

ENQUETE INTEGREE SUR LES CONDITIONS DE VIE DES MENAGES AU RWANDA

Plan d'enquête et plan de sondage

Rapport du Consultant Chris Scott, Juillet 1997¹

1. CADRE

Ce document décrit le plan de sondage proposé pour l'Enquête Intégrée sur les Conditions de Vie des Ménages (EICV) au Rwanda. L'enquête répond au besoin du Gouvernement de connaître les conditions de vie des ménages rwandais. Des données actualisées dans ce domaine lui sont indispensables pour la détermination des objectifs de planification du développement. On demande en particulier les nouvelles pondérations pour l'indice des prix et la révision du secteur des ménages dans les comptes nationaux.

Les objectifs détaillés de l'enquête ainsi que les grandes lignes du plan de travail sont précisés dans le document du projet. Le présent document a pour but d'entrer dans le détail en précisant les modalités exactes à suivre à chaque étape.

2. CONTRAINTES

Deux types de contraintes déterminent le plan de sondage à adopter. D'une part on doit consulter les objectifs de l'enquête: ceux-ci déterminent à une première approximation la taille de l'échantillon et sa répartition dans l'espace et le temps. D'autre part on doit prendre en compte les modalités exactes du travail de terrain: cette considération permettra de raffiner le plan de sondage et d'en fixer les paramètres exacts.

En ce qui concerne les objectifs il y en a un qui est déterminant quant au plan de sondage. Le Gouvernement veut des résultats de l'enquête au niveau de chaque préfecture. Il exige d'ailleurs des résultats utiles au niveau de la ville de Kigali et enfin au niveau du restant du secteur urbain regroupé. Le tableau 1 donne la liste des domaines concernés, avec des estimations des effectifs de ménages qui résident dans chacun.

La tâche du sondeur sera d'élaborer un plan de sondage qui assure dans chacun de ces 13 domaines un échantillon suffisant pour permettre des analyses utiles. Or, dans une enquête par sondage l'erreur de sondage est fonction de la taille de l'échantillon, et non de la fraction de sondage. L'expérience des autres enquêtes analogues dans des pays africains indique qu'un échantillon minimal de l'ordre de 500 ménages est nécessaire pour soutenir des analyses solides, et ceci pour chacun des domaines visés. Il en résulte qu'ils nous faudra un échantillon d'environ 13 x 500, soit de l'ordre de 6.500 ménages au minimum.

Tableau 1. Les Domaines d'étude pour l'EICV

¹ Ce rapport est le fruit d'une mission du consultant du 17 au 31 juillet 1997 au Rwanda. Le consultant tient à remercier le Directeur de la Statistique ainsi que ses collaborateurs, rwandais et expatriés, qui ont contribué à l'élaboration du plan que nous décrivons ici.

Domaine Nombre estimé de ménages
(voir note)

URBAIN

Kigali-ville	74 840
Autre urbain	<u>36 393</u>
	111 233

RURAL

Préfectures de:

Butare	112 587
Byumba	125 718
Cyangugu	104 801
Gikongoro	91 000
Gisenyi	140 531
Gitarama	162 571
Kibungo	71 386
Kibuye	68 014
Kigali-rural	116 249
Ruhengeri	157 384
Umutara	<u>33 889</u>
	1 184 130

NOTE Les chiffres proviennent des travaux préparatifs en vue de l'Enquête Socio-Démographique de 1996. On estime que la mise-à-jour qui est en cours au moment de la rédaction du présent document ajoutera au moins 300.000 ménages.

Un échantillon de cet ordre pour une enquête aussi complexe sera exceptionnelle en Afrique. Seule la discipline rigoureuse permettra de la mener à bien. Le consultant aurait recommandé un échantillon plus modeste mais cela impliquerait des objectifs plus limités que ceux qui ont été fixés par le Gouvernement.

Si nous admettons donc l'échantillon minimal dans les limites des objectifs déterminés on doit allouer environ 500 ménages à chacun des domaines. Nous nous permettrons cependant un petit écart en ce qui concerne le secteur Kigali-ville: cette ville étant la seule d'une taille importante au Rwanda il semble souhaitable de la sur-échantillonner afin de lui assurer un échantillon plus importante: pour elle seule on augmentera l'échantillon de 500 à 700.

Enfin les objectifs de l'enquête déterminent également la répartition de l'échantillon dans le temps. La représentativité dans cette dimension exige que l'enquête soit étalée sur une année complète.

Afin de préciser les paramètres de l'échantillon il faudra étudier de plus près les modalités exacte de la collecte des données.

3. METHODOLOGIE DE LA COLLECTE DES DONNEES

Le questionnaire proposé est long, étant constitué de 12 modules. Un point important concernant la répartition du travail dans le temps c'est que deux des modules sont constitués d'une longue liste de

questions sur les produits (a) achetés et (b) auto-consommés. Or les études menées dans d'autres pays africains montrent que, au moins pour les produits de consommation fréquente, les enquêtés ont du mal à se rappeler de leurs activités au delà d'une période de 3 jours. Par contre, si on devait se limiter à 3 jours de consommation pour chaque ménage on ne disposerait pas d'un échantillon adéquat pour estimer le niveau de vie de chaque ménage: la période de référence serait trop courte. On préférerait une période par ménage de 31 jours (au moins en milieu urbain) afin d'être sûr que le jour de paie soit inclus pour chaque salarié. La question se pose donc: comment peut-on concilier une période de mémoire courte (3 jours au maximum) avec une période de référence par ménage qui s'étale sur un mois? Une réponse est la suivante: on fera de multiples visites auprès de chaque ménage à 3 jours d'intervalle: à chaque visite on interrogera le ménage sur les achats des 3 derniers jours. En ajoutant ces résultats on obtiendra les achats d'un mois entier pour chaque ménage échantillon.

Le fait de faire des multiples visites à chaque ménage présente un autre avantage: on peut profiter de chaque visite pour traiter d'un des modules (logement, santé, éducation, poids et taille des enfants...) sans que le ménage se sente surchargé. Chaque interview est relativement courte même si la totalité des données demandées est volumineuse.

Cette méthode est proposée dans cette enquête pour le secteur urbain. En ce qui concerne le secteur rural les salariés sont rares. Il devient donc moins significatif que la période totale de référence soit d'un mois entier. Nous proposons au lieu qu'on fasse des visites à deux jours d'intervalle (au lieu de trois) et qu'on fasse 8 visites, afin de couvrir une période d'un demi-mois pour chaque ménage.

Dans les deux cas, urbain et rural, on profitera des multiples visites pour faire passer un ou deux modules du questionnaire. Des consignes seront préparées pour les enquêteurs en leur précisant quels modules doivent être passés à chaque visite. Les deux modules "dépenses journalières" et "autoconsommation" seront traités à toutes les visites.

Nous sommes maintenant en mesure de préciser les modalités exactes dans les secteurs urbain et rural respectivement.

Milieu urbain - Kigali-ville

1. Chaque enquêteur travail pendant 33 jours dans une zone.
2. On lui affecte 9 ménages de la zone. Il les divise en 3 lots de 3 ménages.
3. Le 1er jour il fait une visite aux 3 ménages du 1er lot
Le 2e jour 2e lot
Le 3e jour 3e lot

Le 4e jour il fait une 2e visite aux ménages du 1er lot
Le 5e jour2e..... 2e lot
Le 6e jour2e..... 3e lot

Etc.

Le 31e jour il fait sa 11e visite aux ménages du 1er lot
Le 32e jour11e..... 2e lot
Le 33e jour11e..... 3e lot
4. Il répète le même cycle dans une autre zone de Kigali, et ainsi de suite. A la fin de l'année il aura effectué 10 cycles de ce genre, ce qui correspond à 90 ménages répartis dans 10 zones.

5. Huit enquêteurs suivent le même plan dans d'autres zones de Kigali. Ils se regroupent en 2 équipes, chacune sous un contrôleur. Echantillon total: 720 ménages répartis dans 80 zones, le travail étant étalé uniformément sur la totalité de l'année.

Milieu urbain - Autres villes

L'échantillon de zones, ou cellules, est tiré dans les autres villes du Rwanda, proportionnellement à leur populations.

1. Un enquêteur donné sera affecté à une zone/cellule d'une ville donnée pendant 33 jours.
2. Voir Kigali-ville, point 2.
3. Voir Kigali-ville, point 3.
4. Il répète le même cycle dans une autre zone/cellule de l'échantillon, dans une autre ville, et ainsi de suite jusqu'à la fin de l'année (10 cycles dans 10 cellules/zones, 90 ménages).
5. Cinq enquêteurs suivent des cycles de ce genre dans des différentes zones/cellules de l'échantillon. On ajoute un enquêteur de réserve et un contrôleur spécialiste pour le secteur "autres villes". Echantillon total: 450 ménages dans 50 zones/cellules étalés sur une année.
6. Il peut arriver parfois qu'une zone/cellule est tirée deux fois. Dans un tel cas on tirera 18 ménages au lieu de 9; l'enquêteur les divisera en deux groupes de 9 et traitera de chaque groupe dans un cycle séparé de 33 jours, en restant 66 jours dans la même ville.
7. Enfin il arrivera souvent que plusieurs zones/cellules sont tirées dans la même ville. Il convient de ne pas les traiter dans des mois consécutifs, car cela diminuera l'efficacité de l'échantillonnage: les zones/cellules d'une même ville devraient être étalées sur l'année. (On n'applique pas ce raisonnement au cas 6 précédant parce que là où il s'agit d'une même zone/cellule tirée deux fois on aimerait profiter du même travail de dénombrement: il importe que la liste du dénombrement soit aussi à jour que possible.)

Milieu rural

1. Les enquêteurs travaillent par équipe de 4 sous un contrôleur. Cette équipe restera toujours dans la même préfecture. Il y aura donc 11 équipes.
2. Dans chaque préfecture on suit le même cycle. On tirera 40 cellules au début. On les divise en 10 grappes de 4 cellules. Chaque grappe sera affectée à un des cycles de l'enquête (égal à 32 jours).
3. On décrira le programme à suivre pour une de ces grappes dans une des préfectures. Appelons les 4 cellules de la grappe A, B, C et D. Le travail sur le terrain sera facilité si A est proche à B et si C est proche à D. Il importe moins que A+B soit proche à C+D.
4. On affecte deux enquêteurs à la cellule A et deux autres à la cellule B. Ils y resteront pendant 16 jours. Au bout de cette période le groupe travaillant dans A sera transféré à C et le groupe de B sera transféré à D.
5. Ces 4 enquêteurs constituent une équipe et travaillent sous un même contrôleur: celui-ci est muni d'une voiture et sera installé pendant le mois à un endroit proche aux 4 cellules. Il a également à sa charge un enquêteur de réserve.
6. A la fin des 32 jours l'ensemble de l'équipe passe à une autre grappe de 4 cellules, toujours dans la même préfecture.

7. Le programme de travail du cycle de 16 jours au sein d'une cellule donnée est le suivant:

Douze ménages sont tirés par cellule. Ceux-ci seront divisés en deux groupes, un groupe pour chacun des 2 enquêteurs. Pour minimiser les déplacements journaliers on créera ces groupes de manière à constituer deux quartiers, sans chevauchement. Ensuite l'enquêteur divisera son échantillon de 6 ménages en 2 lots, de 3 ménages chacun. L'enquêteur traitera d'un lot les jours pairs et l'autre les jours impairs. Il enquêtera donc les 3 ménages du 1er lot les jours 1, 3, 5, 7, 9, 11, 13 et 15, soit 8 visites en tout. Il enquêtera les 3 ménages du 2e lot les jours 2, 4, 6, 8, 10, 12, 14, et 16.

Il enquêtera toujours 3 ménages au cours d'une journée. Le système ressemble à celui proposé pour le secteur urbain. Mais la rareté des salariés en milieu rural semble justifier une période de référence plus courte: 15 jours au lieu de 31. Enfin cette réduction de la période auprès de chaque ménage exige une augmentation du rythme des visites: on passera tous les 2 jours, au lieu de tous les 3, afin d'assurer suffisamment de visites pour la présentation des nombreux modules.

8. En résumé pour le secteur rural: il y a une équipe de 4 enquêteurs par préfecture, qui tourne toujours au sein de cette préfecture. Il y a 11 préfectures et 10 périodes d'environ un mois (33 jours) dans l'année. Dans chacune de ces "mois" un enquêteur traitera de 2 x 6 ménages. L'échantillon total pour le secteur rural sera donc de $11 \times 10 \times 4 \times 2 \times 6 = 5280$ ménages dans 440 cellules. Le nombre d'équipes étant 11 il y a 44 enquêteurs et 11 contrôleurs avec 11 enquêteurs de réserve dans les équipes. On ajoutera un réserve au siège de 2 enquêteurs et un contrôleur.

Résumé: ensemble des strates

Le tableau 2 résume les paramètres du travail que nous venons de préciser.

Tableau 2. Paramètres principaux de l'enquête

Strate (1)	Enquêteurs		Contrô- Réserve	leurs (4)	Echantillon	
	Princ. (2)	+			(3)	Cell/zones (5)
Kigali-ville	8	+	2	2	80	720
Autres villes	5	+	1	1	50	450
Rural: préfectures						
Butare	4	+	1	1	40	480
Byumba	4	+	1	1	40	480
Cyangugu	4	+	1	1	40	480
Gikongoro	4	+	1	1	40	480
Gisenyi	4	+	1	1	40	480
Gitarama	4	+	1	1	40	480
Kibungo	4	+	1	1	40	480
Kibuye	4	+	1	1	40	480
Kigali-rural	4	+	1	1	40	480
Ruhengeri	4	+	1	1	40	480
Umutara	4	+	1	1	40	480
TOTAL URBAIN	13	+	3	3	130	1170
TOTAL RURAL	44	+	11	11	440	5280
TOTAL	57	+	14	14	570	6450
Réserve au siège	2		1			

4. PLAN DE SONDAGE

Bases de sondage

Les bases de sondage existent à deux niveaux: base aréolaire et base ménage. Au niveau aréolaire on cherchera une liste d'unités qui couvrent l'ensemble de chaque strate sans chevauchement, qui sont d'une taille assez homogène, et qui sont bien délimitées, si possible sur les cartes. Les unités de taille modeste, un ou deux cent ménages par exemple, seraient l'idéal.

Au Rwanda nous disposons d'un ensemble d'unités qui correspond bien à ces conditions, au moins en milieu rural: il s'agit des cellules. Ces cellules, qui correspondent approximativement à ce qu'on appellerait "villages" ou "localités" dans des autres pays, se trouvent partout. Il sont au nombre d'environ 9.000. Ils sont à peu près de la taille souhaitée et chaque personne sait à quelle cellule il appartient.

Le seul inconvénient avec cette unité c'est qu'il n'a pas le même caractère à Kigali. Dans la capitale les cellules sont en moyenne environ 5 fois plus peuplées et cette taille est trop grande pour qu'on les utilise comme unités finales de sondage aréolaire. On sera obligé plutôt de se baser sur la zone, l'unité de subdivision des cellules qu'on trouve partout là où les cellules sont très grandes.

Les cellules, et les zones à leur sein, ont l'avantage d'être des unités de l'administration. Le responsable de la cellule/zone maintient à jour une liste des ménages qui y habitent. Ces listes peuvent être imparfaite, surtout à Kigali, mais leur précision est entièrement suffisante pour permettre qu'on utilise le nombre de ménages enregistrés par le responsable comme mesure de taille de l'unité. Disposer d'une mesure de taille pour chaque unité aréolaire présente le grand avantage pour le sondeur que cela lui ouvre la porte à la méthode classique de sondage avec probabilités proportionnelles à la taille. Avant d'étudier cette possibilité, il convient de terminer la discussion sur les bases de sondage en examinant l'échantillonnage au niveau des ménages.

Au Rwanda deux options s'offrent. En premier lieu on pourrait tirer l'échantillon dans chaque cellule/zone à partir de la liste des ménages qui est détenue par le responsable de la cellule/zone. Ou bien on pourrait faire comme dans les autres pays africains (qui ne possèdent pas de telles listes): entreprendre quelques semaines avant l'enquête une opération de dénombrement dans chaque unité aréolaire de l'échantillon afin de fournir la liste des ménages souhaitée. Après des discussions approfondies dans la Direction de la Statistique on a choisi la deuxième option. Bien que la première soit plus simple on ne peut, à l'heure actuelle, être sûr de l'exactitude des listes existantes. En adoptant la deuxième solution on se permettra après l'enquête d'examiner cette question en comparant le nombre de ménages signalés par le responsable, avec le nombre trouvés par le dénombreur. Au cas où ils s'accordent bien on pourrait se permettre dans la prochaine enquête d'utiliser la méthode plus économique. Dans le cas contraire on saura au moins qu'on a évité l'erreur.

Théorie

Nous recommandons le plan de sondage classique "échantillon autopondéré avec probabilités proportionnelles à a taille". Nous exposons d'abord la théorie: ensuite nous verrons les modifications qui s'imposent dans la pratique.

On suppose un échantillon à deux degrés de sondage. Au 1er degré on sélectionnera un certain nombre d'unités aréolaires (les unités primaires, ou UP). Au 2e degré on tirera des ménages au sein de chaque UP tirée.

Soit M_i le nombre de ménages existants dans l'UP i . M_i est donc la taille de l'UP i .

On tirera les UP avec probabilités proportionnelles à la taille M_i . La probabilité de tirage pour l'UP i sera donc

$$p_{1i} = k M_i \quad \text{-----} (1)$$

Plus exactement, p_{1i} est la probabilité pour que l'unité i soit tirée dans l'échantillon, k étant une constante. L'indice 1 est introduit pour signaler le 1^{er} degré de sondage.

On peut déterminer la constante k de la manière suivante. Faisons la somme de l'équation (1) sur la totalité de la population. On doit obtenir à gauche le nombre d'unités qui seront tirées, disons a . A droite on aura kM , où M est le nombre total de ménages dans la population, soit la somme des M_i . Nous avons donc $k = a/M$, d'où:

$$p_{1i} = aM_i/M \quad \text{-----} \quad (2)$$

Quant au 2^e degré de sondage, on tirera un certain nombre b_i de ménages dans l'UP i qui en contient M_i . La probabilité de tirage pour un ménage donné dans cette UP sera donc

$$p_{2i} = b_i/M_i \quad \text{-----} \quad (3)$$

Il s'agit ici de la probabilité conditionnelle, qui suppose que l'UP i ait déjà été tirée.

La probabilité globale pour un ménage d'être tiré est le produit de p_1 et p_2 , d'où

$$F_i = p_{1i} \cdot p_{2i} = (aM_i/M) \cdot (b_i/M_i) = ab_i/M \quad \text{-----} \quad (4)$$

Il en suit que si on tire un nombre constant b de ménages dans chaque UP la probabilité globale sera pour tout ménage

$$F = ab/M \quad \text{-----} \quad (5)$$

Autrement dit, tous les ménages auront la même probabilité d'être tirée. Un tel échantillon est appelé autopondéré. Il représente directement la population sans qu'on ait besoin de le pondérer. Les moyens, rapports, pourcentages ou taux trouvés dans l'échantillon peuvent être cités tels quels, comme estimations des moyens, taux, etc correspondants dans la population.

A part l'avantage d'être autopondéré le fait qu'on aura un nombre fixe b de ménages à étudier dans chaque UP présente un avantage considérable d'ordre logistique.

Malheureusement dans la pratique ce modèle convenable ne correspond pas toujours à la réalité. Le M_i qui figure à la formule (2) doit être connu pour chaque unité avant qu'on puisse entamer le tirage au 1^{er} degré. Dans beaucoup d'enquêtes ces chiffres sont assez inexacts. C'est ainsi qu'on préfère d'utiliser les données issues du dénombrement dans les UP-échantillon comme M_i dans la formule (3). Au moins ceux-ci seront à jour. Si ces valeurs divergent (appelons M_i' celui qui vient du dénombrement) la formule (4) deviendra

$$F_i = (ab_i/M) \cdot (M_i/M_i') \quad \text{-----} \quad (6)$$

Même si on admet b constant, nous voyons que cela ne nous donne plus l'auto pondération. La F_i va varier dans chaque UP. On sera obligé d'introduire le coefficient de pondération

$$W_i = M_i'/M_i \quad \text{-----} \quad (7)$$

propre à chaque UP de l'échantillon.

Dans la présente enquête nous ne savons pas si ces poids seront nécessaires. Tout dépend de l'importance de la divergence entre le M_i fourni par le responsable de la cellule et le M_i' du dénombreur. Si ces chiffres s'accordent approximativement (par exemple, si on ne trouve aucune UP avec une divergence dépassant 5%) on pourra être assez sûr que la pondération, théoriquement nécessaire, n'aurait qu'un

impact négligeable sur les résultats. C'est une question qu'on ne pourra aborder qu'après de disposer des résultats de l'enquête.

La discussion qui précède suppose qu'on se situe dans une seule strate. En fait, comme on voit au tableau 2, notre échantillon sera constitué dans 13 strates. Si on désigne la strate par l'indice h , on doit insérer cet indice dans chacune des formules de (1) à (7). Par exemple, la formule (5) devient

$$F_h = a_h b_h / M_h \quad \text{----- (5)}$$

L'autopondération dont on a parlé plus haut dépend de la constance de F . Or même si F_{hi} est constante sur i , elle ne sera pas constante sur h dans notre enquête car le paramètre M_h varie selon h . Par conséquent notre échantillon, même s'il est autopondéré au sein de chaque strate, ne sera jamais autopondéré à travers les strates. Il y aura donc un poids à donner à chaque strate si on veut réunir les données provenant de plusieurs strates. On parlera encore des pondérations au chapitre 6.

Application: 1er degré de sondage

Afin d'appliquer la méthodologie décrite plus haut il faudra disposer des éléments suivants:

- La liste des cellules/zones par strate, avec la taille M_{hi} de chacune telle que signalée par le responsable. Au moment de la rédaction de ce rapport ces listes ont été commandées. On peut espérer les voir au siège au cours du mois d'août. Dès qu'une des listes soit disponible pour une strate on pourra faire la somme des tailles $M_h = \sum M_{hi}$. En même temps on calculera les cumuls des tailles qu'on ajoutera dans une colonne de droite sur la liste.
- Les effectifs a_h de cellules/zones à tirer. Ce sont les chiffres qu'on trouve dans la colonne (5) du tableau 2.

Avec ces données en main on procédera à un tirage des cellules/zones avec probabilités proportionnelles à la taille, en appliquant la méthode de sondage systématique (c'est-à-dire, tirage à intervalle fixe). Le travail se fera indépendamment dans chaque préfecture, les préfectures étant les strates. Le fait de tirer par la méthode systématique contribue un élément de stratification supplémentaire dans la mesure où les cellules figurent dans la liste selon un ordre significatif. Dans cette hypothèse, tirer à intervalle fixe signifie tirer dans toutes les parties de la liste, ce qui assure une bonne répartition de l'échantillon sur la variable qui détermine l'ordre des unités dans la liste. C'est ce qu'on appelle la stratification implicite. Dans notre cas nous savons peu de choses à l'avance sur les caractéristiques des unités (cellules) mais on peut au moins les ordonner selon la hiérarchie administrative (préfecture-commune-secteur-cellule). Beaucoup de variables sont liées statistiquement à la situation géographique: une telle stratification implicite tendra donc vers une diminution de l'erreur de sondage en assurant une bonne répartition de l'échantillon dans l'espace.

Les étapes du tirage de l'échantillon au 1er degré sont les suivantes.

Vérifier d'abord que la liste des cellules/zones sur laquelle on fera le tirage est bien classée dans l'ordre hiérarchique administratif avant de calculer les cumuls.

Dès qu'on dispose de M_h et a_h pour une strate h on pourra calculer l'intervalle de sondage $I_h = M_h/a_h$, arrondi à l'entier.

Trouver un nombre aléatoire C_h entre 1 et I_h .

Ecrire la série de sondage:

$$C_h : C_h + I_h : C_h + 2I_h : C_h + 3I_h : \text{etc.}$$

En approchant cette série avec la colonne des cumuls des tailles, on tirera les cellules/zones. La méthode en est la suivante. Pour chaque terme de la série chercher parmi les cumuls le premier qui est égal ou supérieur à ce terme. La ligne qui contient ce cumul correspond à la cellule tirée.

Après avoir effectué le tirage, vérifier qu'on a effectivement tiré le nombre a_h de cellules/zones souhaité.

Vérifier le tirage de chaque unité. L'expérience montre qu'au premier essai des gens se trompent d'une ligne de temps en temps en effectuant un tirage de ce genre.

Application: 2e degré de sondage (ménages)

On enverra d'abord des agents dans toutes les cellules/zones tirées pour effectuer le dénombrement: c'est-à-dire faire la liste de tous les ménages habitant la cellule/zone. Ces visites seront planifiées pour être effectuées environ 2 mois avant l'enquête dans la cellule/zone concernée. Les listes seront ramenées au siège dès qu'elles seront complètes.

On tirera l'échantillon de 2e degré en sélectionnant les ménages à partir de ces listes. A cette fin on aura besoin des éléments suivants pour chaque unité i tirée:

La liste des ménages obtenue lors du dénombrement
 Son effectif - nombre de ménages M_i'
 Le nombre de ménages à tirer dans chaque unité.

C'est à ce point qu'on doit soulever la question de ménages défaillants, qu'il s'agit de refus, d'absents, ou de non trouvés. De différentes enquêtes ont adopté des différentes méthodes pour traiter de ce problème. Dans une enquête qui suppose un nombre fixe de ménages à enquêter dans chaque aire, il convient de remplacer chaque ménage défaillant par un autre qui peut être enquêté. C'est notre cas. Enfin pour permettre de tels remplacements il convient de sélectionner dès le départ un excédent de ménages. On constituera ainsi pour chaque cellule/zone un petit échantillon de réserve.

En milieu urbain il est prévu de tirer $b = 9$ ménages à enquêter par cellule/zone. On ajoutera un réserve de 3. Total à tirer: $b' = 12$.

En milieu rural il est prévu d'enquêter $b = 12$ par cellule. On ajoutera 4 de réserve. Total à tirer: $b' = 16$.

Le tirage sera fait au siège. On divise M_i' par b' et on arrondit à l'entier pour obtenir l'intervalle de tirage l_i . On cherche un nombre aléatoire c_i entre 1 et l_i pour débiter. On numérote les ménages de 1 à M_i' . Les ménages tirés sont ceux portant les numéros

$$c_i : c_i + l_i : c_i + 2l_i : c_i + 3l_i : \text{etc.}$$

Après le tirage vérifier qu'on a effectivement tiré les b' souhaités. Dans le cas autre, ou bien on a fait une erreur, ou bien il s'agit d'une chance de l'arrondi qui a donné 1 de trop ou 1 de moins. S'il s'agit d'une erreur, la rectifier. Si on a obtenu un de plus, en supprimer un au hasard. Si on a tiré un de moins, tirer un ménage de plus dans la liste, au hasard.

Numéroter les ménages tirés de 1 à 12 (milieu urbain) ou de 1 à 16 (milieu rural).

Il convient d'affecter à la liste des réserves les ménages numérotés 4, 8 et 12 en milieu urbain, et les ménages numérotés 4, 8, 12 et 16 en milieu rural. Ces listes de réserves seront détenues par le contrôleur. Sur la liste destinée à l'enquêteur on fera figurer seuls les ménages choisis initialement pour l'enquête (numéros non divisible par 4). Au cas où l'enquêteur ne peut obtenir l'interview il doit demander au contrôleur un remplaçant: le contrôleur doit se satisfaire quant à l'impossibilité de l'interview initiale avant d'admettre le remplacement. Dans l'affirmatif, il donnera à l'enquêteur le nom du remplaçant le plus proche dans la liste.

Lorsque l'enquêteur utilise un ménage de remplacement il ne lui donne pas le numéro du ménage remplacé: au contraire, il doit maintenir le numéro (divisible par 4) qui lui est signalé par le contrôleur. De cette manière on sait qu'un remplacement a eu lieu.

5. Les UP de taille anormale

Certaines des UP sont si grandes que le travail de dénombrement exhaustif semble excessif. Le cas est fréquent dans la ville de Kigali. Ailleurs il est rare.

Etudions d'abord le cas rural. Il convient de fixer une règle qui stipule que toute cellule dépassant une certaine taille (taille signalée par le responsable) doit être scindée. Nous proposons de fixer la limite à 300 ménages. Rien n'empêche cependant de fixer ce seuil plus bas, ou plus haut: cela dépend des organisateurs de l'enquête. Il est pourtant souhaitable de prendre une décision définitive avant de commencer le sondage. Plus le seuil est petit, moins on aura de travail de dénombrement, mais plus on aura de travail de délimitation des segments, avec des travaux associés au bureau, et plus on aura de chances de se tromper dans cette étape.

Au cas où une cellule dépasse le seuil fixé, il doit être scindée dans des sous-cellules, ou zones. Dans la liste des cellules avec leurs tailles cette cellule sera remplacée par les zones qui la constituent. Il importe de noter, cependant, que ce travail est inutile si la cellule n'est pas tirée dans l'échantillon. On doit donc d'abord faire le tirage, et ensuite chercher les cellules tirées qui dépassent la limite de taille.

En supposant qu'on trouve une cellule tirée qui est trop grande comment doit on procéder? Soit la taille de cette cellule M_i selon le responsable. On doit la scinder en deux sous-cellules (trois si nécessaire) de manière à ce que la somme des tailles de celles-ci soit égale à M_i . A cette fin il faudra reprendre le contact avec le responsable pour lui demander des précisions sur cette subdivision. Il se peut par exemple qu'il a déjà établi 4 zones A, B, C, D qui ont été déléguées à des adjoints. On lui demandera la taille de chacune. Deux problèmes peuvent se poser.

1) Il nous faudra des zones d'une taille inférieure à la limite (disons 300). Il est fort probable qu'un découpage en 2 sous-cellules suffira à cette fin. Si par exemple il nous signale:

<u>Zone</u>	<u>Taille</u>
A	100
B	120
C	140
D	150

Pour nous il suffit de créer deux sous-cellules, qu'on pourra appeler A+B et C+D. Selon le responsable leurs tailles seront respectivement 220 et 290. Le premier principe est donc: ne créer pas plus de sous-cellules que nécessaire. Dans cet exemple, on en fera 2.

2) Il se peut que la somme des tailles des 4 zones qu'ils nous donne ne soit pas égale à la taille de la cellule entière qu'il nous a donnée initialement. Par exemple, s'il a signalé $M_i = 450$ initialement, la somme des tailles de ses 4 zones est maintenant 510. Que doit-on faire? On doit se rappeler que ce sont les M_i , des estimations, qui seront mise à jour bientôt par les M_i' . Rien ne nous empêche de les ajuster. En effet,

le sondage a été fait en supposant $M_i = 450$. Si on maintient le tirage des autres cellules il faut que se chiffre soit maintenu, autrement toute la série de sondage changera. On doit donc ajuster les chiffres sur la 4 zones, obtenu plus récemment du responsable, pour maintenir le total de 450. En gardant leurs tailles relatives, on obtiendra:

<u>Zone</u>	<u>Taille ajustée</u>
A	88
B	106
C	124
D	<u>132</u>
	450

Le 2e principe est donc: en remplaçant la cellule trop grande par des sous-cellules qui la composent, veillez à ce que la taille de l'ensemble ne change pas.

EXEMPLE

Dans cet exemple nous présentons un extrait imaginaire de la liste des cellules, avec leurs taille et les cumuls.

<u>Cellule</u>	<u>Taille</u>	<u>Cumul</u>	<u>Tirage</u>
.	.	.	
.	.	.	
Mumba	121	19810	
Kenzi	210	20030	
Gataru	410	20440	X 20360
Kagaro	180	20620	
.	.	.	

La série de sondage a donné lieu au terme 20360. En rapprochant ce terme avec la série des cumuls on a sélectionné la cellule de Gataru. On voit que celle-ci dépasse la limite-taille admise. On doit donc diviser Gataru en deux morceaux. En communiquant avec le responsable de la cellule on a pu scinder Gataru ainsi:

Sous-cellule de Gatara: 200 ménages
Sous-cellule de Karamo: 200 ménages.

La taille totale étant réduit de 410 à 400, on ajustera ces chiffres pour donner 205 à chaque sous-cellule, ce qui maintient la taille globale inchangé à 410.

On remplace la cellule de Gataru par ces deux sous-cellules. Les tailles et les cumuls changent ainsi:

<u>Taille</u>	<u>Cumul</u>	<u>Tirage</u>
Gatara	205	20235
Karamo	205	20440
		X 20360

Le numéro de tirage 20360 ne change pas. Cette fois-ci il tombe sur la sous-cellule Karamo, qui est donc l'unité sélectionnée. Le restant des cumuls ne changent pas et tout le restant de l'échantillon est maintenu.

Cette méthode convient très bien là où les cellules d'une taille aberrante sont rares. A la ville de Kigali il en est autrement. Dans cette ville les cellules sont toutes, ou presque toutes, très grandes et il faudra envisager l'utilisation des zones pratiquement partout. En effet les zones à Kigali correspondent aux

cellules ailleurs. Au moment de la rédaction de ce rapport on ne sait pas encore si certaines zones restent trop grandes de manière à ce qu'on doit chercher le moyen de les segmenter encore.

Enfin, pour les autres villes, il semble probable que les cellules ressemblent celles du milieu rural. Encore une fois, à l'heure actuelle on ne dispose pas des données précises permettant de fixer un seuil approprié pour déterminer la segmentation.

Ces dernières questions doivent être décidées après le départ du consultant.

6. Pondération et extrapolation

La pondération à appliquer sera dans tous les cas l'inverse de la probabilité de sondage. On peut écrire:

$$W_{hi} = 1/F_{hi} .$$

Cette formule définit le facteur d'extrapolation **W**. En multipliant les données chacune par son **W** propre à lui, et en faisant la somme sur l'échantillon, on peut extrapoler les résultats de l'échantillon jusqu'à la population totale.

La grande majorité des estimations dans une enquête sont des taux, des moyens, des ratios, pourcentages, and proportions. Dans tous ces cas il y aura un numérateur et un dénominateur les deux avec le même régime de poids. Si on multiplie tous les poids par un constant K ceci paraîtra en haut et en bas de la fraction et n'aura aucun effet. Pour cette raison les poids sont généralement considéré comme relatifs. On a besoin du poids absolus seulement pour estimer les totaux. Le **W** cité plus haut est absolu: on peut représenter par un petit **w** le poids relatif.

$$w_{hi} = K W_{hi}, K \text{ étant une constante arbitraire.}$$

Dans la présent enquête nous avons deux source des poids. D'abord il y a un poids pour chaque strate:

$$W_{1h} = M_h/a_h b_h$$

Les **M_h** seront connus lorsqu'on aura reçu au siège la totalité des données sur le nombre de ménages dans chaque cellule.

Les **a_h** sont les nombres de cellules ou de zones qui figurent dans la colonne (5) du tableau 1.

Les **b_h** sont normalement 9 pour les strates urbaines ou 12 pour les strates rurales. Il peut arriver, rarement, que dans certaines cellules ces chiffres ne sont pas atteints, pour une raison quelconque. (Questionnaire perdu, par exemple.) Dans un tel cas remplacer ce **b_h** par l'effectif réellement présent dans le fichier.

Enfin il peut y avoir des poids au niveau de la cellule, comme nous l'avons déjà vu (formule 7), au cas où les quantités **M_i'** se révèlent significativement distinctes des **M_i**. Dans ce cas on aura:

$$W_{2i} = M_i'/M_i$$

Le poids globale, qui s'attache à chaque ménage, en fonction de la strate **h** et la cellule **i** dans laquelle il réside, est le produit des deux poids **W₁** et **W₂**.

Enfin, rappelons que dans les villes autres que Kigali la strate est la 2ème: elle n'est pas déterminée par la préfecture.

Au cas où on désire étudier des préfectures complètes, on doit additionner les données rurales de la préfecture, pondérées selon la préfecture, avec celles de la ou les ville(s) qui appartient à cette préfecture, pondérées avec le poids de la strate 2 (autres villes).

7. Répartition de l'échantillon dans le temps

Le plan de travail décrit dans les chapitres précédents propose:

- En milieu urbain: $33 \times 10 = 330$ jours de travail à accomplir par chaque enquêteur au cours de l'année
- En milieu rural: $16 \times 2 \times 10 = 320$ jours de travail à accomplir chaque équipe au cours de l'année

Les objectifs de l'enquête exigent que les résultats se rapportent à une année complète (pas nécessairement du 1er janvier au 31 décembre, mais en tout cas sur une période de 365 jours). La période de 33 jours consécutifs pour chaque enquêteur en milieu urbain, et de 16 jours pour chaque équipe en milieu rural, doit rester intacte dans tous les cas, mais on n'a pas défini encore la manière d'utiliser les jours intervenants. En milieu urbain on dispose d'un reliquat de 35 jours par enquêteur disponibles dans l'année et de 9 périodes non définies entre les blocs de travail. En milieu rural on a 45 jours de disponibles par équipe dans l'année et de 19 périodes non définies entre les blocs de travail. On pourrait donc donner, en milieu urbain 4 jours de congé à chaque enquêteur après son bloc de 33 jours de travail, et en milieu rural 2 ou 3 jours de congé après chaque bloc de 16 jours.

Il serait également possible de donner des congés relativement longs, en réduisant le temps libre entre les blocs de travail. Par exemple, en milieu urbain on pourrait donner un jour de repos entre blocs (total: 9 jours) plus une période continue de 26 jours à une période à choisir.

Quel que soit la décision il est souhaitable que les congés des différents enquêteurs urbains (13 agents en tout) ne se chevauchent pas. Quant aux équipes rurales les 2 ou 3 jours entre les cellules sont minimales et peuvent se chevaucher entre équipes sans gêne. Par contre, si on envisage de réduire ces intervalles pour permettre des congés plus prolongés il faudrait que les congés des 11 équipes rurales ne se chevauchent pas. Des trous, impliquant plusieurs équipes à la fois et d'une durée de plus de 3 jours, sont à éviter.

Une fois que l'échantillon primaire est tiré il serait bon de préparer un calendrier détaillé du travail de terrain, en fixant les dates et les équipes proposées pour le travail dans chaque cellule/zone.

8. Réponses partielles

Il arrivera parfois qu'un ménage s'absente, ou déclare le refus, au cours de la période prévue pour l'enquête. Que doit on en faire?

Si cela arrive avant le début de la période on doit évidemment remplacer le ménage.

Si cela arrive avant que le ménage ait achevé une semaine de travail, on le remplacera en rejetant les données déjà obtenues. Quant au ménage remplaçant, on s'efforcera d'inclure le programme complet pour lui dans la période réduite qui reste, à l'aide des questions retrospectives sur la semaine manquante.

Si la défaillance commence lors de la deuxième semaine, on peut tenter la même solution, ou bien à la rigueur on pourra reporter tout le cycle de travail chez le remplaçant d'une semaine ou plus. On peut appuyer sur l'enquêteur de réserve dans un tel cas.

Si la défaillance se produit lors de la 3e semaine, nous sommes déjà nécessairement dans le cas urbain, car le cycle rural ne dure que 2 semaines. C'est le cas le plus difficile. On peut tenter une des

solutions plus haut mais on sera probablement obligé de laisser tomber ce ménage sans remplacement. A noter que le poids W_j pour les ménages de cette strate doit être augmenté dans ce cas dans la proportion de 9 à 8.

Enfin si c'est la 4ème semaine qui manque on pourra accepter les données dont on dispose, extrapoler les données sur les dépenses journalières pour la semaine qui manque, et noter SANS REPONSE pour les modules qu'on n'a pu poser.

Ces règles n'ont rien de formel. Beaucoup dépendra de la situation exacte trouvée dans chaque cas. Le contrôleur de l'équipe essaiera de négocier la meilleure solution possible.

9. Le dénombrement des ménages

Le dénombrement des ménages est une opération majeure. Il touche au moins 10 fois plus de ménages que l'enquête proprement dite. De ce fait il importe de veiller à ce qu'on ne gaspille pas des ressources en posant des questions inutiles lors de cette opération.

Il suffit de noter le nom du chef de ménage et une indication qui permet de trouver sa résidence (lot? n? ? rue?). En se limitant à ces données un agent pourra lister une centaine par jour.

Dès qu'on demande à l'agent dénombreur de poser des questions au ménage on risque de réduire son rendement à une trentaine de ménages par jour.

Afin d'aider les agents dans la recherche des ménages, on pourrait envisager que le dénombreur attache une étiquette sur la porte de chaque ménage dénombré. Les étiquettes, collés au verso, pourraient être pré-imprimées avec les numéros de 001 à 300 en grands chiffres, les dimensions de la feuille étant de l'ordre de 5cm x 8cm. Il portera ce chiffre sur son formulaire de dénombrement. L'agent doit numéroter également les maisons inoccupées, mais il doit signaler INOCCUPE sur le formulaire dans la colonne réservée au nom. L'échantillonnage ne tiendra pas compte des tels logements.

Enfin il importe, surtout en milieu urbain, que le dénombrement se fasse quelques semaines seulement avant l'enquête dans chaque zone concernée. Il serait aussi regrettable de faire le dénombrement dans une cellule trop tôt, que trop tard.