

ENQUETE PRIORITAIRE 98

Plan de sondage proposé : Juillet 1998

1. CADRE

Dans le cadre du suivi et de l'évaluation des conditions de vie des populations une série d'enquêtes auprès des ménages sont prévues. Il s'agit des enquêtes Prioritaire et Intégrale. L'Enquête Prioritaire (EP) est une enquête de suivie, d'un contenu assez léger et basée sur un échantillon relativement important (entre 7 et 8 mille ménages). L'exécution de la première enquête de ce type a été confiée à l'ISTEEBU sur financement de la Banque Mondiale à travers le projet TWITEZIMBERE. Les travaux de collecte dureront quatre mois (Août-Novembre) et couvrent tout le territoire national. Le présent document est une proposition d'un plan de sondage qui est en réalité une adaptation du plan de sondage proposé par Chris SCOTT lors de sa mission effectuée au BURUNDI en Avril 1993.

2. FACTEURS DETERMINANTS DU PLAN DE SONDAGE

Deux types de contraintes déterminent le choix du plan de sondage de l'EP: celles qui traduisent les objectifs spécifiques de l'enquête et celles qui résultent des moyens logistiques choisis pour son exécution.

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 dizaine de groupes socio-économiques bien définis. Or l'expérience des autres enquêtes africaines indique qu'un échantillon d'environ 400 ménages est nécessaire pour chaque sous-population ainsi visée. En préparant 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 sera effectuée par 13 équipes (6 en Mairie de Bujumbura et 7 dans le reste du pays), chacune composée de 4 enquêteurs, d'un anthropomètre et un contrôleur. Chaque équipe du milieu rural sera munie d'un véhicule tandis qu'en mairie de Bujumbura les équipes seront déplacées au moyen de mini-bus. Un dénombrement des ménages sera organisé dans chaque aire de sondage sélectionnée. Cette opération sera complétée dans l'ensemble de l'échantillon avant que l'enquête proprement dite ne commence (la veille du jour de l'interview). A partir des listes de ménages ainsi dressées, on tirera dans chaque aire un échantillon de 10 ménages à enquêter en milieu rural et 20 ménages en milieu urbain. L'enquête proprement dite durera environ 4 mois.

3. BASES DE SONDAGE

Il convient de distinguer les deux cas: base de sondage aréolaire et base de sondage pour les ménages.

Au niveau aréolaire on cherchera un ensemble d'unités qui couvrent la totalité de la surface burundaise sans lacune et sans chevauchement. Les limites de chaque unité doivent être bien définies et faciles à repérer. A condition de ne pas compromettre cette dernière condition il est préférable que les unités soient d'une taille aussi petite que possible, surtout parce que les unités moins grandes impliquent un travail moindre au niveau du dénombrement des ménages. On aura intérêt en plus à ce que les unités ne varient pas excessivement de taille et d'ailleurs à ce que cette taille (nombre de ménages) soit connue à l'avance pour chaque unité afin qu'on puisse en tenir compte lors du tirage de l'échantillon. Enfin, il serait bon que la liste des unités qui constitue la base de sondage comporte des données descriptives de chaque unité permettant la stratification de la base en fonction de caractéristiques significatives.

Le découpage du territoire national utilisé pour le recensement mené en août 1990 offre en effet deux ensembles d'unités qui correspondent assez bien aux exigences qu'on vient de citer: les sous-collines (SC) et les zones de dénombrement (ZD). Examinons de plus près ce découpage.

Au premier niveau on a 15 provinces plus la Mairie de Bujumbura.

Au deuxième niveau, chaque province est divisée en communes, appelées zones dans la ville de Bujumbura. Le nombre total de communes est de 116 et le nombre de zones à Bujumbura-Mairie est de 9.

Au troisième niveau, les communes se divisent en collines et les zones de Bujumbura-Mairie se divisent en quartiers. Il y a environ 2446 collines (nombre de collines actuellement habitées) et 90 quartiers.

Enfin les collines sont divisées en sous-collines, dont il y en a approximativement 8304 (soit en moyenne un peu plus de 3 par colline).

A un troisième niveau parallèle, les communes ont été divisées, pour les seuls besoins du recensement de 1990, en zones de dénombrement. Il en est de même pour les zones de Bujumbura-Mairie, celles-ci étant également divisées en îlots. Le nombre total de ZD était de 4685, dont 198 dans la ville de Bujumbura. En dehors de Bujumbura-Mairie, les ZD étaient dans la grande majorité des cas constituées de petits groupes de 1, 2 ou 3 sous-collines.

Ce système couvre le territoire national en entier, aux seules exceptions des deux parcs nationaux de la Ruvubu et de la Kibira, qui n'abritent en principe aucune population humaine.

Que devons-nous alors choisir comme bases de sondage? Analysons cela cas par cas.

3.1 Base de sondage en milieu rural.

Les limites de toutes les sous-collines sont connues avec exactitude tandis qu'une tentative de localiser les limites des ZD se baseraient sur des cartes confectionnées lors de la préparation du recensement général de la population et de l'habitat d'Août 1990 (il y a 9 ans). Ces cartes semblent être actuellement d'une utilité très limitée vue les mutations socio-politiques que vit le pays depuis bientôt cinq ans et les travaux de cartographie qui couvriraient tout le pays sont loin d'envisagés dans le cadre de cette étude. Suite aux différentes restructurations administratives opérées depuis 1990 et surtout depuis le début de la crise socio-politique d'Octobre 1993, de nouvelles collines et sous-collines ont été créées, des sous-collines ont été fusionnées et d'autres ne sont plus habitées de telle sorte qu'il est quasiment impossible de

reconstituer les ZD telles qu'elles étaient lors du recensement.

Par contre, les travaux de la mise à jour de la base de sondage de Mars 1998 ont abouti à la constitution d'un fichier de toutes les collines et sous-collines que compte le pays (celles actuellement habitées) ainsi que le nombre de ménages pour chaque sous-collines. Cette base comporte 2446 collines et 8304 sous collines.

Qu'en est-il de la taille et de l'homogénéité des sous collines?

Il est évident que les sous-collines sont de petites tailles par rapport aux ZD et le nombre de ménages par sous-colline est bien connu, ce qui constitue un grand avantage (nombre moyen de ménages par sous colline est de 118). Bien que les ZD soient, par construction, plus homogènes que les SC, la variabilité de taille des sous-collines n'est pas aussi grande qu'on le croyait (la moyenne calculée sur toutes les sous-collines est de 118 ménages avec un écart-type de 88,28 soit un coefficient de variation 0,75). Cette variabilité peut être d'ailleurs corrigée en appliquant la méthode de segmentation pour les sous-collines de taille jugée trop grande. En outre les sous-collines de taille trop petite seront jugées comme abérrantes et exclues de la base de sondage.

Compte tenu de tout ce qui précède, nous pensons qu'il est plus judicieux de prendre la SC comme unité de sondage.

3.2 Base de sondage en mairie de Bujumbura

On prendra comme base de sondage en mairie de Bujumbura les résultats des travaux de cartographie récemment effectués dans le cadre de l'actualisation de la carte censitaire de la ville de Bujumbura. Ces travaux ont abouti à une liste de 177 ZD. Ainsi, l'unité de sondage en mairie de Bujumbura est la ZD.

En ce qui concerne la base de sondage au niveau des ménages il s'agit de l'opération du dénombrement des ménages déjà cité.

En milieu rural et dans les petites villes les agents devront passer de rigo en rigo pour enregistrer les noms des chefs de ménage la veille du jour de l'interview. Afin d'assurer un dénombrement complet il faudra qu'un agent

dénombrer soit accompagné par le chef de colline ou par son délégué.

A Bujumbura-Mairie, et peut-être dans certains centres urbains de l'intérieur du pays, l'unité de listing sera en premier lieu la parcelle. L'agent dénombrer passera dans chaque parcelle dans l'aire qui lui a été affectée, accompagné du chef de 10 parcelles ou de chef de rue, en demandant le nom de chaque ménage habitant la parcelle.

4. AU NIVEAU AREOLAIRE, UN SEUL DEGRE DE SONDAGE OU DEUX?

Si nous admettons le choix de la Sous-Colline (SC) comme l'unité principale de sondage aréolaire on peut toutefois se demander si on n'aurait pas intérêt à introduire un degré de sondage supplémentaire soit avant soit après la SC.

L'introduction d'une unité de sondage supérieure à la SC aurait deux conséquences significatives: on réduirait quelque peu la distance moyenne entre les SC voisines dans l'échantillon et on augmenterait légèrement l'erreur de sondage. Or en sélectionnant environ 452 SC réparties sur l'ensemble du Burundi (Bujumbura-Mairie exclue) on trouvera que la distance entre les SC-échantillons est déjà très réduite et le gain en la réduisant davantage serait négligeable. Il semble donc préférable de maintenir la SC comme unité primaire de sondage.

On introduira, selon le cas, une unité aréolaire inférieure à la SC en créant exprès des segments à l'intérieur des SC tirées afin d'en tirer un dans chaque SC (pour les SC de Plus 200 ménages). La taille minimale pour chaque segment est fixée à 100 ménages.

Nous resterons donc avec la SC comme la seule unité aréolaire de sondage à employer avec éventuellement possibilité de segmenter certaines SC.

5. LES GROUPES SOCIO-ECONOMIQUES

5.1 Définition des groupes socio-économiques: principes généraux

Le choix des GSE résulte d'une négociation entre analyste et sondeur. Rappelons d'abord qu'il s'agit de trouver une répartition de l'échantillon qui assurera un minimum de 400 ménages pour chaque GSE, ce chiffre étant

considéré comme le minimum qui 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 la taille globale de l'échantillon (entre 7000 et 8000 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 8.000 (taille maximale admissible pour l'échantillon) par 400, soit 20. Mais ce serait faux car il y a une autre contrainte. Le sondeur ne sera pas en mesure d'identifier les GSE avec précision au moment du sondage. On verra dans le chapitre 5.2 que les GSE qui intéressent les analystes seront définis chacun en fonction de deux critères, dont l'un est géographique et l'autre socio-économique. Par exemple, on voudrait étudier les ménages résidant à Bujumbura et rémunérés par le secteur public, ou bien les ménages de la zone des plaines 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 sera dotée d'un taux unique 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. C'est ainsi que, en exigeant un minimum de 400 ménages pour tous les GSE on sera amené à en tirer beaucoup plus que cela pour certains GSE.

5.2 Choix des GSE

Pour se fixer une base de travail l'analyste choisira d'abord un découpage géographique et un classement socio-économique. Une première proposition à cet effet est donnée en bas. A noter que ce schéma reste provisoire jusqu'à-ce-qu'on ait démontré qu'il permet de construire un échantillon qui satisfait aux conditions imposées.

Découpage géographique

BUJUMBURA-Mairie

RESTANT DU BURUNDI: Zones des plaines
Montagnes et zone de transition
Plateaux occidentaux
Plateaux orientaux

Classement socio-économique

Pour Bujumbura: Secteur public
(Mairie) Autre formel
Secteur informel
Autres

Pour le restant
du Burundi: Cultivateurs, culture d'exportation
Cultivateurs, autres
Non cultivateurs

En croisant ces deux classifications on aurait en principe $4 + 3 \times 4 = 16$ groupes. Avant de voir si ce schéma peut donner un échantillon admissible, cherchons à raffiner un peu les définitions.

Classement Buja/Reste. Il s'agit de la Mairie de Bujumbura, ville capitale qui englobe 4,45% de la population nationale. La deuxième ville étant 10 fois plus petite, et la totalité des villes autres que Bujumbura constituant moins de 2% de la population, il semble approprié de créer les GSE urbains uniquement dans la ville de Bujumbura.

Grandes zones agro-écologiques. Nous avons cherché un ensemble de 4 grandes zones pour distinguer les régions agro-écologiques.

Un découpage du Burundi en 11 "régions naturelles" a été proposé dans une étude de l'Institut des Sciences Agronomiques du Burundi intitulée Contribution à la connaissance des régions naturelles du Burundi (ISABU, juillet 1992). Ce même découpage a été appliqué par l'ISTEEBU dans les analyses de son enquête sur le budget et la consommation des ménages ruraux (1986-1990). Pour les besoins de la présente enquête on propose un regroupement des 11 régions naturelles en 4 grandes zones, de la manière suivante:

Régions

ZONE I, Plaines: Bugesera, Buragane, Imbo, Moso
ZONE II, Montagnes et trnsition: Bututsi, Mugamba, Mumirwa
ZONE III, Plateaux occidentaux: Buyenzi, Kirimiro

ZONE IV, Plateaux orientaux:Buyogoma, Bweru

Si on se base ici strictement sur le découpage de ISABU les zones coupent parfois des limites des communes, d'ailleurs les zones ne tiennent pas compte des SC. Comme notre unité de sondage sera la SC cela présente une difficulté pour nous. Nous avons donc décidé de modifier les zones légèrement de manière à ce que toutes les zones soient constituées par des groupements de communes entières. Cet objectif a été atteint en affectant les communes coupées à la zone dans laquelle se trouve la plus grande fraction de sa population. Le classement définitif des communes en fonction des zones, adoptée pour la présente enquête, figure à l'Annexe 1 de ce rapport.

Classes socio-économiques. Celles-ci seront définies en fonction de l'emploi du chef du ménage. La première classe, secteur publique, serait normalement constituée de l'ensemble des agents de l'état (sous statut ou sous contrat). L'effectif total de ceux-ci à Bujumbura-Mairie est de 9305 (source: Fonction Publique) mais il n'y a aucun moyen de savoir combien de ces personnes sont des chefs de ménage. Ces effectifs ne sont guère suffisants pour qu'on les considère comme GSE en tant que tel. On propose d'y ajouter le groupe para-public qui est constitué de 7673 personnes. On a ainsi un total de 16978 personnes. D'après le recensement, si on considère l'ensemble des salariés (publics, para-public et privés) à Bujumbura-Mairie on trouve 45% seulement qui sont des chefs de ménages. En supposant que ce pourcentage s'applique au sein des secteurs public et para-public, on aura un total de 7640 ménages dans ce GSE.

Classement des cultivateurs. Les cultures d'exportation au Burundi sont le thé, le café et le coton. Ces cultures ne sont pas limitées à des grandes plantations: chacune peut être cultivée par des exploitants familiaux. Des données sur la répartition des cultures sont disponibles à partir de l'enquête budgets-consommation des ménages. Celles-ci montrent que la majorité des ménages agricoles au Burundi cultivent à la fois des cultures vivrières et une ou plusieurs des cultures d'exportation. On propose donc le classement suivant:

Cultivateurs: secteur exportation - ceux qui cultivent au moins une des 3 cultures citées plus haut, et

Cultivateurs: autres - ceux qui ne cultivent aucune de ces 3 cultures.

Il reste un petit groupe de non cultivateurs. Dans les données de l'EBC deux groupes de ce genre peuvent être identifiés dans le milieu rural: (A) Ceux qui ne citent pas une activité agricole comme leur activité économique principale et (B) Ceux qui ne cite l'agriculture ni comme activité principale ni comme activité secondaire. Les vrais non cultivateurs sont sans doute les (B); pourtant le groupe plus large (A) est probablement plus intéressant pour les analystes. Dans les tableaux qui suivront, le groupe (A-B), soit ceux qui citent l'agriculture comme activité secondaire, sera ajouté aux "cultivateurs autres" et le groupe (B) sera appelé "non cultivateurs".

On s'efforcera cependant de donner des renseignements relatifs à la taille de l'échantillon aussi bien pour A que pour B.

Les statistiques sur les nombres de ménages dans les différentes catégories socio-économiques en milieu rural proviendront de deux sources: les totaux par grande zone seront tirés du recensement et la répartition entre les catégories viendront de l'EBC.

Il est évident que certaines des données citées sont assez douteuses: il a fallu faire des estimations et des hypothèses qui sont loin d'être sûres. Par conséquent nous n'avons pas hésité à mélanger des données relatives à des dates différentes: autrement dit, nous faisons l'hypothèse, faute de mieux, que les caractéristiques observées dans les différentes enquêtes et recensements n'ont pas varié de façon notable au cours des années récentes. Dans la mesure où nous avons tort l'échantillon de l'enquête pourra se révéler insuffisant en matière de taille dans quelques domaines mais cela n'implique nullement un biais quelconque dans les conclusions tirés de l'enquête.

6. REPARTITION DE L'ECHANTILLON

Le tableau 1 présente un premier essai en vue d'une répartition de l'échantillon entre les différentes zones et catégories définies plus haut. La structure du tableau est conçue pour permettre au sondeur de voir l'impact précis sur l'échantillon des décisions prises relatives au choix des GSE. A partir de cela, il peut déterminer si une modification quelconque serait souhaitable afin d'obtenir un échantillon plus convenable. Les strates géographiques sont disposées en colonnes: les lignes correspondent aux classes socio-économiques. Chaque case du tableau représente un GSE potentiel mais on pourra regrouper les cases, soit horizontalement soit verticalement, pour constituer les GSE plus convenables. L'indice h désigne la strate et l'indice g la caractéristique socio-économique.

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

N_{gh} : le nombre de ménages de la population dans la strate h et possédant la caractéristique socio-économique g

n_{gh} : le nombre de ménages de l'échantillon dans la strate h et possédant la caractéristique socio-économique g

$F_{gh} = n_{gh}/N_{gh}$: le taux de sondage dans la strate h pour la catégorie socio-économique g

On commence avec les N . Ces quantités proviendront des diverses sources citées plus haut: recensement, enquêtes..

Après avoir inscrit les N_{gh} on passe aux n_{gh} . On inscrira 400 pour le total de chaque GSE. Pour cet essai tous les GSE sont constitués d'une case unique du tableau, sauf peut-être en ce qui concerne la catégorie non cultivateurs, catégorie qu'on doit laisser de côté pour l'instant.

Ensuite pour chacune des cases où on a inscrit $n = 400$ on calculera la quantité $F_{gh} = n_{gh}/N_{gh}$ et on inscrira cette quantité en dessous du n correspondant.

Enfin on fait le total $N_h = \sum_g N_{gh}$ pour chaque colonne en l'inscrivant en dessous du tableau. On cherche alors la plus grande fraction de sondage, $F_{\max, h}$ dans chaque colonne h

et on l'inscrit en dessous du N_h . C'est ce 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 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.

Tableau 1 REPARTITION DE L'ECHANTILLON

Nombre de ménages

Secteur	BUJUMBURA (Mairie)	R E S T A N T D U B U R U N D I			
		Plaines	Montagnes et trans.	Plateau	Plateau occid.
<u>oriental</u>					
Public et	N 6643				
para-	n 400				
<u>public</u>	F 60,21				
Privé	N 12769				
formel	n 400				
	F 31,32				
	N 16776				
Informel	n 400				
	F 23,84				
	N 8218				
Autres	n 400				
	F 48,67				
Cultivateurs:		N 105381	115497	204339	105919
secteur		n 400	400	400	400
400					
exportation		F 3,79	3,46	1,96	
<u>3,78</u>					
Cultivateurs:		N 109831	54648	91591	
75208					
autres		n 400	400	400	
400					
		F 3,64	7,32	4,37	
<u>5,32</u>					
Non culti-		N 26600	18884	25105	11663

vateurs (B)				
	N_h	44407	241812	189029 321035
192790				
	$F_{\max, h}$	60,21	3,79	7,32 4,37
5,32				
ECHANTILLON TOTAL: <u>Bujumbura</u> 2674 <u>Restant</u> 4729 <u>Ensemble</u> 7403				

NOTE: N = Nombre de ménages existant
n = Nombre de ménages souhaités dans l'échantillon
F = 1000 fois les taux de sondage

On verra que l'échantillon total se situe bien dans les limites que nous nous sommes fixées. Par conséquent aucune modification du schéma n'est nécessaire et on peut accepter les GSE tels qu'ils ont été proposés plus haut.

Cependant, l'on doit tenir compte de deux contraintes. La première contrainte est la taille globale de l'échantillon. Elle a été fixée à 7200 ménages. La seconde contrainte est qu'on doit tirer 10 ménages par SC et 20 ménages par ZD. Il faudra alors ajuster les n_h au multiple de 10 le plus près pour le milieu rural et au multiple de 20 le plus près pour la mairie puis ramener le nombre total de ménages à enquêter à 7200 (seules les strates du milieu rural seront concernées par la diminution du nombre de ménages). On répartit proportionnellement alors les 4520 ménages entre les différents GSE du milieu rural. Cela changera à leur tour les F_h à appliquer. Le dernier cadre du tableau 1 se trouve ainsi modifié légèrement comme il est montré au tableau 2.

Tableau 2 REPARTITION DEFINITIF DE L'ECHANTILLON

	BUJUMBURA	R E S T A N T	D U	B U R U N D
I	(Mairie)	Plaines	Montagnes	Plateau
			et trans. occidentale	Plateau oriental

n_h	2680	880	1320	1340	980
F_h	60,35	3,64	6,98	4,17	5,08

ECHANTILLON TOTAL: Bujumbura 2680 Restant 4520 Ensemble
7200

Nous sommes maintenant en mesure de donner (tableau 3) la liste définitive des GSE avec le nombre estimé de ménages qui seront tirés dans l'échantillon dans chaque GSE, calculé à partir des fractions F_h qui figurent au tableau 2. (Certains de ces effectifs sont légèrement inférieurs à 400 en conséquence de la mise-au-point que nous venons de signaler.) Rappelons, d'ailleurs que certaines des estimation sont calculées d'une manière assez grossière.

Nous avons ajouté à la liste des GSE les deux catégories de "non cultivateurs" mentionnées au chapitre 5, mais dans chaque cas la zone rurale entière (c'est à dire, "restant du Burundi") a été incluse, sans ventilation entre les 4 grandes zones. Sous cette forme on verra que les deux définitions, A et B, donnent lieu à des catégories dont l'échantillon sera adéquat pour l'analyste.

Enfin il importe de noter qu'en visant certains GSE explicitement on n'exclut pas les analyses sur d'autres catégories. En effet toute catégorie dont l'échantillon dépasse 400 (approximativement) peut être considérée comme analysable avec à peu près la même précision que celle que nous avons exigée pour les GSE visés.

Tableau 3 LISTE DES GSE RETENUS

Définition du GSE l'échantillon	Ménages prévus dans (Recensement 1990)
------------------------------------	-------------------------------------------------

BUJUMBURA

1. Secteur public et para-public	401
2. Privé formel	771
3. Secteur informel	1012
4. Autres (inactifs)	496

RESTANT DU BURUNDI

Plaines:

5. Cultivateurs, secteur d'exportation	383
6. Cultivateurs, autres	416

Montagnes et zone de transition:

7. Cultivateurs, secteur d'exportation	806
8. Cultivateurs, autres	382

Plateau occidental:

9. Cultivateurs, secteur d'exportation	852
10. Cultivateurs, autres	381

Plateau oriental:

11. Cultivateurs, secteur d'exportation	538
12. Cultivateurs, autres	383

Ensemble des régions (sauf Bujumbura)

13. Non cultivateurs (A)	586
14. Non cultivateurs (B)	393

NOTE: Le groupe (A) des non cultivateurs sont ceux qui ne citent pas d'activité agricole comme activité principale. Le groupe (B) est constitué de ceux qui ne le cite ni comme activité principale ni comme activité secondaire.

—

7. MODALITES DU TIRAGE DE L'ECHANTILLON

7.1 Principes

La répartition de l'échantillon définie au tableau 2 implique une stratification de la base de sondage aréolaire en 5 strates: la ville de Bujumbura et 4 strates essentiellement rurales. Le tirage de l'échantillon se fera indépendamment dans chaque strate exception faite pour la mairie.

7.1.1. *Modalités de tirage de l'échantillon en Mairie de BUJUMBURA*

La description de la méthodologie qui suit concerne l'échantillonnage en milieu urbain.

On tire un échantillon de ZD avec probabilité proportionnelle à la taille, c'est-à-dire au nombre de ménages recensés dans la ZD. Soit M_i le nombre de ménages recensés dans la ZD et soit M_i' l'effectif de ménages enregistrés lors du dénombrement dans cette ZD. Ces deux quantités seront souvent différentes, que ce soit à cause des mouvements des populations ou des erreurs dans l'une ou l'autre opération.

On montre facilement que la probabilité pour que la ZD soit incluse dans l'échantillon est donnée par la formule:

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

dans une strate donnée, où a = le nombre de ZD tirées dans la strate et la somme $\sum M_i$ s'étend sur la totalité des ZD dans la strate (qu'elles soient tirées ou non). La quantité p_{1i} représente la probabilité relative au 1er degré de sondage et à la ZD i.

Au 2ème degré de sondage on tirera exactement 20 ménages dans chaque ZD à partir des M_i' ménages dénombrés dans cette ZD. La probabilité pour qu'un ménage donné soit tiré dans la ZD i tirée est donc:

$$p_{2i} = 20 / M_i'$$

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

$$F_i = p_{1i} p_{2i} = 20 (a / \sum M_i) . (M_i / M_i')$$

Rappelons qu'il s'agit ici des quantités relatives à une seule strate.

D'une manière plus générale on peut écrire, pour la strate h:

$$F_{hi} = 20 (a_h / \sum_i M_{hi}) . (M_{hi} / M_{hi}')$$

Remarquons que dans le cas où M_{hi} et M_{hi}' sont toujours identiques cette probabilité est constante pour tous les ménages de la strate. Un tel échantillon est appelé "autopondéré" à l'intérieur des strates. Mais puisqu'en fait M_{hi} et M_{hi}' seront souvent différents il faudra dans notre cas pondérer les données lors de l'analyse. Les coefficients de pondération à appliquer sont l'inverse des probabilités globales données par la formule précédente, soit:

$$W_{hi} = 1/F_{hi} = (\sum_i M_{hi} / a_h) . (M_{hi}' / M_{hi}) / 20$$

Ces poids devront être calculés pour chaque SC de l'échantillon et inscrits dans le fichier des données de l'enquête. Il est donc indispensable d'enregistrer et de garder les valeurs M_i (qui proviennent du recensement) dès que le tirage des ZD est au point, et d'enregistrer et garder les M_i' à partir du moment où le dénombrement des ménages en vue de l'enquête ait été complété.

Enfin rappelons que dans toutes les formules citées plus haut les sommes Σ s'étendent sur la totalité des ZD de la strate, qu'elles soient tirées ou non dans l'échantillon.

7.1.2. Modalités de tirage de l'échantillon en milieu rural.

La description de la méthodologie qui suit concerne l'échantillonnage en milieu rural.

Comme en milieu urbain, on tire un échantillon de SC avec probabilité proportionnelle à la taille, c'est-à-dire au nombre de ménages recensés dans la SC. Soit M_i le nombre de ménages recensés dans la SC et soit M_i' l'effectif de ménages enregistrés lors du dénombrement dans cette SC. Ces deux quantités seront souvent différentes, que ce soit à cause des mouvements des populations ou des erreurs dans l'une ou l'autre opération. Par rapport à la taille de la SC, soit S_i le nombre de segments de la SC

On montre facilement que la probabilité pour que la SC soit incluse dans l'échantillon est donnée par la formule:

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

dans une strate donnée, où a = le nombre de SC tirées dans la strate et la somme $\sum M_i$ s'étend sur la totalité des SC dans la strate (qu'elles soient tirées ou non). La quantité p_{1i} représente la probabilité relative au 1er degré de sondage et à la SC i.

Au 2ème degré de sondage on tirera exactement 10 ménages dans chaque SC à partir des M_i' ménages dénombrés dans cette SC. La probabilité pour qu'un ménage donné soit tiré dans la SC i tirée est donc:

$$p_{2i} = 10 / M_i' S_i$$

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

$$F_i = p_{1i} p_{2i} = 10 (a / \sum M_i) . (M_i / M_i' S_i)$$

Rappelons qu'il s'agit ici des quantités relatives à une seule strate. D'une manière plus générale on peut écrire, pour la strate h :

$$F_{hi} = 10 (a_h / \sum_i M_{hi}) . (M_{hi} / M_{hi}' S_{hi})$$

Si $S_{hi}=1$, c'est à dire qu'il n'y a pas eu de segmentation, on retrouve alors la méthode appliquée au 7.1.1.

Les coefficients de pondération à appliquer sont l'inverse des probabilités globales données par la formule précédente, soit:

$$W_{hi} = 1/F_{hi} = (\sum_i M_{hi} / a_h) . (M_{hi}' S_{ih} / M_{hi}) / 10.$$

Ces poids devront être calculés pour chaque SC de l'échantillon et inscrits dans le fichier des données de l'enquête. Il est donc indispensable d'enregistrer et de garder les valeurs M_i (qui proviennent du recensement) dès que le tirage des SC est au point, et d'enregistrer et de garder les M_i' à partir du moment où le dénombrement des ménages en vue de l'enquête ait été complété.

7.2 Application

Les paramètres du sondage sont implicites dans les résultats qui figurent au tableau 2. Les valeurs de n_h représentent l'effectif de ménages qu'il faudra tirer dans chaque strate. En divisant ces chiffres par 20 ou par 10 selon le cas on obtient le nombre de ZD ou de SC à tirer.

Tableau 4 PARAMETRES DU SONDAGE

	BUJUMBURA R E S T A N T D U B U R U N D I					
	(Mairie)	Plaines	Montagnes	Plateau	Plateau	TOTAL
	et trans. occidental oriental					
N_h	44.407	241.812	189.029	321.035	192.790	989.073
n_h	2.680	880	1.320	1340	980	7200
îlots/SC à tirer = a_h	134	88	132	134	98	452

7.3 Modalités du tirage au 1er degré

Le tirage des SC avec probabilités proportionnelles à la taille se fera par la méthode classique. On obtient d'abord la liste des SC de chaque strate, avec la "taille" (nombre de ménages recensés) de chaque SC. La première tâche est de réordonner la liste dans un ordre géographique.

Nous avons effectué cela au niveau des communes seulement, à l'aide d'une carte du Burundi sur laquelle les limites de toutes les communes sont indiquées. (Ordonner les SC elles-mêmes ne serait guère praticable et ajouterait peu à la précision du sondage une fois que les communes ont été disposées convenablement.)

Ensuite les tailles des SC seront cumulées dans la liste des SC ainsi regroupées, au sein de chaque strate, de manière à ce qu'on trouve un cumul en face de chaque SC dans la liste. Le cumul recommencera à zéro au début de chaque strate. On obtient alors pour chaque strate une série de

sondage à partir de du nombre I_h qu'on trouve au tableau 4.

La série commence avec un nombre aléatoire entre 1 et I_h : à partir de ce point on ajoute systématiquement le pas I_h . Pour chaque terme de cette série on cherche dans la colonne des cumuls le premier cumul qui est égale ou qui dépasse ce terme. La SC correspondant à ce cumul est tirée. Le processus se répète avec chaque strate.

7.4 Modalités du tirage au 2e degré

Dans chaque SC tirée on devra organiser un dénombrement des ménages sur le terrain. Bien qu'on vise un échantillon de 20 ou de 10 ménages selon le cas, on en tirera 24 pour le cas de 20 ménages et de 12 pour le cas de 10 ménages afin de disposer d'une réserve pour le remplacement éventuel des ménages défaillants.

Le tirage se fait selon la méthode systématique à probabilité égale. Les étapes sont les suivantes:

- 1) On numérote les ménages dans la liste, de 1 à M_i' .
- 2) On calcule le pas de sondage, $I_{2i}=M_i'/24$ ou $I_{2i}=M_i'/12$. (L'indice 2 ici est utilisée pour signaler le 2ème degré de sondage.)
- 3) On choisit un nombre aléatoire entre 1 et I_{2i} . Ce sera le 1er ménage tiré.
- 4) On tire les autres ménages à l'intervalle I_{2i} à partir du premier.
- 5) Les 4 ménages de réserve sont désignés par un sous-échantillonnage systématique, à l'intervalle de 6 parmi les 24 ou les 12 tirés. Leurs noms seront alors extraits de la liste et donnés au contrôleur de l'équipe qui seul est habilité à autoriser l'utilisation d'un ménage remplaçant.

7.5 Extrapolation

L'extrapolation des données se fera à l'aide des coefficients de pondération dont la formule a été donnée vers la fin du chapitre 7.1.1 et 7.1.2. A noter qu'on doit utiliser 20 ou 10 et non 24 ou 12.

7.6 Paramètres des travaux de terrain

1. En Mairie de Bujumbura

Les travaux de collecte en mairie seront organisés comme suit :

- 1.0.5 jours par équipe et par ZD pour les contacts administratifs et sensibilisation. ;
- 2.1.5 jours par équipe pour le dénombrement dans une ZD et le tirage des ménages échantillons (l'on estime qu'un

agent enquêteur peut enregistrer 50 ménages par jour mais on préfère ajouter une demi-journée pour le cas des ménages où tous les membres du ménage sont absents pour des raisons de service : c'est le cas de la plupart des quartiers résidentiels).

3.1.5 jours par équipe et par ZD pour les interviews.

Soit au total 3.5 jours par équipe et par ZD. Avec 6 équipes, il leur faudra 78 jours pour parcourir les 134 ZD en quatre mois soit en moyenne 19 jours de travail par mois ; ce qui semble raisonnable compte tenu de la disponibilité des fonctionnaires pendant les jours et heures de travail. Les agents seront obligés de travailler plus les samedi et dimanche.

2 En milieu rural.

En milieu rural, chaque équipe prendra chaque en charge deux SC à la fois et les travaux seront organisés de la façon suivante :

- 1 1 jour par équipe pour le déplacement, l'installation et les prises de contacts ;
- 2 1.5 jours pour le dénombrement et le tirage ;
- 3 1 jour pour les interviews.

Soit 3.5 jours par équipe et pour 2 SC. Avec 7 équipes, il leur faudra 113 jours pour couvrir en quatre mois les 452 SC soit en moyenne 28 jours de travail par mois. Cette moyenne de jours de travail ne pose non plus de problème quand on sait que les agents vont s'adapter avec le temps dans le remplissage des questionnaires et à la longue dégager un temps de repos.

N.B Rappelons qu'une équipe est composée de quatre enquêteurs, d'un anthropomètre et d'un contrôleur.

ANNEXE : CLASSEMENT DEFINITIF DES COMMUNES EN FONCTION DES QUATRE ZONES AGRO-ECOLOGIQUES

1. ZONES DE PLAINES

	PROVINCE	COMMUNE	REGION	MENAGE
1	BUBANZA	GIHANGA	1	7 615
2	BUJA-RURAL	BUGARAMA	1	848
3	BURURI	RUMONGE	1	12 845
4	CANKUZO	CENDAJURU	1	3 837
5	CANKUZO	GISAGARA	1	10 787
6	CIBITOKI	BUGANDA	1	882
7	CIBITOKI	RUGOMBO	1	7 417
8	GITEGA	BUKIRASAZI	1	5 102
9	GITEGA	BURAZA	1	9 352
10	GITEGA	ITABA	1	8 825
11	GITEGA	MAKEBUKO	1	11 100
12	KIRUNDO	BUGABIRA	1	12 210
13	KIRUNDO	BUSONI	1	18 493
14	KIRUNDO	BWAMBARANGWE	1	8 746
15	KIRUNDO	GITOBE	1	8 603
16	KIRUNDO	KIRUNDO	1	14 652
17	KIRUNDO	NTEGA	1	16 090
18	MAKAMBA	KAYOGORO	1	10 397
19	MAKAMBA	KIBAGO	1	4 410
20	MAKAMBA	MABANDA	1	4 441
21	MAKAMBA	MAKAMBA	1	8 191
22	MAKAMBA	VUGIZO	1	6 208
23	MUYINGA	MUYANGE-GASHOHO	1	9 791
24	RUTANA	BUKEMBA	1	3 170
25	RUTANA	GIHARO	1	9 866
26	RUTANA	GITANGA	1	5 135
27	RUYIGI	GISURU	1	10 564
28	RUYIGI	KINYINYA	1	5 802
29	RUYIGI	NYABITSINDA	1	6 433
TOTAL				241 812

2. MONTAGNES ET ZONES DE TRANSITION

	PROVINCE	COMMUNE	REGION	MENAGE
1	BUBANZA	BUBANZA	2	2 534
2	BUBANZA	MPANDA	2	1 375
3	BUJA-RURAL	ISARE-MUGARURO	2	3 931
4	BUJA-RURAL	KABEZI	2	20 281
5	BUJA-RURAL	KANYOSHA	2	13 955
6	BUJA-RURAL	MUGONGOMANGA	2	3 611
7	BUJA-RURAL	MUHUTA	2	9 961
8	BUJA-RURAL	MUKIKE	2	3 656
9	BUJA-RURAL	MUTAMBU	2	7 990
10	BURURI	BURAMBI	2	1 098
11	BURURI	BURURI	2	9 156
12	BURURI	BUYENGERO	2	5 044
13	BURURI	MATANA	2	6 262
14	BURURI	MUGAMBA	2	8 445
15	BURURI	RUTOVU	2	6 937
16	BURURI	SONGA	2	7 153
17	BURURI	VYANDA	2	4 564
18	CIBITOKÉ	BUKINANYANA	2	6 353
19	CIBITOKÉ	MABAYI	2	8 520
20	CIBITOKÉ	MUGINA	2	10 915
21	GITEGA	RYANSORO	2	5 379
22	GITEGA	NYARUSANGE	2	6 528
23	KAYANZA	MURUTA	2	8 425
24	MURAMVYA	BUKEYE	2	7 026
25	MURAMVYA	GISOZI	2	4 076
26	MURAMVYA	MURAMVYA	2	8 498
27	MURAMVYA	RUSAKA	2	7 356
TOTAL				189 029

4 PLATEAUX OCCIDENTAUX

	PROVINCE	COMMUNE	REGIO N	MENAGE
1	GITEGA	BUGENDANA	3	14 955
2	GITEGA	GIHETA	3	10 891
3	GITEGA	GISHUBI	3	8 799
4	GITEGA	GITEGA	3	19 752
5	GITEGA	MUTAHO	3	9 278
6	KARUZI	BUGENYUZI	3	14 554
7	KARUZI	GIHOGAZI	3	11 502
8	KAYANZA	BUTAGANZWA	3	8 263
9	KAYANZA	GAHOMBO	3	6 989
10	KAYANZA	GATARA	3	9 503
11	KAYANZA	KABARORE	3	10 587
12	KAYANZA	KAYANZA	3	14 909
13	KAYANZA	MATONGO	3	11 211
14	KAYANZA	MUHANGA	3	10 335
15	KAYANZA	RANGO	3	11 265
16	MURAMVYA	BISORO	3	5 598
17	MURAMVYA	KAYOKWE	3	6 134
18	MURAMVYA	KIGANDA	3	7 311
19	MURAMVYA	MBUYE	3	8 251
20	MURAMVYA	NDAVA	3	9 985
21	MURAMVYA	NYABIHANGA	3	11 368
22	MURAMVYA	RUTEGAMA	3	7 155
23	NGOZI	BUSIGA	3	10 202
24	NGOZI	GASHIKANWA	3	9 056
25	NGOZI	KIREMBA	3	12 430
26	NGOZI	MARANGARA	3	12 295
27	NGOZI	MWUMBA	3	11 458
28	NGOZI	NGOZI	3	16 372
29	NGOZI	NYAMURENZA	3	10 279
30	NGOZI	RUHORORO	3	10 348
TOTAL				321 035

4. PLATEAUX ORIENTAUX

	PROVINCE	COMMUNE	REGION	MENAGE
1	CANKUZO	CANKUZO	4	6 807
2	CANKUZO	KIGAMBA	4	5 108
3	CANKUZO	MISHIHA	4	5 247
4	KARUZI	BUHIGA	4	8 105
5	KARUZI	GITARAMUKA	4	8 809
6	KARUZI	MUTUMBA	4	5 507
7	KARUZI	NYABIKERE	4	6 157
8	KARUZI	SHOMBO	4	6 159
9	KIRUNDO	VUMBI-BUKUBA	4	12 793
10	MUYINGA	BUHINYUZA	4	6 126
11	MUYINGA	BUTIHINDA	4	13 848
12	MUYINGA	GASORWE	4	10 671
13	MUYINGA	GITERANYI	4	10 499
14	MUYINGA	MUYINGA	4	17 095
15	MUYINGA	MWAKIRO	4	5 750
16	NGOZI	TANGARA	4	11 056
17	RUTANA	MPINGA-KAYOVE	4	16 125
18	RUTANA	MUSONGATI	4	7 592
19	RUTANA	RUTANA	4	1 481
20	RUYIGI	BUTAGANZWA	4	10 293
21	RUYIGI	BUTEZI	4	4 610
22	RUYIGI	BWERU	4	5 901
23	RUYIGI	RUYIGI	4	7 051
TOTAL				192 790