EPM-1999: les salariés (30,9%), le secteur primaire (agriculture, élevage, pêche et chasse) avec 38,7% et les autres (30,4%).

D'autres regroupements de GSE à partir des résultats de l'EPM-1999 ont également été étudiés avec l'équipe de la DSM. Comme nous le verrons plus loin, les regroupements ci-dessus ont été retenus car ils permettent d'obtenir une taille d'échantillon ne dépassant la limite des 6000 ménages signalée par la DSM.

En utilisant les projections de population faites par l'INSTAT pour 2001, la taille moyenne des ménages (5,1) et la répartition en pourcentage des ménages par milieu de résidence selon l'EPM-1999, il a été possible d'obtenir une répartition du nombre estimé de ménages par milieu de résidence.

**Tableau 2:** Distribution de la population et des ménages par région

Région	Projections de population pour 2001 <sup>2</sup>	Nombre estimé de ménages en 2001
Antananarivo	4585000	899020
Antsiranana	1189000	233137
Fianarantsoa	3369000	660588
Mahajanga	1734000	340000
Toamasina	2595000	508824
Toliara	2230000	437255
Madagascar	15702000	3078824

**Tableau 3:** Distribution des ménages par milieu de résidence

Milieu de résidence	Ménages selon l'EPM-1999	Nombre estimé de ménages en 2001
Capitale	5.8%	178572
AGCU	4.7%	144705
CUS	12.9%	397168
Rural	76.6%	2358379
Madagascar	100.0%	3078824

L'expérience des enquêtes du même type a montré que pour faire des analyses fiables pour une GSE il faut au moins 400 ménages dans la GSE. Ainsi, la taille finale de l'échantillon est affectée par le choix et le nombre des GSE. Une rencontre avec le directeur de la DSM et son équipe a permis de retenir le regroupement des GSE mentionné ci-dessus parmi un certain nombre de regroupements de GSE qui ont été étudiés sur la base des résultats de l'EPM-1999.

En croisant le milieu de résidence avec le regroupement de GSE retenu, un taux de sondage permettant d'obtenir 400 ménages pour chaque GSE ou groupe de GSE a été utilisé séparément à l'intérieur de chaque milieu de résidence. Ce taux de sondage correspond au taux de sondage maximum pour le milieu de résidence. Les tableaux 4, 5 et 6 présentent par milieu de résidence le calcul de la taille de l'échantillon. Pour ces tableaux, les notations sont:

- N : nombre de ménages estimé pour 2001
- f : taux de sondage permettant d'obtenir 400 ménages pour le GSE
- $f_{\max}$  : taux de sondage maximum pour le milieu de résidence
- n : nombre de ménages obtenu en appliquant  $f_{\max}$  au nombre total de ménages existants dans le milieu de résidence.

<sup>2</sup> Source: INSTAT, Projections et Perspectives Démographiques, Rapport d'Analyse, Volume 2, Tome VIII, Janvier 1997

Le nombre de grappes est obtenu en divisant le nombre total de ménages par le nombre de ménages à sélectionner par grappe. Comme pour l'EPM-1999, 16 ménages seront sélectionnés par grappe urbaine (capitale, AGCU, et CUS) et 18 ménages par grappe rurale.

Tableau 4: Calcul de la taille pour la capitale et AGCU

GSE	Capitale	AGCU
Salariés	N = 83215 f = 400/83215	63670 400/63670
Autres	N = 95357 f = 400/95357	81035 400/81035
	$f_{\max} =$	0.0048068
Total	N = 178572 n = 858	144705 909
Nombre de ménages (arrondi)	864	912
Nombre de grappes	54	57

Tableau 5: Calcul de la taille pour les CUS

GSE	CUS
Salariés	N = 122725 f = 400/122725
Secteur primaire	N = 153704 f = 400/153704
Autres	N = 120739 f = 400/120739
	$f_{\max} =$
Total	N = 397168 n = 1316
Nombre de ménages (arrondi)	1312
Nombre de grappes	82

Tableau 6: Calcul de la taille pour le milieu rural

GSE	Rural
Secteur primaire	N = 1879628 f = 400/1879628
Autres	N = 478751 f = 400/478751
	$f_{\max} =$
Total	N = 2358379 n = 1970

Nombre de ménages (arrondi)	1980
Nombre de grappes	110

Cette taille d'échantillon de 5068 ménages est dans les limites fixées par les contraintes budgétaires, et reste également très proche de la taille de l'échantillon de l'EPM-1999 (5120 ménages). Le tableau 7 donne la répartition finale de l'échantillon par région et milieu de résidence.

**Tableau 7:** Répartition finale de l'échantillon de ménages et grappes

Région	Capitale		AGCU		CUS		Rural		Total	
	Grappes	Ménages	Grappes	Ménages	Grappes	Ménages	Grappes	Ménages	Grappes	Ménages
Antananarivo	54	864	11	176	16	256	21	378	102	1674
Fianarantsoa			10	160	15	240	20	360	45	760
Toamasina			10	160	14	224	19	342	43	726
Mahajanga			9	144	13	208	17	306	39	658
Toliara			8	128	14	224	18	324	40	676
Antsiranana			8	128	11	176	15	270	34	574
Madagascar		864	56	896	83	1328	110	1980	303	5068

#### 4. Structure de l'échantillon

L'échantillon de l'EPM-2001 est un échantillon aréolaire, stratifié, et tiré à 2 degrés. L'unité primaire de sondage est l'UPS telle que définie dans l'échantillon maître. Au premier degré 303 UPS seront sélectionnées avec la même probabilité car les UPS de l'échantillon maître ont été sélectionnées avec une probabilité proportionnelle au nombre de ménages recensés en 1993. Un dénombrement des ménages dans chaque UPS sélectionnée permettra d'obtenir une liste de ménages à partir de laquelle des ménages seront sélectionnés au deuxième degré (16 ménages par grappe urbaine et 18 par grappe rurale). Avant le dénombrement, les UPS ayant un grand nombre de ménages seront divisées en plusieurs segments dont un seul sera retenu pour l'EPM-2001.

#### 5. Tirage des UPS et des ménages

##### Tirage des UPS

L'échantillon sera tiré indépendamment pour chaque milieu de résidence à l'intérieur d'une région selon la méthode du tirage systématique décrite ci-après.

- 1) Classer d'abord les UPS de l'échantillon maître. Dans les AGCU, les UPS seront classées par ordre géographique et par classe socio-économique. Dans les CUS, les villes seront d'abord classées selon la taille de la ville, et dans chaque ville les UPS seront ensuite classées par ordre géographique. En milieu rural, les UPS seront classées par ordre géographique.

- 2) Calculer I, le pas de tirage des UPS. I est obtenu en divisant le nombre total d'UPS existantes dans le milieu de résidence de la région par le nombre correspondant de UPS à sélectionner. Arrondir la valeur de I à l'entier le plus près.
- 3) Choisir un nombre aléatoire R compris entre 1 et le pas de tirage I calculé en 2).
- 4) Les UPS sélectionnées sont celles qui ont pour rang dans la liste ordonnée: R, R + I, R + 2I, R + 3I, R + 4I, etc.

### Tirage des ménages

A la suite du dénombrement des ménages, 16 et 18 ménages seront tirés respectivement par UPS urbaine et rurale. Pour tenir compte des non-réponses, il y aura par UPS 2 ménages de réserve qui seront utilisés pour remplacer les ménages ne pouvant être enquêtés.

Dans la capitale, les AGCU et CUS, on sélectionnera d'abord 18 ménages par UPS en utilisant la méthode du tirage systématique. Ensuite, 2 ménages seront identifiés comme ménages de réserve de la manière suivante. Choisir un nombre aléatoire R compris entre 1 et 9. Sur la liste des 18 ménages déjà sélectionnés, les ménages ayant pour rang R et R + 9 seront choisis comme ménages de réserve.

En milieu rural, 20 ménages seront d'abord sélectionnés en utilisant le tirage systématique. Les 2 ménages de réserve seront sélectionnés de la même manière qu'en milieu urbain. Le nombre aléatoire R sera compris entre 1 et 10 et les ménages de rang R et R + 10 serviront de ménages de remplacement.

## **6. Segmentation des grandes UPS**

Certaines UPS sélectionnées contiennent un grand nombre de ménages et exigeraient un travail énorme si tous leurs ménages devaient être systématiquement dénombrés. Pour limiter le travail de terrain dans une UPS, chaque UPS ayant plus de 250 ménages sera divisée en 2 ou plusieurs segments dont un seul sera retenu dans l'échantillon. La règle de segmentation est la suivante:

Nombre de ménages	Nombre de segments à créer
0 – 250 ménages	0
251 – 500 ménages	2
501 – 750 ménages	3
751 – 1000 ménages	4
etc.	etc.

La procédure complète de segmentation est décrite dans le manuel de cartographie et de dénombrement des ménages.

## 7. Probabilités de sondage et coefficients de pondération

Les probabilités de sondage sont calculées pour chaque degré de tirage. Les notations sont les suivantes:

- $P_i$  : probabilité d'inclusion de la  $i^{\text{ème}}$  UPS dans l'échantillon maître
- $P_{1hi}$  : probabilité de sondage au premier degré de la  $i^{\text{ème}}$  UPS de la strate  $h$
- $P_{2hi}$  : probabilité de sondage au deuxième degré de la  $i^{\text{ème}}$  UPS de la strate  $h$

Soient  $a_h$  le nombre de UPS tirées dans la strate  $h$ ,  $A_h$  le nombre total de UPS dans la strate, et  $t_{hij}$  la taille estimée en proportion du segment  $j$  choisi pour la  $i^{\text{ème}}$  UPS de la strate  $h$ . On notera que  $t_{ihj}=1$  si l'UPS n'a pas été segmentée et la somme des  $t_{ihj}$  est égale à 1 pour chaque UPS.

Au premier degré, la probabilité d'inclusion de cette UPS dans l'échantillon est donnée par:

$$P_{1hi} = P_i \times \frac{a_h}{A_h} \times t_{ihj}$$

Au deuxième degré, un nombre  $b$  de ménages seront tirés à partir des  $L_{hi}$  ménages nouvellement dénombrés par l'équipe EPM-2001 dans la  $i^{\text{ème}}$  UPS de la strate  $h$  lors de l'opération de mise à jour des cartes. Donc :

$$P_{2hi} = \frac{b}{L_{hi}}$$

où  $b = 16$  pour l'ensemble du milieu urbain et  $b = 18$  pour le milieu rural.

Pour assurer la représentativité de l'échantillon au niveau national, l'utilisation de coefficients de pondération des données de l'EPM-2001 est nécessaire. La composante principale du coefficient de pondération est l'inverse du produit des probabilités de sondage aux 2 degrés. Elle est calculée pour chaque grappe selon la formule suivante:

$$W_{hi} = \frac{A_h \times L_{hi}}{P_i \times a_h \times t_{ij} \times b}$$

Compte tenu du fait que certains ménages identifiés ne seraient pas enquêtés, cette composante principale doit être ajustée avec le taux de réponse des ménages pour obtenir le coefficient final de pondération des données selon la formule suivante:

$$W_{hi}^f = W_{hi} \times \frac{n_h}{n_{\cdot}}$$

où  $n_h$  est le nombre total de ménages sélectionnés dans la strate  $h$ , et  $n'_h$  est le nombre de ménages avec une interview complète dans la strate  $h$ .

Tous les ménages d'une même grappe seront ainsi affectés du même coefficient de pondération. Un fichier Excel permettant de calculer les coefficients de pondération des données de l'enquête sera construit ultérieurement.