

PN-ARW-420

81163

ASPAP/DAI  
RAPPORT No.

ASPAP/DSA Rapport Technique:

RECOMMANDATIONS SUR LES PROCEDURES D'ESTIMATION POUR  
L'ENQUETE NATIONALE AGRICOLE DU RWANDA

Projet des Enquêtes Agricoles et Analyse  
des Politiques Economiques du secteur rural  
(ASPAP)

Contrat USAID No. 696-0126

David J. Megill\*  
Consultant de courte durée de DAI\*\*

Kigali, Rwanda  
Juin 1992

\*1504 Kenwood Ave., Alexandria, VA 22305, USA

\*\*Development Alternatives, Incorporated (DAI), 7250 Woodmont  
Avenue, Suite 200, Bethesda, MD 20814, USA, principal contractant  
institutionnel pour ASPAP, USAID Contract No. 696-0126-C-00-7777-00

## TABLE DES MATIERES

	Page No.
1. Présentation Générale.....	1
2. Finalisation du plan de sondage pour l'Enquête Nationale Agricole.....	2
3. Calcul des Facteurs de Pondération pour le nouvel échantillon....	3
4. Ajustement des Pondérations pour les DR échantillons manquant....	4
5. Ajustement des Pondérations pour les ménages de l'échantillon non interviewés.....	5
6. Ajustement Facultatif des Facteurs de Pondération basés sur des Projections de la Population.....	6
7. Estimations de l'Enquête.....	7
8. Calcul des Variances pour les Estimations de l'Enquête.....	8
ANNEXE A. Tableaux des écarts types (erreurs d'échantillonnage) pour l'Enquête Nationale Agricole de 1990.	

## 1. Présentation Générale

L'objectif de ce rapport est de fournir une documentation sur la méthodologie d'estimation des données de l'Enquête Nationale Agricole en se basant sur le nouveau plan de sondage et aussi de présenter les procédures de calcul des Facteurs de Pondération appropriés ainsi que les variances des estimation de l'enquête. Ces procédures d'estimation sont basées sur le nouveau plan de sondage décrit dans le précédent rapport technique de David Megill, "Recommandations préliminaires pour le plan de sondage du Programme National des Enquêtes Agricoles au Rwanda" (Mars 1992).

Au cours de la précédente mission de Megill dont l'objectif était de mettre en place un plan de sondage, la Division des Statistiques Agricoles (DSA) n'avait pas encore décidé sur le nombre final de districts de recensement (DR) qui devaient être inclus dans l'échantillon. Par conséquent un plan de sondage flexible basé sur la sélection de 4 sous-échantillons de 78 DR chacun a été élaboré. Suivant la disponibilité des ressources pour les travaux de collecte, l'échantillon pour une quelconque enquête agricole devrait être basée sur l'un ou plusieurs sous-échantillons. Les procédures d'estimation seront similaires quelque soit la taille de l'échantillon final avec les facteurs de pondérations dépendant sur le nombre de sous-échantillons inclus dans l'enquête. Les procédures d'estimation décrites dans ce rapport seront rendues flexibles pour qu'elles soient facilement adaptées à l'échantillon final.

Tout le travail durant cette mission a été encore réalisé en collaboration étroite avec Mr. Samuel Munyaneza, Statisticien de la DSA. De grands remerciements sont adressés à Mr. Anastase Murekez, Directeur de l'ASPAP, et à tout le personnel de la DSA ainsi qu'au Dr. Tardif-Douglin, Représentant du Contractant, Development Alternatives, Incorporated (DAI), pour leur assistance et leur soutien inestimables.

## 2. Finalisation du Plan de Sondage pour l'Enquête Nationale Agricole

Après avoir considéré les différentes alternatives telles que décrites dans le rapport de Megill sur les recommandations préliminaires du plan de sondage, la DSA est actuellement entrain de finaliser le plan de sondage pour la prochaine enquête. Etant donné que les ressources disponibles sont actuellement limitées pour l'Enquête Nationale Agricole, l'échantillon va, dans un premier temps, porter sur un sous-échantillon de 78 DR. Dans les années à venir, il sera possible d'inclure d'autres sous-échantillons dans l'enquête suivant la disponibilité des ressources en vue d'augmenter la précision, au niveau régional, des estimations de la production des cultures.

L'analyse déjà menée pour comparer les données provenant de la collecte mensuelle de la production agricole à celles provenant de la collecte hebdomadaire montre une sous-estimation de la production agricole de la plupart des cultures par la collecte mensuelle, même si dans beaucoup de cas les différences n'étaient pas significatives statistiquement parlant. La DSA a décidé de continuer la collecte mensuelle de la production agricole pour voir si cette tendance va continuer au cours du temps. Etant donné l'inquiétude à propos du biais potentiel dans les données collectées mensuellement, la DSA est encore entrain d'étudier la possibilité de combiner les données provenant des deux méthodes. Les estimations de la production agricole sont actuellement basées uniquement sur les données collectées de façon hebdomadaire.

Etant données les difficultés de se déplacer d'un DR à un autre, chaque enquêteur va collecter les données dans un seul DR. En vue de maintenir un volume de travail raisonnable pour l'enquêteur dans chaque DR, 16 ménages échantillons seront sélectionnés pour la collecte hebdomadaire de la production agricole dans chaque DR, ainsi que 16 ménages échantillons pour la collecte mensuelle. Ceci correspond au volume de travail actuel de l'enquêteur; ce qui semble être tout à fait raisonnable. A partir des listes actualisées des chefs de ménages, un échantillon de 4 grappes de 12 ménages chacune sera sélectionnée dans chaque DR. Chaque grappe sera ensuite divisée en deux groupes de 6 ménages chacun (4 à enquêter et 2 pour des remplacements éventuels) pour la collecte mensuelle et pour la collecte hebdomadaire. Cette façon de former les grappes permet d'avoir un plus haut niveau de corrélation entre l'échantillon utilisé pour la collecte hebdomadaire et celui utilisé pour la collecte mensuelle des données, ce qui devrait améliorer l'analyse des différences dans les résultats provenant des deux méthodes de collecte.

Etant donné un échantillon de 16 ménages (pour la collecte hebdomadaire) dans chacun des 78 DR échantillons, la taille totale de l'échantillon au niveau

national sera de 1.248 ménages, sans compter les non-réponses. Si la DSA décide plus tard de combiner les données provenant de la collecte hebdomadaire de celles provenant de la collecte mensuelle pour la tabulation des estimations de la production agricole, la taille de l'échantillon va doubler. Cependant, comme c'est indiqué dans le rapport de Megill sur les recommandations du plan de sondage, le gain en précision correspondant ne serait que de 8 pourcent en moyenne, à cause des effets de grappes qui sont assez grands. La décision de combiner les données sur la production agricole collectées mensuellement avec celles collectées hebdomadairement dépendra de la grandeur du biais dans les données mensuelles qui est actuellement inconnu. En poursuivant la collecte des données mensuelles, il serait certainement possible, au cours du temps, d'estimer le biais approximatif. Si le biais est petit par rapport au gain en précision provenant d'un échantillon plus grand, les données devraient être combinées et traitées ensemble.

Durant cette mission, Megill a travaillé avec Munyaneza sur le calcul des variances pour les estimations de la production de 1990 à l'aide du logiciel PC CARP. Ces résultats, présentés en Annex A, sont consistants avec ceux produits avant pour les données de 1990. Les coefficients de variation (C.V) pour beaucoup de cultures étaient assez élevés au niveau de la *préfecture*. Des résultats similaires peuvent être attendus à partir du nouvel échantillon basé toujours sur 78 DR.

Bien que l'échantillon de 78 DR devrait produire des résultats fiables au niveau préfectoral, Murekezi a indiqué qu'il est important que la DSA puisse, plus tard, fournir des résultats raisonnables aux niveaux sous-régionaux quand les ressources auront permis d'augmenter le nombre de DR échantillons. L'échantillon de 312 DR correspondant à tous les quatre sous-échantillons devrait fournir des estimations fiables de la production agricole au niveau de la *préfecture*. En général, un minimum de 20 DR échantillons serait requis pour toute zone géographique afin de fournir des estimations modérément fiables. Ceci signifie que pour pouvoir fournir des résultats pour chacune des 32 *sous-préfectures*, un minimum de 640 DR échantillons serait requis dans tout le pays. Quant aux estimations au niveau *communal*, il serait nécessaire de stratifier l'échantillon total par *commune* et de déterminer le nombre de DR échantillons requis pour chaque commune individuellement. Comme il y a 145 *communes*, le nombre moyen de DR par commune est d'environ 43, bien que le nombre de DR varie par *commune*. Puisque le taux de sondage pour les DR approcherait 50 pourcent dans beaucoup de *communes*, le facteur de correction de la population finie au premier degré de sondage réduirait considérablement la variance, réduisant ainsi le nombre de DR échantillons requis dans plusieurs *communes*. Même en tenant en considération le facteur de correction de la population finie au premier degré de sondage, un échantillon d'au moins 2.000 DR serait requis pour fournir des résultats fiables au niveau de la *commune*.

Si un projet quelconque souhaite avoir des résultats fiables pour une *sous-préfecture* ou *commune* donnée, il sera nécessaire d'augmenter le nombre de DR échantillons dans cette zone. Dans un tel cas, la même base de sondage utilisée pour l'Enquête Nationale Agricole (établie à partir des données du recensement général de la population de 1991) peut servir pour sélectionner l'échantillon supplémentaire de DR pour cette zone. Une fois la partie de la base de sondage correspondant à cette zone est extraite, les mêmes procédures de sélection de l'échantillon peuvent être utilisées pour sélectionner un plus grand nombre de DR avec probabilité proportionnelle à la taille (PPT).

### 3. Calcul des Facteurs de Pondération

En vue d'étendre les données des Enquêtes Nationales Agricoles au niveau national ou régional, il est nécessaire d'appliquer un poids (facteur d'extension) aux données provenant de chaque enregistrement de collecte. Le poids fondamental pour un ménage échantillon serait égal à l'inverse de sa probabilité de sélection (calculé en multipliant les probabilités à chaque degré de sélection). Puisque les DR ont été sélectionnés avec probabilité proportionnelle à la taille (PPT), un poids à part devra être calculé pour chaque DR. Au second degré de sélection, un nombre fixé (16) est sélectionné dans chaque DR. La probabilité globale de sélection pour les ménages échantillons dans chaque DR échantillon serait la suivante:

$$p_{hi} = \frac{n_h \times M_{hi}}{M_h} \times \frac{m_{hi}}{M'_{hi}},$$

où:

$p_{hi}$  = probabilité de sélection des ménages échantillons dans le i-ième DR échantillon de la h-ième strate (*préfecture*)

$n_h$  = nombre de DR échantillons sélectionnés dans la strate h

$M_{hi}$  = nombre de ménages recensés en 1991 (mesure de la taille) pour le i-ième DR dans la strate h

$M_h$  = nombre total de ménages recensés en 1991 pour la strate h (mesure de la taille cumulée)

$m_{hi}$  = nombre de ménages échantillons sélectionnés dans le i-ième DR échantillon de la strate h

$M'_{hi}$  = nombre de ménages provenant de la nouvelle liste (liste actualisée) des ménages pour le i-ième DR échantillon dans la strate h

Le poids fondamental ( $W_{hi}$ ) pour chaque ménage échantillon serait donc l'inverse de cette probabilité, à savoir:

$$W_{hi} = \frac{M_h}{n_h \times M_{hi}} \times \frac{M'_{hi}}{m_{hi}},$$

où  $W_{hi}$  est le poids (facteur de pondération) pour chaque ménage échantillon dans le i-ième DR échantillon de la strate h.

Il est à noter que si le nombre de ménages échantillons listés dans un DR échantillon est égal au nombre correspondant de ménages recensés en 1991 (c'est à dire,  $M'_{hi} = M_{hi}$ ) et si le nombre de ménages sélectionnés dans chaque DR échantillon est fixé (par exemple 16), chaque ménage échantillon dans une strate aurait le même poids. Les poids variéeraient légèrement entre les strates vue la façon dont les DR échantillons ont été attribués. En pratique, cependant,  $M'_{hi}$  est rarement égal à  $M_{hi}$ , ce qui fait que les poids variéeront également par DR à l'intérieur de chaque strate.

Un fichier *spreadsheet* de LOTUS a été étab'i pour contenir l'information sur l'échantillonnage et pour le calcul des facteurs de pondération pour les DR échantillons. Puisque la valeur de  $M'_{hi}$  pour chaque DR échantillon sera déterminée en fonction de la nouvelle liste (liste actualisée) des ménages, il sera nécessaire de mettre à jour ce *spreadsheet* après l'actualisation de la liste afin de calculer les facteurs de pondération appropriés. Il est aussi important de comparer la valeur de  $M'_{hi}$  avec celle de  $M_{hi}$  pour chaque DR échantillon, pour être sûr que le DR échantillon a été convenablement couvert. S'il y a une très grande différence (disons au delà de 20 pourcent) entre le nombre de ménages de la liste actualisée et le nombre correspondant trouvé lors du recensement de 1991, il se pourrait qu'on ait mal interprété les limites du DR, ainsi le superviseur devrait vérifier la liste sur terrain. Ce *spreadsheet* sera également mis à jour avec l'information sur le nombre de ménages effectivement enquêtés et il ajustera automatiquement les facteurs de pondération pour les non-réponses, en se basant sur l'ajustement décrit dans les deux prochaines sections. Il est tellement flexible que les facteurs de pondération peuvent être facilement révisés lorsque des sous-échantillons supplémentaires sont inclus dans l'enquête. Une copie de ce *spreadsheet* est incluse en Annex B, de même qu'une liste d'instructions pour la mise à jour du *spreadsheet* pour chaque enquête.

#### 4. Ajustement des Pondérations pour les DR échantillons manquant

Lorsqu'un DR échantillon n'est pas couvert entièrement à cause des problèmes de sécurité, il est possible de le remplacer par un autre DR. Cette procédure est basée sur l'hypothèse que le DR échantillon choisi pour remplacer est représentatif du DR échantillon qu'il remplace. Vu les problèmes qui rendent difficile la couverture du DR échantillon remplacé, ses caractéristiques sont probablement quelque peu différentes de celles des DR qui peuvent être enquêtés sans problème, introduisant ainsi un petit biais dans les résultats d'enquête. Si un DR échantillon manquant n'est pas remplacé, les facteurs de pondération seraient ajustés en se basant sur l'hypothèse que les DR échantillons enquêtés sont représentatifs des DR échantillons manquant, ce qui conduit à un même biais. La raison pour laquelle on remplace les DR échantillons manquant c'est pour maintenir la taille de l'échantillon.

Une façon de remplacer les DR échantillons manquant serait de sélectionner un DR échantillon à partir des sous-échantillons qui ne sont pas utilisés, puisque ces DR ont été sélectionnés en utilisant les mêmes procédures (avec probabilité proportionnelle à la taille). Dans ce cas, il ne serait pas nécessaire d'ajuster les facteurs de pondération. En calculant les facteurs de pondération pour le DR échantillon qui remplace,  $M_h$ ,  $M'_h$  et  $m_h$  correspondraient aux valeurs pour le nouvel DR échantillon. Il est recommandé de sélectionner le DR échantillon remplaçant dans la même commune que le DR échantillon original, lorsque un tel DR échantillon est disponible dans l'autre sous-échantillon; si non, un DR échantillon d'une commune voisine peut être utilisé comme remplaçant. De cette façon la représentativité géographique de l'échantillon sera maintenu. En tout cas, le DR remplaçant devrait être sélectionné dans la même strate (préfecture).

Dans les enquêtes précédentes, les DR échantillons manquant n'ont pas été remplacés. Par exemple pour l'enquête de 1989 seulement 76 des 78 DR échantillons ont été couverts. Si les DR échantillons ne sont pas remplacés dans les enquêtes futures, il sera nécessaire d'ajuster les facteurs de pondération. Dans ce cas, le facteur de pondération fondamental ( $W_h$ ) pour les ménages dans la strate (préfecture) h devrait être multiplié par le facteur suivant:

$$F_{1h} = \frac{n_h}{n'_h}$$

où:

$F_{1h}$  = facteur d'ajustement pour le facteur de pondération fondamental des ménages échantillons dans la strate h

$n'_h$  = nombre de DR échantillons effectivement couverts pendant l'enquête dans la strate h

On peut voir que le facteur de pondération ajusté ( $W'_h$ ) peut être exprimé comme suit:

$$W'_{hi} = W_{hi} \times F_{1h} = \frac{M_h}{n_h \times M_{hi}} \times \frac{M'_{hi}}{m_{hi}}$$

Par conséquent cet ajustement peut être réalisé en remplaçant tout simplement  $n_h$  par  $n'_h$  dans le facteur de pondération.

#### 5. Ajustement des Pondérations pour les ménages de l'échantillon non interviewés

Les procédures actuelles d'enquête incluent le remplacement de chaque ménage non-enquêté par un autre ménage de la même grappe. Pour la collecte hebdomadaire, des grappes de 6 ménages chacune sont sélectionnées: quatre ménages à enquêter et deux pour des remplacements éventuels. Cependant, puisque le fichier des données analysées n'inclut que les ménages qui ont été suivis toute la saison (tous les 6 mois), tout remplacement au cours de la saison implique des données incomplètes aussi bien pour le ménage original (remplacé) que pour le ménage remplaçant. Afin d'éviter les complications dans l'analyse des données d'enquête, les enregistrements pour ces ménages échantillons sont exclus lors de la tabulation

des données. La raison pour laquelle les ménages non-enquêtés sont remplacés est pour assurer que les ménages remplaçant sont bien en place (et sont familiers avec les interviews) pour la saison suivante; ceci permet également de maintenir un volume de travail constant pour l'enquêteur dans chaque DR échantillon.

Après la collecte des données pour une saison, la tabulation de la production agricole totale par mois pour la saison est effectuée en excluant les enregistrements pour les ménages remplaçant et pour ceux remplacés (avec moins de 6 mois de données). A la fin de l'année agricole (Septembre à Août), les données provenant des deux saisons sont combinées, et la tabulation de la production agricole est effectuée pour les ménages ayant 12 mois complets de données. Dans ce cas, il est nécessaire d'ajuster les facteurs de pondération séparément pour la tabulation des données par saison et pour l'année. Dans chaque cas, les procédures d'ajustement seront similaires. Il est recommandé de faire l'ajustement des facteurs de pondération pour les ménages non-interviewés au niveau du DR.

L'ajustement du facteur de pondération dépend aussi de la raison pour la non-réponse. Le facteur de pondération devrait être ajusté pour tenir compte des ménages qui ont refusé de répondre, des absences temporaires (pas à la maison), etc. Si une habitation est abandonnée et qu'aucun autre ménage n'est venu l'occuper, cette habitation ne peut pas être une unité d'échantillonnage valide. Le deuxième ajustement du facteur de pondération (pour les ménages non-enquêtés) est défini comme suit:

$$F_{2hi} = \frac{m'_{hi}}{m''_{hi}},$$

où:

$F_{2hi}$  = facteur d'ajustement pour le facteur de pondération des ménages dans le i-ième DR échantillon dans la strate h

$m'_{hi}$  = nombre d'unités de sondage (ménages) valides sélectionnés dans le i-ième DR échantillon dans la strate h ( en excluant les habitations abandonnées dans l'échantillon)

$m''_{hi}$  = nombre de ménages avec des données pour toute la période de référence (ménages inclus dans la tabulation des données)

Le facteur de pondération final est calculé comme suit:

$$W''_{hi} = W'_{hi} \times F_{2hi}$$

Pour que l'échantillon reste représentatif dans le temps, il est important de mettre à jour les listes des ménages des DR échantillons au moins tous les 2 ans. Puisque la valeur de  $M'_{hi}$  va changer pour refléter l'augmentation (ou la diminution) du nombre de ménages, les facteurs de pondération seront mis à jour en conséquence.

Après chaque enquête il sera nécessaire de mettre à jour le fichier *spreadsheet* de LOTUS utilisé pour maintenir l'information d'échantillonnage et pour calculer les facteurs de pondération, avec le nombre exact de ménages enquêtés. Ce *spreadsheet* recalculera ensuite les facteurs de pondération basés sur l'information mise à jour (voir Annexe B).

## 6. Ajustement Facultatif des Facteurs de Pondération basés sur des Projections de la Population

Etant donné que les facteurs de pondération  $W''_{hi}$  sont basés sur les probabilités de sélection, ils produiront des estimations non biaisées (en supposant que les listes des ménages sont précises). Cependant, s'il est possible d'obtenir de bonnes projections chaque année sur le nombre de ménages selon la répartition rural, rural de fait /urbain de droit et rural de droit/urbain de fait pour les

DR dans chaque préfecture, un ajustement facultatif (en plus de  $F_{1h}$  et  $F_{2h}$ ) pour les ménages échantillons dans chaque préfecture pourraient être calculés comme suit:

$$F_{3h} = \frac{M'_h}{\hat{M}_h},$$

où:

$M'_h$  = nombre projeté des ménages ruraux dans la strate (préfecture)h

$$\hat{M}_h = \sum_{i=1}^{n'_h} m''_{hi} \times W''_{hi}$$

Le dénominateur dans  $F_{3h}$  correspond au nombre total des ménages pondérés dans la strate à partir des données de l'enquête. Ce facteur d'ajustement serait appliqué aux facteurs de pondération de tous les ménages échantillons dans la strate correspondante. En supposant que le nombre projeté des ménages ruraux est précis, ce type de proportion d'ajustement a comme effet la réduction de l'erreur d'échantillonnage. Le nombre projeté de ménages ( $M'_h$ ) devrait probablement être estimé en obtenant d'abord les projections de la population pour chaque préfecture et en les divisant par la taille moyenne des ménages donnée par le Recensement Général de la Population et de l'Habitat. Puisque les projections de la population aux niveaux régionaux sont particulièrement sujettes à des erreurs, et que la population ciblée ne concerne que les milieux ruraux et semi-ruraux dans chaque préfecture, la DSA devrait être prudente dans l'utilisation du facteur d'ajustement  $F_{3h}$  du facteur de pondération. Si les projections de la population sont imprécises, cet ajustement pourrait biaiser les résultats de l'enquête.

## 7. Estimations de l'Enquête

Quelques unes des plus importantes estimations de l'enquête seront sous forme des totaux, telle que la production agricole totale d'une culture. L'estimation d'enquête d'un total sera obtenu de la manière suivante:

$$\hat{Y} = \sum_{h=1}^{10} \sum_{i=1}^{n'_h} \sum_{j=1}^{m''_{hi}} W''_{hi} \times Y_{hij},$$

où  $Y_{hij}$  = valeur de la variable Y pour le j-ième ménage échantillon dans le i-ième DR échantillon dans la strate h

Dans le cas d'une estimation d'enquête du nombre total de ménages ayant une caractéristique particulière, la variable Y peut être définie comme suit:

Y = 1 si le ménage a une caractéristique particulière  
= 0 autrement

D'autres estimations d'enquête peuvent être sous forme de proportions, définies comme suit:

$$\hat{R} = \frac{\hat{Y}}{\hat{X}},$$

où  $\hat{Y}$  et  $\hat{X}$  sont des estimations pondérées des totaux, calculées comme indiqué ci-dessus.

Les estimations des moyennes et des proportions sont des types spéciaux de ratios. Dans le cas de l'estimation de la moyenne, la variable X au dénominateur est égale à 1 pour chaque ménage, ainsi le total correspondant est égal à la somme des pondérations. Pour l'estimation de la proportion des ménages ayant une certaine caractéristique, la variable au numérateur est égale à 0 ou 1, suivant qu'un ménage échantillon a cette caractéristique ou pas.

### 8. Calcul des Variances pour les Estimations de l'Enquête

Dans les publications des résultats d'enquête, il est important d'inclure un chapitre sur la précision des données de l'enquête. En plus d'une présentation des tableaux des erreurs de sondage pour la plupart des estimations importantes de l'enquête, les différentes sources d'erreurs non dues à l'échantillonnage devraient être décrites.

L'écart type, ou la racine carré de la variance, est utilisé pour mesurer l'erreur de sondage, bien qu'il pourrait aussi inclure une petite partie de l'erreur non due à l'échantillonnage. L'estimateur de la variance devrait tenir compte des différents aspects du plan de sondage, telles que la stratification et les grappes. Un des logiciels de micro ordinateur qui peut être utilisé pour calculer les variances des estimations d'enquête basées sur un échantillon stratifié à deux degrés telles celles de l'Enquête Nationale Agricole est PC CARP. Ce logiciel peut être utilisé pour calculer les variances des totaux, moyennes, proportions et ratios, aussi bien pour les coefficients de régression. Dans le cas des estimations des sous populations, les variances peuvent être calculées pour chaque catégorie d'une variable de classification, et il est possible de classer différentes variables en les croisant. PC CARP utilise un estimateur de la variance d'un échantillon en grappes. Une copie de PC CARP est disponible à la DSA. Ce logiciel exige un micro ordinateur de type IBM ou IBM compatible avec un math coprocessor; la plupart des micro ordinateurs à la DSA remplissent cette condition. Il est recommandé d'utiliser une machine plus rapide (avec un chip de 386), spécialement lorsqu'on utilise un grand fichier de données ou lorsque plusieurs classifications croisées sont incluses dans l'analyse, vu que certaines étapes de PC CARP pourraient exiger un temps considérable à la machine.

Les formules suivantes sont utilisées pour estimer la variance de l'échantillon stratifié et en grappes:

(1) Variance pour l'estimation d'enquête d'un total

$$VAR(\hat{Y}) = \sum_{h=1}^{10} \left[ \frac{n'_h}{(n'_h-1)} \times \sum_{f=1}^{n'_h} \left( \hat{Y}_{hf} - \frac{\hat{Y}_h}{n'_h} \right)^2 \right],$$

où:

$$\hat{Y}_{hf} = \sum_{j=1}^{m''_{hf}} W''_{hf} \times Y_{hfj}$$

$$\hat{Y}_h = \sum_{f=1}^{n'_h} \sum_{j=1}^{m''_{hf}} W''_{hf} \times Y_{hfj}$$

(2) Variance pour l'estimation d'enquête d'un ratio

$$VAR(\hat{R}) = VAR\left(\frac{\hat{Y}}{\hat{X}}\right) = \frac{1}{\hat{X}^2} [VAR(\hat{Y}) + \hat{R}^2 \times VAR(\hat{X}) - 2 \times \hat{R} \times COV(\hat{X}, \hat{Y})],$$

où:

$$COV(\hat{Y}, \hat{X}) = \sum_{h=1}^{10} \left[ \frac{n'_h}{(n'_h-1)} \times \sum_{i=1}^{n'_h} \left( X_{hi} - \frac{X'_h}{n'_h} \right) \times \left( Y_{hi} - \frac{Y'_h}{n'_h} \right) \right]$$

$Var(Y)$  et  $Var(X)$  sont des variances des estimations des totaux, calculées en utilisant la formule spécifiée précédemment.

Les fichiers d'entrée de données pour PC CARP doivent commencer avec les données suivantes sur l'enregistrement pour chaque ménage échantillon: (1) numéro de la strate; (2) numéro du DR; et (3) facteur de pondération. Pour les données de 1989 et 1990 il y a 21 strates identifiées dans les fichiers de données; ces strates sont décrites dans le rapport de Karen Stanecki's intitulé "Sample Design and Variance Estimation Procedures for the Rwanda Crop Forecasting Survey" (Avril 1985). Il a été nécessaire de combiner deux strates (16 et 17) qui n'avaient qu'un seul DR chacun. Dans le cas du nouveau plan de sondage, il y a 10 strates, correspondant aux préfectures. Ces données sur chaque enregistrement seraient suivies par les variables de classification, telles que préfecture ou classe de taille des ménages, et les variables d'analyse à utiliser dans les tabulations, telle que la production agricole pour chaque culture. Puisque PC CARP ne peut pas transformer les données (en ajoutant des variables, par exemple), les fichiers d'entrée de données de PC CARP devraient être générés avec les données déjà transformées.

PC CARP a été utilisé précédemment pour calculer les erreurs de sondage pour les estimations de la production agricole de certaines cultures à partir des données de 1989. Ces erreurs de sondage se trouvent dans l'Annex A du précédent rapport technique de Megill sur les recommandations d'échantillonnage. Pendant la présente visite, Megill a travaillé encore avec Muryaneza pour le calcul des erreurs de sondage à partir des données de 1990 en utilisant PC CARP. Ces erreurs de sondage sont présentées en Annex A du présent document. Muryaneza devrait maintenant avoir suffisamment d'expérience pour générer des fichiers d'entrée de données de PC CARP et calculer les erreurs de sondage pour les données futures.

L'output de PC CARP n'est pas facile à lire et interpréter, puisque les nombres sont présentés en notation scientifique, les noms des variables ne peuvent être de plus de 8 caractères en longueur, et les catégories pour les variables de classification sont uniquement identifiées par leurs codes. Par conséquent, il est recommandé que l'output de PC CARP soit transformé en un format publiable en utilisant par exemple LOTUS. Un tel format est illustré par les tableaux en Annexe A, qui comprennent le nombre d'observations, la valeur de l'estimation, l'erreur de sondage, le coefficient de variation (C.V), l'intervalle de confiance à 95 pourcent et l'effet du plan de sondage (*design effect*).

ANNEXE A

Tableau 1.1. Estimations de la production totale pour certaines cultures avec les écarts types correspondants, les coefficients de variation, les intervalles de confiance de 95% et les effets du plan de sondage (Design Effects:DEFF), basés sur les données de l'Enquête sur la Production Agricole de 1990  
RWANDA

Nombre d'observations = 1,208

Culture	Estimat. de la Production (Tonnes)	Ecart Type (Tonnes)	C.V. (%)	Interv.de Conf. de 95%		DEFF
				Limite Inférieure	Limite Supérieure	
<i>Haricots Total</i>	205,907	16,103	7.8%	174,345	237,469	4.00
<i>Pois Total</i>	11,036	1,200	10.9%	8,683	13,388	2.99
<i>Sorgho</i>	141,835	23,131	16.3%	96,498	187,172	5.82
<i>Mais Total</i>	95,973	18,624	19.4%	59,469	132,477	3.62
<i>Patate Douce</i>	819,279	61,040	7.5%	699,642	938,916	8.72
<i>Manioc Total</i>	264,952	27,393	10.3%	211,262	318,642	4.45
<i>Pomme de Terre</i>	285,030	128,688	45.1%	32,805	537,261	6.34
<i>Banane a Cuire</i>	602,793	58,522	9.7%	488,090	717,496	2.54
<i>Banane a Bière</i>	1,916,750	145,777	7.6%	1,631,027	2,202,473	4.40
<i>Café</i>	40,971	9,003	22.0%	23,326	58,616	9.93
<i>Blé</i>	7,389	2,742	37.1%	2,014	12,765	5.68
Nombre Total de Ménages	1,271,900	44,204	3.5%	1,185,260	1,358,540	-

Tableau 1.2. Estimations de la production totale pour certaines cultures avec les écarts types correspondants, les coefficients de variation, les intervalles de confiance de 95% et les effets du plan de sondage (Design Effects:DEFF), basés sur les données de l'Enquête sur la Production Agricole de 1990 BUTARE

Nombre d'observations = 151

Culture	Estimat. de la Production (Tonnes)	Ecart Type (Tonnes)	C.V. (%)	Interv.de Conf.de 95%		DEFF
				Limite Inférieure	Limite Supérieure	
<i>Haricots Total</i>	14,852	4,313	29.0%	6,398	23,306	5.34
<i>Pois Total</i>	890	370	41.6%	165	1,614	4.87
<i>Sorgho</i>	14,089	4,531	32.2%	5,209	22,969	4.36
<i>Mais Total</i>	2,123	1,184	55.8%	-199	4,444	7.26
<i>Patate Douce</i>	122,828	20,404	16.6%	82,837	162,819	3.13
<i>Manioc Total</i>	43,798	5,714	13.0%	32,598	54,998	0.88
<i>Pomme de Terre</i>	4,899	1,255	25.6%	2,439	7,359	1.88
<i>Banane a Cuire</i>	22,835	6,779	29.7%	9,548	36,121	4.29
<i>Banane a Bière</i>	194,903	42,941	22.0%	110,738	279,068	4.07
<i>Café</i>	3,155	1,169	37.1%	863	5,446	5.53
<i>Blé</i>	-	-	-	-	-	-
Nombre Total de Ménages	156,274	15,818	10.1%	125,271	187,277	1.73

Tableau 1.3. Estimations de la production totale pour certaines cultures avec les écarts types correspondants, les coefficients de variation, les intervalles de confiance de 95% et les effets du plan de sondage (Design Effects:DEFF), basés sur les données de l'Enquête sur la Production Agricole de 1990  
BYUMBA

Nombre d'observations = 126

Culture	Estimat. de la Pro duction (Tons)	Ecart Type (Tons)	C.V. (%)	Inter.de Conf. de 95%		DEFF
				Limite Inférieure	Limite Supérieure	
<i>Haricots Total</i>	37,581	8,136	21.6%	21,636	53,527	2.63
<i>Pois Total</i>	1,578	394	25.0%	806	2,350	1.95
<i>Sorgho</i>	43,172	18,867	43.7%	6,192	80,151	5.84
<i>Mais Total</i>	8,268	2,568	31.1%	3,235	13,300	3.00
<i>Patate Douce</i>	91,468	19,035	20.8%	54,158	128,777	4.02
<i>Manioc Total</i>	18,436	3,415	18.5%	11,742	25,129	0.67
<i>Pomme de Terre</i>	10,353	2,520	24.3%	5,414	15,292	1.21
<i>Banane a Cuire</i>	71,180	5,860	8.2%	59,695	82,664	0.40
<i>Banane a Bière</i>	223,214	41,196	18.5%	142,470	303,958	2.21
<i>Café</i>	2,283	1,108	48.5%	111	4,456	3.26
<i>Blé</i>	1,916	1,390	72.5%	-808	4,640	6.23
<b>Nombre Total de ménages</b>	<b>141,439</b>	<b>13,909</b>	<b>9.8%</b>	<b>114,178</b>	<b>168,700</b>	<b>1.46</b>

Tableau 1.4. Estimations de la production totale pour certaines cultures avec les écarts types correspondants, les coefficients de variation, les intervalles de confiance de 95% et les effets du plan de sondage (Design Effects:DEFF), basés sur les données de l'Enquête sur la Production Agricole de 1990 CYANGUGU

Nombre d'observations = 94

Culture	Estimat. de la Production (Tons)	Ecart Type (Tons)	C.V. (%)	Inter.de Conf. de 95%		DEFF
				Limite Inférieure	Limite Supérieure	
<i>Haricots Total</i>	6,682	2,352	35.2%	2,072	11,292	6.61
<i>Pois Total</i>	566	245	43.2%	87	1,046	2.66
<i>Sorgho</i>	1,113	492	44.2%	148	2,078	1.26
<i>Mais Total</i>	5,153	1,431	27.8%	2,349	7,957	2.71
<i>Patate Douce</i>	35,591	11,620	32.7%	12,815	58,366	5.90
<i>Manioc Total</i>	23,382	9,440	40.4%	4,879	41,884	5.27
<i>Pomme de Terre</i>	3,058	1,036	33.9%	1,028	5,087	2.36
<i>Banane a Cuire</i>	37,056	15,308	41.3%	7,052	67,059	6.06
<i>Banane a Bière</i>	96,430	34,814	36.1%	28,194	164,666	5.13
<i>Café</i>	5,740	2,920	50.9%	17	11,463	7.86
<i>Blé</i>	33	33	100.0%	-32	97	0.39
<b>Nombre Total de ménages</b>	<b>90,281</b>	<b>15,205</b>	<b>16.8%</b>	<b>60,479</b>	<b>120,083</b>	<b>2.62</b>

Tableau 1.5. Estimations de la production totale pour certaines cultures avec les écarts types correspondants, les coefficients de variation, les intervalles de confiance de 95% et les effets du plan de sondage (Design Effects:DEFF), basés sur les données de l'Enquête sur la Production Agricole de 1990 GIKONGORO

Nombre d'observations = 91

Culture	Estimat. de la Production (Tons)	Ecart Type (Tons)	C.V. (%)	Inter.de Conf. de 95%		DEFF
				Limite Inférieure	Limite Supérieure	
Haricots Total	2,905	527	18.1%	1,872	3,937	1.08
Pois Total	1,178	433	36.7%	330	2,026	2.22
Sorgho	4,616	767	16.6%	3,113	6,118	1.02
Mais Total	3,066	1,243	40.6%	629	5,503	3.44
Patate Douce	79,021	17,332	21.9%	45,051	112,991	2.96
Manioc Total	5,504	1,733	31.5%	2,107	8,901	2.57
Pomme de Terre	8,032	3,115	38.8%	1,927	14,138	2.76
Banane a Cuire	3,085	521	16.9%	2,064	4,107	0.98
Banane a Bière	50,043	4,940	9.9%	40,361	59,724	0.25
Café	1,436	400	27.9%	652	2,221	1.41
Blé	923	464	50.3%	13	1,833	2.82
Nombre Total de ménages	100,963	7,184	7.1%	86,882	115,044	0.53

Tableau 1.6. Estimations de la production totale pour certaines cultures avec les écarts types correspondants, les coefficients de variation, les intervalles de confiance de 95% et les effets du plan de sondage (Design Effects:DEFF), basés sur les données de l'Enquête sur la Production Agricole de 1990 GISENYI

Nombre d'observations = 128

Culture	Estimat. de la Production (Tons)	Ecart Type (Tons)	C.V. (%)	Inter.de Conf. de 95%		DEFF
				Limite Inférieure	Limite Supérieure	
<i>Haricots Total</i>	9,007	3,350	37.2%	2,441	15,574	6.66
<i>Pois Total</i>	866	348	40.2%	184	1,548	1.91
<i>Sorgho</i>	1,431	879	61.5%	-293	3,154	4.33
<i>Mais Total</i>	30,826	16,664	54.1%	-1,834	63,486	3.63
<i>Patate Douce</i>	66,665	22,867	34.3%	21,846	111,484	7.10
<i>Manioc Total</i>	5,963	1,987	33.3%	2,069	9,858	1.93
<i>Pomme de Terre</i>	68,005	39,149	57.6%	-8,727	144,736	5.20
<i>Banane a Cuire</i>	16,598	7,775	46.8%	1,359	31,837	8.63
<i>Banane a Bière</i>	124,842	73,412	58.8%	-19,045	268,729	6.66
<i>Café</i>	8,525	7,368	86.4%	-5,916	22,965	16.38
<i>Blé</i>	256	167	65.1%	-71	583	1.08
<b>Nombre Total de ménages</b>	<b>124,690</b>	<b>15,953</b>	<b>12.8%</b>	<b>93,422</b>	<b>155,958</b>	<b>2.15</b>

Tableau 1.7. Estimations de la production totale pour certaines cultures avec les écarts types correspondants, les coefficients de variation, les intervalles de confiance de 95% et les effets du plan de sondage (Design Effects:DEFF), basés sur les données de l'Enquête sur la Production Agricole de 1990  
GITARAMA

Nombre d'observations = 155

Culture	Estimat. de la Production (Tons)	Ecart Type (Tons)	C.V. (%)	Inter.de Conf. de 95%		DEFF
				Limite Inférieure	Limite Supérieure	
<i>Haricots Total</i>	21,069	3,098	14.7%	14,998	27,141	2.41
<i>Pois Total</i>	998	257	25.7%	494	1,502	2.34
<i>Sorgho</i>	10,474	2,327	22.2%	5,912	15,035	2.52
<i>Mais Total</i>	3,106	794	25.6%	1,549	4,663	2.58
<i>Patate Douce</i>	125,313	17,453	13.9%	91,105	159,521	2.20
<i>Manioc Total</i>	66,675	19,666	29.5%	28,129	105,220	5.59
<i>Pomme de Terre</i>	4,929	1,206	24.5%	2,565	7,293	2.41
<i>Banane a Cuire</i>	42,993	6,660	15.5%	29,940	56,046	1.86
<i>Banane a Bière</i>	327,479	50,498	15.4%	228,503	426,455	2.54
<i>Café</i>	7,450	2,551	34.2%	2,450	12,449	3.57
<i>Blé</i>	687	678	98.6%	-641	2,016	1.98
Total Number of Households	164,541	10,674	6.5%	143,619	185,463	0.75

Tableau 1.8. Estimations de la production totale pour certaines cultures avec les écarts types correspondants, les coefficients de variation, les intervalles de confiance de 95% et les effets du plan de sondage (Design Effects:DEFF), basés sur les données de l'Enquête sur la Production Agricole de 1990 KIBUNGO

Nombre d'observations = 95

Culture	Estimat. de la Production (Tons)	Ecart Type (Tons)	C.V. (%)	Inter.de Conf. de 95%		DEFF
				Limite Inférieure	Upper Supérieure	
<i>Haricots Total</i>	25,936	3,682	14.2%	18,719	33,153	1.37
<i>Pois Total</i>	662	245	37.1%	181	1,143	1.92
<i>Sorgho</i>	20,079	6,327	31.5%	7,677	32,480	3.59
<i>Mais Total</i>	3,560	677	19.0%	2,234	4,886	1.11
<i>Patate Douce</i>	44,857	11,238	25.1%	22,830	66,885	3.06
<i>Manioc Total</i>	22,753	7,738	34.0%	7,587	37,920	3.32
<i>Pomme de Terre</i>	3,720	1,170	31.4%	1,427	6,012	1.46
<i>Banane a Cuire</i>	246,683	36,289	14.7%	175,556	317,810	1.10
<i>Banane a Bière</i>	333,971	35,753	10.7%	263,895	404,047	0.61
<i>Café</i>	4,774	1,737	36.4%	1,370	8,178	2.71
<i>Blé</i>	312	217	69.4%	-113	738	0.51
<b>Nombre Total de ménages</b>	<b>96,650</b>	<b>15,337</b>	<b>15.9%</b>	<b>66,590</b>	<b>126,709</b>	<b>2.50</b>

Tableau 1.9. Estimations de la production totale pour certaines cultures avec les écarts types correspondants, les coefficients de variation, les intervalles de confiance de 95% et les effets du plan de sondage (Design Effects:DEFF), basés sur les données de l'Enquête sur la Production Agricole de 1990 KIBUYE

Nombre d'observations = 88

Culture	Estimat. de la Production (Tons)	Ecart Type (Tons)	C.V. (%)	Inter.de Conf. de 95%		DEFF
				Limite Inférieure	Limite Supérieure	
<i>Haricots Total</i>	6,902	545	7.9%	5,833	7,971	0.36
<i>Pois Total</i>	1,916	292	15.2%	1,344	2,488	0.85
<i>Sorgho</i>	3,455	1,890	54.7%	-249	7,158	7.51
<i>Mais Total</i>	18,663	5,293	28.4%	8,288	29,037	1.90
<i>Patate Douce</i>	57,771	6,252	10.8%	45,517	70,025	0.81
<i>Manioc Total</i>	7,960	2,616	32.9%	2,832	13,088	2.53
<i>Pomme de Terre</i>	12,197	1,236	10.1%	9,774	14,620	0.18
<i>Banane a Cuire</i>	5,487	3,144	57.3%	-676	11,650	7.35
<i>Banane a Bière</i>	39,570	15,118	38.2%	9,939	69,201	5.17
<i>Café</i>	175	89	50.8%	1	350	1.51
<i>Blé</i>	676	255	37.7%	177	1,176	2.05
<b>Nombre Total de ménages</b>	<b>91,620</b>	<b>3,687</b>	<b>4.0%</b>	<b>84,394</b>	<b>98,846</b>	<b>0.15</b>

variation, les intervalles de confiance de 95% et les effets du plan de sondage (Design Effects:DEFF), basés sur les données de l'Enquête sur la Production Agricole de 1990 KIGALI

Nombre d'observations = 154

Culture	Estimat. de la Production (Tons)	Ecart - Type (Tons)	C.V. (%)	Inter.de Conf. de 95%		DEFF
				Limite Inférieure	Upper Supérieure	
<i>Haricots Total</i>	55,385	9,076	16.4%	37,597	73,174	2.70
<i>Pois Total</i>	865	357	41.3%	165	1,565	2.48
<i>Sorgho</i>	27,365	7,631	27.9%	12,408	42,321	4.26
<i>Mais Total</i>	4,072	973	23.9%	2,164	5,979	3.28
<i>Patate Douce</i>	80,805	14,564	18.0%	52,259	109,351	3.85
<i>Manioc Total</i>	64,520	11,769	18.2%	41,454	87,587	2.79
<i>Pomme de Terre</i>	5,862	3,287	56.1%	-580	12,304	4.48
<i>Banane a Cuire</i>	126,076	38,948	30.9%	49,738	202,414	8.36
<i>Banane a Bière</i>	393,032	65,479	16.7%	264,693	521,371	2.72
<i>Café</i>	6,976	2,425	34.8%	2,223	11,730	4.68
<i>Blé</i>	233	117	50.4%	3	463	0.91
Nombre Total de ménages	164,911	19,410	11.8%	126,868	202,954	2.49

Tableau 1.11. Estimations de la production totale pour certaines cultures avec les écarts types correspondants, les coefficients de variation, les intervalles de confiance de 95% et les effets du plan de sondage (Design Effects:DEFF), basés sur les données de l'Enquête sur la Production Agricole de 1990 RUHENGARI

Nombre d'observations = 126

Culture	Estimat. de la Production (Tons)	Ecart Type (Tons)	C.V. (%)	Inter.de Conf. de 95%		DEFF
				Limite Inférieure	Limite Supérieure	
<i>Haricots Total</i>	25,587	7,187	28.1%	11,500	39,674	2.92
<i>Pois Total</i>	1,517	664	43.8%	215	2,819	5.35
<i>Sorgho</i>	16,044	7,050	43.9%	2,226	29,862	4.32
<i>Mais Total</i>	17,138	5,249	30.6%	6,849	27,427	4.52
<i>Patate Douce</i>	114,961	36,149	31.4%	44,109	185,813	6.05
<i>Manioc Total</i>	5,962	4,243	71.2%	-2,354	14,277	7.68
<i>Pomme de Terre</i>	163,979	122,450	74.7%	-76,023	403,981	6.42
<i>Banane a Cuire</i>	30,802	12,702	41.2%	5,905	55,698	4.37
<i>Banane a Bière</i>	133,262	52,343	39.3%	30,671	235,853	7.72
<i>Café</i>	457	276	60.4%	-84	997	4.17
<i>Blé</i>	2,353	2,181	92.7%	-1,923	6,628	8.27
<b>Nombre Total de ménages</b>	<b>140,529</b>	<b>15,282</b>	<b>10.9%</b>	<b>110,577</b>	<b>170,481</b>	<b>1.77</b>

## ANNEX B

## RWANDA NATIONAL AGRICULTURAL SURVEY

## SPREADSHEET FOR MAINTAINING SAMPLING INFORMATION AND CALCULATING WEIGHTS

## SUBSAMPLE NO. 2

PREF.	COMM.	NAME OF COMMUNE	SECT.	NAME OF SECTEUR	DR	ZONE	R_U	TOTAL NO. OF POP.	NO. OF HH'S IN FRAME FOR DR	NO. OF HH'S LISTED IN DR	NO. OF HH'S IN SAMPLE FOR DR	BASIC SAMPLING WEIGHT	NO. OF VALID SAMPLE HH'S	NO. OF COMPLETED INTERVIEWS	NON-RESP. ADJ.	FINAL WEIGHT
1	1	GISHAMVU	10	SHOLI	25	700	0	906	204							
1	5	MARABA	3	GISAKURA	21	700	0	1047	217							
1	15	NYAKIZU	7	MARABA	20	700	0	881	193							
1	18	RUNYINYA	7	MBASA	16	700	0	892	185							
1	10	MUYIRA	5	MUNYINYA	11	800	0	1341	270							
1	19	RUSATIRA	2	GAHANA	10	807	0	1200	260							
1	8	MUGUSA	3	CYAYI	33	809	0	952	229							
1	13	NTYAZO	12	NYAHURE	11	809	0	1131	274							
1	9	MUYAGA	7	NYERANZI	28	900	0	1276	310							
2	3	CYUMBA	7	MUSENDA	2	600	0	1031	216							
2	7	KIBALI	8	RUHENDA	19	600	0	1229	256							
2	12	MUKARANGE	1	BUGWE	5	600	0	989	213							
2	1	BUYOGA	3	KAVUMU	29	607	0	1009	206							
2	2	BWISIGE	3	GIHUKU	31	1106	0	1016	216							
2	16	RUTARE	7	MWENDO	23	1106	0	1180	241							
2	14	MUVUMBA	5	RUKOMO	2	1206	0	1615	353							
2	13	MURAMBI	1	GAKENKE	74	1211	0	1127	215							
2	15	NGARAMA	1	GAKOMA	17	1211	0	1530	291							
3	1	BUGARAMA	4	KIBANGIRA	3	102	0	1379	357							
3	3	GAFUNZO	4	GABIRO	3	200	0	1196	242							
3	6	GISUMA	5	GASHIRABWOBA	32	200	0	1062	196							
3	9	KARENGERA	3	GASUMO	10	201	0	1250	257							
3	10	KIRAMBO	2	GAHISI	30	300	0	1324	287							
3	7	KAGANO	10	RAMBIRA	37	302	0	1135	216							
4	4	KIVU	7	RWISHWA	13	500	0	1119	274							
4	9	MUSEBEYA	5	GIKUNGU	24	500	0	1104	260							
4	13	RWAMIKO	4	MATA	3	500	0	1348	271							
4	7	MUKO	2	CYOBE	17	507	0	1243	258							
4	1	KARAMA	1	CYANIKA	13	700	0	1319	273							
4	8	MUSANGE	5	KIBAGA	15	700	0	1177	239							
5	5	KAYOVE	8	KINUNU	44	305	0	1094	217							
5	8	NYANYUMBA	12	RUSHUBI	3	305	0	1104	244							
5	10	RUBAVU	8	MURARA	22	403	0	1195	286							
5	1	GASEKE	3	GISEBEYA	18	500	0	1195	239							
5	4	KARAGO	7	RAMBURA	24	500	0	1225	261							
5	2	GICIYE	1	BIREMBO	49	507	0	1259	294							
5	9	RAMBA	6	NYAMPILI	6	507	0	1152	235							
5	12	SATINSYI	4	HINDIRO	10	705	0	1091	231							
6	12	NYAKABANDA	8	NGARU	37	607	0	1387	315							
6	11	NYABIKENKE	10	RUKARAGATA	34	700	0	1000	204							
6	3	KIGOMA	1	BUTARE	41	800	0	986	225							
6	6	MUKINGI	10	RUTAGARA	6	800	0	1199	239							
6	10	NTONGWE	9	NYAKABUNGO	22	800	0	1157	252							
6	13	NYAMABUYE	14	TAKWE	12	800	0	1401	279							
6	4	MASANGO	7	NYAKOGO	15	807	0	1072	213							
6	9	MUSHUBATI	6	KADUHA	23	807	0	896	176							
6	5	MUGINA	6	MUKINGA	5	809	0	1567	299							
7	11	SAKE	5	MURWA	24	1011	0	1263	298							
7	1	BIRENGA	12	SAKARA	7	1100	0	1119	210							
7	4	KIGARAMA	8	REHERA	30	1100	0	1344	262							
7	10	RUTONDE	9	SOVU	24	1100	0	851	201							
7	3	KAYONZA	7	RWINKWAVU	8	1112	0	1459	331							
7	9	RUSUMO	2	GISENYI	60	1112	0	1091	230							
7	9	RUSUMO	9	NYAMUGARI	73	1112	0	1440	384							

## SUBSAMPLE NO. 2 (Continued)

PREF.	COMM.	NAME OF COMMUNE	SECT.	NAME OF SECTEUR	DR	ZONE	R_U	TOTAL NO. OF POP.	NO. OF HH'S IN FRAME FOR DR	NO. OF HH'S LISTED IN DR	NO. OF HH'S IN SAMPLE FOR DR	BASIC SAMPLING WEIGHT	NO. OF VALID SAMPLE HH'S	NO. OF COMPLETED INTERVIEWS	NON-RESP. ADJ.	FINAL WEIGHT
-------	-------	-----------------	-------	-----------------	----	------	-----	-------------------	-----------------------------	--------------------------	------------------------------	-----------------------	--------------------------	-----------------------------	----------------	--------------

8	1	BWAKIRA	3	MUGUNDA	9	705	0	1548	317
8	2	GISHYITA	8	NGOMA	7	305	0	1775	372
8	3	GISOVU	5	KAVUMU	23	503	0	1255	277
8	4	GITESI	12	RURAGWE	29	503	0	922	197
8	8	RUTSIRO	3	GASOVU	35	503	0	1182	236
8	5	KIVUMU	8	NGOBAGOBA	31	507	0	1224	265
9	1	BICUMBI	9	MMULIRE	71	1110	0	1070	208
9	9	MUGAMBAZI	8	NTYABA	29	700	0	1279	257
9	14	RUTONGO	5	KABUYE	27	700	0	1824	413
9	16	TARE	1	BUMBA	7	706	0	1046	205
9	7	KANZENZE	11	MWOGO	2	910	0	1295	290
9	11	NGENDA	8	NYARUGENGE	67	1009	0	1368	309
9	3	GASHORA	8	RILIMA	34	1011	0	1420	285
9	5	GIKORO	2	DUHA	35	1100	0	1290	259
9	12	RUBUNGO	4	KINYINYA	2	1100	0	1393	324
11	7	KINIGI	3	KIHORA	19	400	0	1058	214
11	6	KIGOMBE	3	GAHONDOGO	107	400	1	600	129
11	5	KIDAHU	5	CYANIKA	18	406	0	1227	263
11	12	NYAKINAMA	4	KABERE III	21	504	0	929	209
11	1	BUTARO	4	KAYANGE	41	600	0	1259	302
11	3	CYERU	12	RUHOMBO	28	600	0	1059	206
11	16	RUHONDO	4	MUKONO	19	607	0	1057	222
11	4	GATONDE	3	GAHANGA	19	706	0	1152	234
11	15	NYARUTOVU	3	GASHENYI	47	706	0	1122	244

SAMPLING PARAMETERS BY STRATUM

STRATUM (PREF.)	NO. OF HH'S IN 1991 CENSUS FRAME	NO. OF DR'S SELECTED IN STRATUM
1	160550	9
2	162939	9
3	103818	6
4	100119	4
5	150163	8
6	170789	9
7	136202	7
8	97042	6
9&10	193398	9
11	161191	9

Instructions pour la Mise à jour du Spreadsheet pour le calcul des Facteurs de Pondération

1. Les huit premières colonnes contiennent une information provenant de la base de sondage et ne devraient pas changer. Par conséquent ces colonnes du *spreadsheet* devraient être protégées.
2. La colonne "No. of Hh's Listed in DR" devrait être mise à jour après l'actualisation des listes des ménages des DR. Dans le présent rapport cette valeur est appelée  $M'_h$ . Cette information devrait être mise à jour au moins une fois tous les deux ans.
3. La colonne "No. of Hh's in Sample for DR" devrait être 16 d'après le plan de sondage, mais ceci pourrait varier selon l'enquête. Dans le présent rapport nous l'appelons  $m_h$ . Si les données collectées hebdomadairement sont combinées avec celles collectées mensuellement,  $m_h$  sera égale à 32 d'après le plan de sondage actuel.
4. La colonne "Basic Sampling Weight" apparaît dans le *spreadsheet* comme une formule. Si le *spreadsheet* est changé, il est important de vérifier que la formule pour chaque case correspond à l'équation pour  $W_h$  spécifiée dans ce rapport.
5. La colonne "No. of Valid Sample Hhs." devrait être mise à jour après chaque enquête. Dans ce rapport elle s'appelle  $m'_h$ . Bien que sa valeur est généralement égale à  $m_h$ , elle exclut tout ménage échantillon qui est abandonné.
6. La colonne "No. of Completed Interviews" devrait être mise à jour après chaque enquête. Dans le rapport elle s'appelle  $m''_h$ . Ceci est le nombre de ménages du DR échantillon avec des données complètes (données de production pour tous les mois) qui sont utilisés dans les tabulations.
7. La colonne "Non-Response Adjustment" apparaît dans le *spreadsheet* comme une formule. Elle correspond à  $F_{2H}$  dans le rapport.
8. La colonne "Final Weight" apparaît aussi dans le *spreadsheet* comme une formule. Elle correspond à  $W''_h$  dans le rapport. Ceci est le facteur de pondération qui devrait être affecté à tous les enregistrements pour les ménages échantillons dans le DR.
9. Dans le tableau sur "Sampling Parameters by Stratum" au-dessous des DR échantillons dans le *spreadsheet*, la colonne pour "No. of Hhs. in 1991 Census Frame" ne devrait pas changer. La colonne pour "No. of DR's Selected in Stratum" devrait correspondre à  $n'_h$  dans le rapport, c'est à dire le nombre de DR échantillons effectivement couverts par l'enquête dans la strate. De cette façon le facteur de pondération sera automatiquement ajusté pour les DR échantillons manquant.
10. Pour affecter les pondérations aux enregistrements de l'enquête, le *spreadsheet* peut être utilisé pour obtenir un fichier ASCII qui contient l'identification géographique pour les DR et les pondérations finales correspondantes. Un programme SPSS peut être élaboré pour l'affectation des pondérations aux enregistrements par DR.
11. Les informations concernant les DR dans les 3 autres sous-échantillons sont incluses dans le *spreadsheet*, au-dessous des informations correspondant au sous-échantillon numéro 2 (sélectionné pour la première enquête). Si un sous-échantillon supplémentaire est utilisée pour une enquête, le *spreadsheet* peut être révisé pour les inclure dans le calcul du facteur de pondération. Dans ce cas, toutes les formules peuvent être copiées pour les DR échantillons supplémentaire; pour les facteurs de pondérations de base ( $W_h$ ), les formules devraient être copiées à partir des cases correspondantes pour cette strate dans le sous-échantillon 2. Il est important de vérifier ces formules après la révision du *spreadsheet*. La colonne pour "No. of DR's Selected In Stratum" devrait être mise à jour pour refléter le nombre total de DR dans tous les sous-échantillons incluses dans l'enquête.
12. Le *spreadsheet* avec les pondérations finales pour chaque enquête devraient être sauvegardé sous un nouveau nom qui identifie l'enquête et l'année. Ces fichiers devraient être sauvegardés sur au moins deux disquettes.