

REPUBLIQUE TOGOLAISE
DIRECTION GENERALE
DE LA STATISTIQUE ET DE LA COMPTABILITE NATIONALE

ENQUETE QUIBB 2006 DU TOGO

Plan de sondage et résultat du tirage de l'échantillon des unités primaires
(révision du document technique antérieur de décembre 2003)

Document technique préparé par Julien Amegandjin

Décembre 2005

Sommaire

Avant-propos	2
1. Introduction	3
2. Base de sondage du premier degré	3
3. Base de sondage du deuxième degré	4
4. Domaines d'étude et strates	5
5. Taille de l'échantillon des ménages	6
6. Modes de tirage	7
7. Répartition des échantillons des unités primaires et des ménages	8
8. Mise en œuvre du tirage des unités primaires	11
9. Estimation du total d'une variable et coefficients de pondération correspondants ...	12

Annexe : Liste des 300 zones de dénombrement (ou unités primaires) composant l'échantillon du premier degré

Avant-propos

Le présent plan de sondage remplace celui qui a été élaboré en décembre 2003. Il est apparu nécessaire de réviser ce dernier pour mettre en conformité le mode de tirage du premier degré de l'enquête QUIBB avec celui de l'enquête MICS3 afin que les deux enquêtes utilisent le même échantillon d'unités primaires. L'on évitera ainsi de réaliser deux fois des travaux de cartographie et de redénombrement des ménages.

Le premier plan de sondage de l'enquête QUIBB fait appel pour le premier degré, au tirage avec probabilités inégales et avec remise tandis que le présent plan de sondage est fondé pour le premier degré, sur le tirage systématique avec probabilités proportionnelles aux tailles des unités. Les travaux cartographiques et le redénombrement des ménages qui devront être réalisés préalablement aux tirages du deuxième degré pour chacune des deux enquêtes seront réalisés dans le cadre de l'enquête MICS3. Il est alors apparu indiqué de rapprocher le plan de sondage de l'enquête QUIBB de celui de l'enquête MICS3 pour que les deux enquêtes utilisent le même échantillon d'unités primaires. S'il doit en être ainsi, les résultats du tirage du premier degré de l'enquête MICS3 vont s'appliquer à l'enquête QUIBB. Une deuxième raison qui a milité en faveur de l'adaptation du plan de sondage de l'enquête QUIBB à celui de l'enquête MICS3 tient à l'avantage du tirage systématique avec probabilités proportionnelles aux tailles des unités sur le tirage avec probabilités inégales et avec remise en termes de précision des estimations.

L'élaboration de ce plan de sondage révisé est le produit d'un appui technique réalisé dans le cadre d'une collaboration bénévole du consultant avec la Direction Générale de la Statistique et de la Comptabilité Nationale.

L'enquête QUIBB qui va être menée en 2006 au Togo est une enquête par sondage aléatoire à couverture nationale basée sur un échantillon de 7500 ménages. Son objectif est l'étude des indicateurs de bien-être des ménages et principalement de ceux relatifs aux niveaux de vie, aux revenus et à l'accès aux infrastructures sociales et économiques.

L'exécution de l'enquête est assurée par la Direction Générale de la Statistique et de la Comptabilité Nationale en coopération avec les représentations au Togo, de la Banque mondiale, du PNUD, du FNUAP et de l'UNICEF.

Le présent document technique rend compte des travaux réalisés au titre de l'élaboration du plan de sondage révisé de l'enquête et des résultats du tirage de l'échantillon des unités primaires. Ce document technique n'aborde pas les questions d'estimation des erreurs de sondage. Pour cet aspect du plan de sondage, l'on se réfèrera à la deuxième partie du plan de sondage de l'enquête MICS3 qui sera élaborée au début de l'année 2006.

1. Introduction

L'enquête QUIBB est réalisée au moyen d'un sondage probabiliste, ce qui rend possibles l'extrapolation des résultats de l'échantillon observé à l'ensemble de l'univers sondé ainsi que le calcul des erreurs de sondage en terme de coefficient de variation ou d'intervalle de confiance.

Il est fait appel à un sondage aréolaire à deux degrés avec stratification au premier degré. Les unités statistiques du premier degré ou unités primaires sont les zones de dénombrement (ZD) définies lors des travaux cartographiques censitaires réalisés en 1996 pour le milieu rural et en 1997 pour le milieu urbain. Elles forment la base de sondage du premier degré du sondage.

Un échantillon de zones de dénombrement est tiré au premier degré. Les unités statistiques du deuxième degré ou unités secondaires sont constituées par les ménages des zones de dénombrement tirées au premier degré. Elles définissent la base de sondage du deuxième degré du sondage.

2. Base de sondage du premier degré

Le Togo est subdivisé en 5 régions administratives et chaque région est subdivisée en préfectures se composant de villes pour le milieu urbain et de cantons pour le milieu rural. Des travaux cartographiques censitaires réalisés en 1996 et en 1997 préalablement au recensement national de l'agriculture de 1996 et au projet de recensement général de la population et de l'habitat de 1997 ont fourni une subdivision du territoire national en zones de dénombrement. Les zones de dénombrement sont des unités aréolaires distinctes, non vides et qui recouvrent l'ensemble du pays. Elles forment une partition du territoire national habité et sont constituées de manière à contenir chacune, environ un millier d'habitants. Les travaux cartographiques ont également permis des estimations de l'effectif de la population pour chaque zone de dénombrement.

Au total, 4 618 zones de dénombrement ont été constituées et la population totale estimée à 4 247 832 habitants. La liste des zones de dénombrement avec leurs caractéristiques constitue la base de sondage du premier degré de l'enquête. Les principales caractéristiques de la zone de dénombrement contenues dans la base de sondage sont :

- le code de la région ;
- le nom de la région ;
- le code de la préfecture (ou de l'arrondissement pour Lomé) ;
- le nom de la préfecture (ou de l'arrondissement pour Lomé) ;
- le code du canton (de la ville ou du quartier pour Lomé) ;
- le nom du canton (de la ville ou du quartier pour Lomé) ;
- le code de la zone de dénombrement ;
- le nom du village (ou du quartier urbain) ;
- le milieu de résidence ;
- l'effectif de la population.

La Direction de la Démographie et des Statistiques Sociales de la Direction Générale de la Statistique et de la Comptabilité Nationale possède un fichier informatique en Excel de cette base de sondage. Lors de l'élaboration en 2003 du premier plan de sondage de l'enquête QUIBB, la base de sondage a été légèrement restructurée et on en possède depuis un fichier en Access. Le tableau 1 donne les principales caractéristiques de cette base de sondage.

Les nouveaux fichiers Excel et Access comportent l'essentiel de ce qui est exigé d'une base de sondage pour l'élaboration d'un plan de sondage. Cependant, ils ne comportent pas l'effectif des ménages par zone de dénombrement, ce qui constitue un inconvénient toutefois surmontable, car l'effectif des ménages de la zone de dénombrement est une variable plus fiable que l'effectif de la population de la zone de dénombrement lorsque ces deux variables proviennent de travaux cartographiques.

Tableau 1 : Structure de la base de sondage des zones de dénombrement (ZD)

Code de la région	Nom de la région	Effectif des préfectures ou arrondissements de Lomé	Effectif des cantons ou quartiers de Lomé	Effectif des ZD	Proportion des ZD (%)	Effectif de la population	Proportion de la population (%)
1	Région Maritime	11	128	1 852	40,1	1 750 471	41,2
2	Région des Plateaux	9	95	1 064	23,0	965 640	22,7
3	Région Centrale	4	36	489	10,6	435 702	10,3
4	Région de la Kara	7	61	607	13,1	564 683	13,3
5	Région des Savanes	4	56	606	13,1	531 336	12,5
Ensemble du pays		35	376	4 618	100,0	4 247 832	100,0

3. Base de sondage du deuxième degré

Les ménages des zones de dénombrement tirées au premier degré constituent la base de sondage du deuxième degré de ce plan de sondage. Si les travaux cartographiques censitaires étaient récents, la base de sondage du deuxième degré de l'enquête QUIBB pourrait être extraite de la base globale des ménages dès la connaissance de l'identité des ZD tirées au premier degré. Mais en 2005, l'identité et la localisation des ménages d'une ZD ainsi que leur effectif ont dû connaître bien des changements. C'est pourquoi une observation nouvelle des ménages s'avère nécessaire dans chacune des ZD tirées au premier degré. Cette nouvelle collecte des données sera axée sur la localisation, l'identité et la taille de chaque ménage de la

ZD enquêtée. Cette actualisation de la base des ménages pour les seules ZD de l'échantillon du premier degré est vue comme une véritable pré-enquête distincte avec des équipes propres de superviseurs, de contrôleurs et d'enquêteurs.

4. Domaines d'étude et strates

Un domaine d'étude est une partie de l'univers sondé pour laquelle sont recherchés des résultats significatifs, c'est-à-dire des estimations séparées et d'une précision suffisante. Dans le cas de l'enquête QUIBB 2006, 5 principaux domaines d'étude sont envisagés au départ, à savoir les 5 régions administratives.

Les 5 domaines d'étude qui se confondent avec les 5 régions administratives correspondent grosso modo aux zones d'action des institutions ou ONG qui interviennent dans le domaine social au Togo. Cependant, la ville de Lomé et sa périphérie sont généralement considérées comme constituant une zone d'intervention particulière. C'est pourquoi cet espace urbain a été traité comme un domaine d'étude à part, non pas seulement pour tenir compte de sa spécificité de capitale, mais aussi en raison de son poids, soit 20,1 % de la population nationale ou 48,6 % de la population de la Région Maritime. Finalement, 6 domaines d'étude sont retenus, la Région Maritime étant subdivisée en deux domaines d'étude : la ville de Lomé et sa périphérie, d'une part, le reste du milieu urbain et le milieu rural de la région, d'autre part.

Le critère de stratification retenu pour cette enquête est le milieu de résidence, ce qui permet de séparer les ménages ruraux des ménages urbains dans chaque région, les enquêtes socio-économiques auprès des ménages ayant montré que bien souvent les ménages d'un même milieu de résidence se révèlent plus homogènes.

La prise en compte de ce critère de stratification a conduit à la définition de 12 strates à raison de deux par domaine d'étude. Le tableau 2 présente les caractéristiques de définition des 6 domaines d'étude et des 12 strates.

Tableau 2 : Caractéristiques de définition des domaines d'étude et des strates

Domaine d'étude		Code de la région	Code de la strate	Dénomination de la strate
Code	Dénomination			
1	Lomé et Périphérie	1	11	Lomé
		1	12	Périphérie de Lomé
2	Reste de la Région Maritime	1	13	Autre milieu urbain de la Région Maritime
		1	14	Milieu rural
3	Région des Plateaux	2	21	Milieu urbain
		2	22	Milieu rural
4	Région Centrale	3	31	Milieu urbain
		3	32	Milieu rural
5	Région de la Kara	4	41	Milieu urbain
		4	42	Milieu rural
6	Région des Savanes	5	51	Milieu urbain
		5	52	Milieu rural

La périphérie de la ville de Lomé est constituée par la partie urbaine de la préfecture du Golfe. Bien qu'elle ne fasse pas partie administrativement de la ville de Lomé, elle est pratiquement vue comme telle et ses habitants se considèrent comme des habitants de la capitale. Elle constitue un espace qui compte un effectif de 216 034 habitants, soit 5,1 % de la population nationale. Les conditions de vie de cette périphérie sont présumées différentes de celles de la population de la ville de Lomé ainsi que de celles du reste du milieu urbain de la Région Maritime. En particulier, l'accès à l'eau potable et à l'électricité ainsi que l'éloignement des individus du centre de la capitale font de cet espace, un groupe particulier de population. C'est pourquoi il a été retenu de l'étudier à part en en faisant une strate.

5. Taille de l'échantillon des ménages

Quelle taille n retenir pour l'échantillon national des ménages pour cette première enquête QUIBB ? La taille de l'échantillon de l'enquête sera définie en adoptant la méthode de détermination de la taille de l'échantillon utilisée pour l'enquête MICS3 ainsi que le résultat obtenu, les deux enquêtes possédant de nombreuses similitudes du fait qu'elles portent sur l'observation d'un grand nombre d'indicateurs de condition de vie.

Pour une précision d'estimation donnée et pour une structure de sondage donnée, il correspond à chaque indicateur une taille d'échantillon. C'est la taille minimum qui assure la précision retenue. Chaque indicateur possède donc sa taille d'échantillon. Mais comme l'enquête n'utilise qu'un seul échantillon, on convient de choisir parmi les indicateurs à observer, celui qui requiert la taille d'échantillon la plus grande.

Dans les documents techniques de l'UNICEF présentés dans le premier atelier sur le MICS3 organisé à Dakar, du 13 au 18 juin 2005, il est recommandé de considérer parmi les sous-populations observées dans l'enquête, celle qui représente la plus faible proportion d'individus. Il s'agit de la sous-population des enfants de 12 mois à 23 mois qui représente selon les pays, 3 % à 5 % de la population totale. La raison est que la taille d'échantillon qui satisfait les exigences de précision pour l'estimation d'un indicateur-clé parmi les indicateurs recherchés pour cette population-cible, sera la plus grande taille possible. Et cette taille d'échantillon satisfera ipso facto les exigences de précision d'estimation pour tout indicateur de toute sous-population représentant une proportion plus élevée dans la population sondée. Parmi les indicateurs recherchés sur la population des enfants de 12 mois à 23 mois (taux de couverture vaccinale pour le DTCoq, la rougeole, le BCG ou la poliomyélite), l'indicateur-clé sera de préférence le taux de couverture vaccinale le plus faible attendu.

La taille n de l'échantillon national des ménages est déterminée par la formule ci-après :

$$n = \frac{4r(1-r)fw}{e^2 r^2 pt}$$

Les variables de la fraction du 2^e membre de l'équation sont définis ou justifiées comme suit :

- 4 est le facteur qui traduit un niveau de confiance de 0,95 pour les intervalles de confiance acceptés (en réalité la vraie valeur de ce facteur est le carré du nombre 1,96 que l'on arrondit à 4) ;
- r est le taux de couverture vaccinale pour l'indicateur-clé retenu ;
- f est l'effet du plan de sondage ou l'effet de la structure de sondage : il s'agit d'un coefficient de redressement pour tenir compte de l'utilisation d'un plan de sondage complexe par rapport au sondage aléatoire simple ;
- w est le facteur de correction qui tient compte des cas de non réponse potentiels de l'enquête ;
- e est l'erreur relative de niveau de confiance 0,95 que l'on est disposé à tolérer ;

- p est la proportion de la sous-population cible des enfants des 12 mois à 23 mois ;
- t est la taille moyenne des ménages au niveau national.

Données retenues pour la détermination de la taille n

Pour les principales vaccinations touchant la sous-population des enfants des 12 mois à 23 mois, les taux de couverture observés lors de l'enquête MICS2 en 2000 et les taux de couverture attendus par le Programme Elargi de Vaccination du Togo pour 2005 se présentent comme suit.

	Taux observés en 2000	Taux attendus en 2005
DTC0q3	56,6 %	80 %
Rougeole	56,7 %	80 %
Polio3	57,2 %	80 %
BCG	81,4 %	80 %
Toutes vaccinations	40,8 %	-

Le niveau très élevé des taux de couverture vaccinale pour les principales vaccinations met quelque peu en défaut la méthode de détermination de la taille n . Il est entendu que la méthode vaut plutôt pour des niveaux de couverture attendus plus petits. Avec les résultats de l'enquête MICS2 de 2000, on peut faire l'hypothèse que l'ensemble des vaccinations aura un taux de couverture de 50 % en 2005 après avoir atteint le niveau de 40,8 % en 2000. Le taux de couverture de l'ensemble des vaccinations est l'indicateur-clé retenu avec $r = 0,50$.

Les valeurs suivantes sont retenues pour les différentes variables :

- $r = 0,50$
- $f = 3$ (une étude récente sur un pays de la région indique que cette variable a une valeur de l'ordre de 3)
- $w = 1,06$ (on anticipe 6 % de cas de non réponse)
- $e = 0,12$ (on est disposé à tolérer une erreur relative de 12 % à un niveau de confiance de 0,95)
- $p = 0,035$
- $t = 5,4$

On obtient pour la taille n de l'échantillon national des ménages la valeur 4674. Si le plan de sondage de l'enquête ne prévoyait pas de domaine d'étude, un échantillon de 4674 ménages au niveau national est suffisant pour satisfaire les exigences de précision de l'enquête. Cette taille d'échantillon demeurera toujours suffisante si le plan de sondage ne prévoit pas plus de 4 domaines d'étude, étant entendu qu'un plan de sondage doit allouer à chaque domaine d'étude un sous-échantillon de 800 à 1200 ménages pour être un plan efficace. La valeur de la taille n doit donc être corrigée pour être portée à 6000 au moins si l'on tient compte des 6 domaines d'étude retenus par le plan de sondage de l'enquête. Il a été convenu avec les acteurs de l'organisation de l'enquête MICS3 de considérer un échantillon de 6600 ménages.

Cette taille d'échantillon pour l'enquête MICS3 est proche de celle retenue pour l'enquête QUIBB dans le premier plan de sondage élaboré en 2003, soit 7500. Les exigences de précision pour une première enquête QUIBB réalisée au Togo sont telles qu'elles ont conduit à cette valeur de taille d'échantillon.

6. Modes de tirage

Les tirages des échantillons sont mis en œuvre de façon indépendante d'une strate à l'autre. Les unités primaires (ou zones de dénombrement) sont tirées selon le mode de tirage systématique avec probabilités proportionnelles aux tailles des unités, un cas particulier de tirage avec probabilités inégales et sans remise. La probabilité de sortie d'une ZD est choisie proportionnelle à la taille de la ZD, taille définie par l'effectif de la population de la ZD. La taille de l'échantillon des ZD est choisie égale à 300 au niveau national.

Pour les tirages du deuxième degré, c'est-à-dire les tirages des unités secondaires (ou ménages), il est fait appel à un sondage aléatoire simple sans remise. Les ménages sont donc tirés avec probabilités égales et sans remise. Pour des questions de commodité dans les estimations, on n'utilisera pas ici le tirage systématique avec probabilités égales à la place du sondage aléatoire simple sans remise.

Un nombre constant de ménages est prélevé dans chaque ZD-unité primaire de l'échantillon du premier degré. Ce nombre est choisi égal à 25, ce qui conduit à l'échantillon global de 7500 ménages envisagé.

Le plan de sondage défini par les modes de tirage retenus pour les deux degrés induit un échantillon auto-pondéré, ce qui signifie que toutes les unités secondaires (ou ménages) ont la même probabilité d'appartenir à l'échantillon. L'avantage d'un tel échantillon est la simplification qu'il apporte dans les estimations : une moyenne d'échantillon d'une strate estime sans biais la moyenne correspondante de la strate. Mais malheureusement, cet avantage d'un échantillon auto-pondéré ne va pas jouer ici à cause de l'utilisation simultanée dans les estimations, de valeurs anciennes et de valeurs actualisées de la taille des ZD. Mais, cela ne constitue aucunement une difficulté pour la mise en œuvre des estimations, en raison de la puissance des outils de calcul dont nous disposons aujourd'hui.

7. Répartition des échantillons des unités primaires et des ménages

7.1 Répartition de l'échantillon des unités primaires entre les domaines d'étude

Le taux de sondage du premier degré n'est pas uniforme vis-à-vis des 6 domaines d'étude retenus. La répartition de l'échantillon des 300 ZD selon le domaine d'étude n'est donc pas proportionnelle à la taille des domaines d'étude. La nécessité de sur-échantillonner certains des domaines d'étude en est la raison.

Certains domaines d'étude sont plus faiblement représentés que d'autres dans l'univers sondé. Il en résulte qu'une répartition proportionnelle de l'échantillon des ZD entre les domaines d'étude ne garantit pas des estimations d'une précision acceptable pour certains des domaines d'étude faiblement représentés. Ainsi, la Région Centrale, la Région de la Kara et la Région des Savanes sont sur-échantillonnées. D'un autre côté, le domaine d'étude constitué de la ville de Lomé et de sa périphérie est également sur-échantillonné parce qu'il représente un grand espace urbain dont les ménages sont présumés peu homogènes, avec pour conséquence une plus forte dispersion pour les variables d'observation. D'avoir sur-échantillonné ce domaine d'étude permet d'atténuer la dispersion des variables d'observation en réduisant les variances estimées y relatives. Le tableau 4 présente la répartition retenue pour l'échantillon des 300 unités primaires entre les 6 domaines d'étude.

7.2 Répartition de l'échantillon des unités primaires entre les strates d'un domaine d'étude

Pour chaque domaine d'étude, il est fait appel à un sondage stratifié représentatif au premier degré. Autrement dit, le sous-échantillon des unités primaires de chacun des domaines d'étude est réparti entre les strates proportionnellement à la taille. Il n'a pas été possible de faire appel à l'une des répartitions optimales des sous-échantillons des unités primaires dans les domaines d'étude faute de disposer de données sur la dispersion des variables au niveau des strates. Le tableau 4 présente la répartition des sous-échantillons d'unités primaires au sein des domaines d'étude ainsi que le taux de sondage du 1^{er} degré. On observe bien que le taux de sondage du 1^{er} degré est bien uniforme au sein de chaque domaine d'étude. Au niveau national, le taux de sondage du 1^{er} degré est de 0,06, ce qui correspond à l'observation en moyenne, de 6 unités primaires sur 100. Pour les domaines d'étude sur-échantillonnés, le taux est égal à 0,07 ou 0,08. Il est égal à 0,05 pour les 2 domaines d'étude sous-échantillonnés.

Tableau 3 : Structure de la base de sondage et de l'échantillon du 1^{er} degré

Dénomination du domaine d'étude	Dénomination de la strate	Effectif des ZD	Proportion des ZD %	Effectif de la population %	Proportion de la population	Répartition des 300 ZD par domaine d'étude			Répartition des ZD d'un domaine d'étude (répartition proportionnelle à la taille en ZD des strates)	Taux de sondage du 1 ^{er} degré
						Répartition proportionnelle à la taille en nombre de ZD	Répartition corrigée			
							Effectif	Effectif		
Lomé et Périphérie	Lomé	644	13,9	635 248	15,0				48	0,074
	Périphérie de Lomé	223	4,8	216 034	5,1				16	0,074
Sous-ensemble de Lomé		867	18,8	851 282	20,0	56	64	21,3	64	0,074
Région Maritime	Autre milieu urbain	93	2,0	89 589	2,1				5	0,054
	Milieu rural	892	19,3	809 600	19,1				48	0,054
Sous-ensemble de la Région Maritime		985	21,3	899 189	21,2	64	53	17,7	53	0,054
Région des Plateaux	Milieu urbain	171	3,7	166 929	3,9				9	0,052
	Milieu rural	893	19,3	798 711	18,8				46	0,052
Sous-ensemble de la Région des Plateaux		1 064	23,0	965 640	22,7	69	55	18,3	55	0,052
Région Centrale	Milieu urbain	128	2,8	115 583	2,7				10	0,082
	Milieu rural	361	7,8	320 119	7,5				30	0,082
Sous-ensemble de la Région Centrale		489	10,6	435 702	10,3	32	40	13,3	40	0,082
Région de la Kara	Milieu urbain	145	3,1	148 796	3,5				11	0,074
	Milieu rural	462	10,0	415 887	9,8				34	0,074
Sous-ensemble de la Région de la Kara		607	13,1	564 683	13,3	39	45	15,0	45	0,074
Région des Savanes	Milieu urbain	66	1,4	63 333	1,5				5	0,071
	Milieu rural	540	11,7	468 003	11,0				38	0,071
Sous-ensemble de la Région des Savanes		606	13,1	531 336	12,5	39	43	14,3	43	0,071
Ensemble du pays		4 618	100,0	4 247 832	100,0	300	300	100,0	300	0,065

7.3 Répartition de l'échantillon des ménages entre les strates d'un domaine d'étude

Au deuxième degré, 25 ménages sont prélevés dans chacune des 300 unités primaires tirées au 1^{er} degré, ce qui est conforme au choix de disposer d'un échantillon de 7500 ménages au niveau national. Le tableau 4 présente la répartition des 7500 ménages entre les strates.

Le taux global de sondage, celui des deux degrés réunis, n'est pas connu sauf de façon approximative, parce que la base de sondage ne contient pas d'information sur l'effectif des ménages. La colonne 7 du tableau 4 fournit pour chaque strate, un taux global de sondage estimé. Ce taux est basé sur l'effectif estimé des ménages calculé à partir des tailles moyennes des ménages par région qui sont :

- 4,8 pour la ville de Lomé ;
- 4,5 pour la Région Maritime ;
- 5,3 pour la Région des Plateaux ;
- 6,7 pour la Région Centrale ;
- 6,2 pour la Région de la Kara ;
- 7,8 pour la Région des Savanes.

Le taux global de sondage au niveau national est de 0,0095, ce qui correspond à l'observation de 9,5 ménages en moyenne sur 1000.

Tableau 4 : Structure de la base de sondage et de l'échantillon du 2^e degré

Dénomination du domaine d'étude	Dénomination de la strate	Effectif de la population	Effectif estimé des ménages	Répartition des 300 ZD (Effectif)	Répartition des 7500 ménages (%)	Taux de sondage global
Lomé et Périphérie	Lomé	635 248	132 343	48	1 200	0,0091
	Périphérie de Lomé	216 034	48 008	16	400	0,0083
Sous-ensemble de Lomé		851 282	180 351	64	1 600	0,0089
Région Maritime	Autre milieu urbain	89 589	19 909	5	125	0,0063
	Milieu rural	809 600	179 911	48	1 200	0,0067
Sous-ensemble de la Région Maritime		899 189	199 820	53	1 325	0,0066
Région des Plateaux	Milieu urbain	166 929	31 496	9	225	0,0071
	Milieu rural	798 711	150 700	46	1 150	0,0076
Sous-ensemble de la Région des Plateaux		965 640	182 196	55	1 375	0,0075
Région Centrale	Milieu urbain	115 583	17 251	10	250	0,0145
	Milieu rural	320 119	47 779	30	750	0,0157
Sous-ensemble de la Région Centrale		435 702	65 030	40	1 000	0,0154
Région de la Kara	Milieu urbain	148 796	23 999	11	275	0,0115
	Milieu rural	415 887	67 079	34	850	0,0127
Sous-ensemble de la Région de la Kara		564 683	91 078	45	1 125	0,0124
Région des Savanes	Milieu urbain	63 333	8 120	5	125	0,0154
	Milieu rural	468 003	60 000	38	950	0,0158
Sous-ensemble de la Région des Savanes		531 336	68 120	43	1 075	0,0158
Ensemble du pays		4 247 832	786 595	300	7 500	0,0095

8. Mise en œuvre du tirage des unités primaires

Les tirages des unités primaires sont réalisés de façon indépendante d'une strate à l'autre. Comme cela est indiqué plus haut, les unités primaires (ou zones de dénombrement) sont tirées selon le mode de tirage systématique avec des probabilités proportionnelles aux tailles des unités. Nous utilisons ici la méthode fondée sur les cumuls des probabilités d'inclusion d'ordre 1 des unités exposée brièvement ci-après.

Soit U une population composée des M unités statistiques d'étiquettes $1, 2, \dots, r, \dots, M$ et de tailles respectives $X_1, X_2, \dots, X_r, \dots, X_M$. On désire sélectionner m unités de la population U selon le mode de tirage systématique avec probabilités proportionnelles aux tailles (tirage SPPT) des unités.

On désigne par π_r la probabilité d'inclusion d'ordre 1 de l'unité r , c'est-à-dire la probabilité pour l'unité r de faire partie de l'échantillon des m unités et par X la somme cumulée totale des tailles X_r , soit

$$X = \sum_{j=1}^M X_j$$

La taille relative p_r de l'unité r est définie par la relation

$$p_r = X_r / X \quad \forall r = 1, 2, \dots, M \quad (1)$$

Pour un tirage SPPT, la probabilité d'inclusion π_r est proportionnelle à la taille relative p_r , ce qui s'exprime par la double relation

$$\pi_r = k p_r = k X_r / X \quad \forall r = 1, 2, \dots, M$$

où k est une constante.

On montre que pour tout plan de sondage de taille fixe m (taille de l'échantillon), on a

$$\sum_{r=1}^M \pi_r = m \quad (2)$$

On en déduit

$$m = \sum_{r=1}^M \pi_r = k \sum_{r=1}^M \frac{X_r}{X} = k$$

D'où

$$\pi_r = m p_r = m X_r / X \quad \forall r = 1, 2, \dots, M \quad (3)$$

La procédure de tirage fondée sur les probabilités d'inclusion exige que soit vérifiée la condition :

$$0 < \pi_r < 1 \quad \forall r = 1, 2, \dots, M \quad (4)$$

Si pour une unité r , $\pi_r > 1$, ou bien cette unité est désignée d'office pour faire partie de l'échantillon avec la probabilité 1 et on aura alors à tirer $m-1$ autres unités parmi les $M-1$ unités restantes, ou bien l'unité est segmentée en 2 nouvelles unités ou plus avant les tirages.

La procédure de tirage est fondée sur les cumuls des probabilités d'inclusion des unités. Pour plus de détails concernant la méthode de tirage, l'on se référera au Manuel de l'utilisateur du logiciel ⁽¹⁾ TIRAGE 1.0.

⁽¹⁾ Le logiciel TIRAGE 1.0 est une réalisation de Julien Amegandjin. La version définitive de ce logiciel a été achevée en octobre 2005.

La liste des 300 unités primaires tirées figure en annexe. Chacune des unités primaires de cette liste est repérée par un numéro d'ordre allant de 1 à 300 et par le champ IDENTITE qui est une concaténation des codes de la région, de la préfecture (ou de l'arrondissement pour Lomé), du canton (ou de la ville, ou du quartier pour Lomé) et de la ZD. La ZD peut être également repérée par un code numérique de 4 chiffres qui est une concaténation du code de la strate et du rang de tirage dans la strate comme le montre le tableau 5 pour les ZD tirées de numéros d'ordre 8 à 12.

Tableau 5 : Autre code de repérage de la ZD pour les ZD de n° 8 à 12

N° d'ordre de la ZD	Code de la strate	Rang du tirage	Autre code de repérage de la ZD	IDENTITE
8	11	8	1108	1238261
9	11	9	1109	1238274
10	11	10	1110	1239288
11	11	11	1111	1239301
12	11	12	1112	1239315

Chaque ZD de la liste est présentée avec ses caractéristiques de la base de sondage ainsi que les caractéristiques du tirage.

9. Estimation du total d'une variable d'étude et coefficient de pondération correspondant

Soit y une variable d'étude du sondage dont on veut estimer le total Y . Rappelons les principales données concernant cette estimation.

- $U = \{1, 2, \dots, M\}$ désigne l'univers sondé ;
- Y est le total de la variable y étendue à l'ensemble de l'univers U , avec

$$Y = \sum_{k \in U} y_k$$

- L'échantillon du 1^{er} degré comporte m unités primaires (ou ZD); i désignant l'identité d'une unité de cet échantillon, on a

$$i = 1, 2, \dots, m$$
- Au 2^e degré, un échantillon de n unités secondaires ou ménages (n étant constant) est sélectionné dans chacune des m UP de l'échantillon du 1^{er} degré. Pour l'UP échantillonnée i , n ménages sont sélectionnés parmi les N_i ménages de l'UP i .

Nous allons utiliser la méthode de Des Raj pour construire l'estimateur \hat{Y} du total Y . Pour ce faire, désignons par \hat{Y}_1 l'estimateur du total de la variable y dans un sondage à un degré avec sélection de m unités par tirage systématique avec probabilités proportionnelles aux tailles des unités. Rappelons que \hat{Y}_1 a pour expression

$$\hat{Y}_1 = \sum_{i=1}^m \frac{y_i}{\pi_i}$$

Selon la méthode de Des Raj, l'estimateur \hat{Y} du total Y de la variable y dans le sondage à deux degrés a pour expression

$$\hat{Y} = \sum_{i=1}^m \frac{\hat{y}_i}{\pi_i} \quad (5)$$

où \hat{y}_i est l'estimateur du total de y dans l'UP i , soit

$$\hat{y}_i = \frac{N_i}{n} \sum_{j=1}^n y_{ij} \quad (6)$$

n et N_i étant définis plus haut.

Si la valeur de l'effectif N_i pour l'UP i est la même lors des tirages des deux degrés, alors \hat{Y} a pour expression

$$\hat{Y} = \sum_{i=1}^m \frac{N}{mN_i} \frac{N_i}{n} \sum_{j=1}^n y_{ij}$$

soit
$$\hat{Y} = \frac{N}{mn} \sum_{i=1}^m \sum_{j=1}^n y_{ij} \quad (7)$$

Le coefficient de pondération W_i pour l'estimation du total Y est une constante et a pour expression

$$W_i = \frac{N}{mn} \quad \forall i = 1, 2, \dots, m \quad (8)$$

W_i est comme il fallait s'y attendre, égal à l'inverse de la probabilité de sélection de toute unité secondaire qui est aussi une constante. L'échantillon est donc auto pondéré, ce qui simplifie grandement les estimations.

Mais si la valeur de l'effectif N_i pour l'UP i a été actualisée avant les tirages du 2^e degré et prend la valeur K_i , alors l'estimateur \hat{Y} a pour expression

$$\hat{Y} = \sum_{i=1}^m \frac{N}{mN_i} \frac{K_i}{n} \sum_{j=1}^n y_{ij}$$

soit
$$\hat{Y} = \frac{N}{mn} \sum_{i=1}^m \frac{K_i}{N_i} \sum_{j=1}^n y_{ij} \quad (9)$$

Le coefficient de pondération W_i pour l'estimation du total Y n'est plus une constante. Il dépend de chaque UP échantillonnée et a pour expression

$$W_i = \frac{N}{mn} \frac{K_i}{N_i} \quad \forall i = 1, 2, \dots, m \quad (10)$$

Les formules (9) et (10) définissent l'estimateur du total d'une variable d'étude y . Mais la base de sondage disponible possède une lacune : les effectifs N_i n'y figurent pas. On a dû utiliser les tailles relatives H_i / H au lieu des tailles relatives N_i / N pour tirer les unités primaires, H_i étant l'effectif de la population dans l'UP i et H , l'effectif total de la population de l'univers. La valeur K_i sera observée lors du redénombrement des ménages dans les UP échantillonnées lors des travaux de cartographie. Le coefficient de pondération à utiliser et l'estimateur du total de y sont alors définis respectivement comme suit :

$$W_i = \frac{H}{mn} \frac{K_i}{H_i} \quad \forall i = 1, 2, \dots, m \quad (11)$$

et
$$\hat{Y} = \frac{H}{mn} \sum_{i=1}^m \frac{K_i}{H_i} \sum_{j=1}^n y_{ij} \quad (12)$$

REPUBLIQUE TOGOLAISE
DIRECTION GENERALE
DE LA STATISTIQUE ET DE LA COMPTABILITE NATIONALE

ENQUETES MICS ET QUIBB De 2006

Révision de la structure
d'identification des ZD de la base de sondage pour deux régions

Document préparé par Julien Amegandjin

Janvier 2006

Sommaire

1. Identification ancienne des ZD	2
2. Nouvelle identification des ZD	2

Annexe : ListeRéviséeDesZDTirées_110106 (document présenté séparément
en fichier Excel)

Si une base de sondage de zones de dénombrement (ZD) jouit d'une identification de ces éléments constitutifs au moyen d'un code numérique, ce code doit être défini de manière à éviter des doublons dans la base, c'est-à-dire des ZD possédant un même identifiant. La base de sondage des ZD du Togo actuellement disponible est issue des travaux de cartographie de 1996 et de 1997. Malgré quelques aménagements apportés à cette base en 2003 lors de l'élaboration du premier plan de sondage du QUIBB et en 2005 lors de l'élaboration de l'enquête MICS3, il est apparu nécessaire de réviser une fois de plus l'identification des ZD de cette base. C'est l'objet du présent document technique.

1. Identification ancienne des ZD

L'identification d'une ZD selon la révision de 2003 se fait au moyen d'un code numérique à 6 chiffres pour toutes les régions à l'exception de la ville de Lomé. Ce code numérique est une concaténation des identifiants partiels suivants :

- code de la région (code à 1 chiffre) ;
- code de la préfecture (code à 1 chiffre) ;
- code du canton ou de la ville (code à 2 chiffres) ;
- code de la ZD (code à 2 chiffres).

Pour la ville de Lomé qui, administrativement ne fait partie d'aucune région, les ZD sont identifiées par un code à 6 chiffres à partir des codes partiels suivants :

- code de l'arrondissement (code à 1 chiffre) ;
- code du quartier (code à 2 chiffres) ;
- code de la ZD (code à 3 chiffres).

Pour éviter les risques de doublon pour des ZD de Lomé et des ZD des autres régions, les ZD de Lomé ont été identifiées par un code numérique à 7 chiffres. Pour ce faire, la ville de Lomé a été considérée comme appartenant à la Région Maritime. L'identifiant d'une ZD de Lomé intègre ainsi les codes de la région, de l'arrondissement, du quartier et de la ZD.

2. Nouvelle identification des ZD

Lors de travaux de simulation d'estimation avec des données fondées sur l'identification présentée plus haut, il est apparu que la base de sondage comporte encore des défauts en terme de doublons au niveau de la Région Maritime. Lors de la conception de la structure de la base de sondage au départ, des cantons de parties rurales et des villes de la préfecture du Golfe ont reçu un même code. Ainsi, 36 doublons figurent dans la base de sondage des ZD de la région Maritime comme le montrent les tableaux 1 et 2.

Pour éviter ces doublons, il est suggéré d'identifier les ZD de la Région Maritime en faisant intervenir, en plus des autres identifiants, la variable "type de milieu de résidence". Une ZD de cette région sera donc identifiée par un code à 7 chiffres au moyen des 5 codes partiels suivants :

- code de la région (code à 1 chiffre) ;
- code de la préfecture (code à 1 chiffre) ;
- code du canton ou de la ville (code à 2 chiffres) ;
- code de la ZD (code à 2 chiffres) ;
- code du type de résidence (code à 1 chiffre).

Etant donné que désormais les ZD de la ville de Lomé et de la Région Maritime sont identifiées par un code à 7 chiffres, des risques de doublon dans l'ensemble de la base de sondage sont nombreux. Pour y remédier, on conviendra d'attribuer le nombre 9 au code de région pour la ville de Lomé. Ces changements entraînent une révision dans l'identification des ZD tirées au premier degré pour les enquêtes MICS et QUIBB de 2006. La liste révisée des ZD tirées qui tient compte des nouveaux codes est présentée sous la dénomination ListeRéviséeDesZDTirées_110106. Elle remplace la première liste des ZD établie en octobre 2005.

Avec les dernières modifications, il est désormais possible d'avoir la base de sondage des ZD en un fichier unique sans doublon.

Tableau 1 : ZD urbaines de la préfecture du Golfe observées comme doublons

IDENTITE	CODE_REG	CODE_PREF	PREFECTURE	CODE_CANT	VILLECANTON	CODE_ZD	TYPE_RES
110501	1	1	GOLFE	5	AFLAO	1	1
110502	1	1	GOLFE	5	AFLAO	2	1
110503	1	1	GOLFE	5	AFLAO	3	1
110504	1	1	GOLFE	5	AFLAO	4	1
110505	1	1	GOLFE	5	AFLAO	5	1
110506	1	1	GOLFE	5	AFLAO	6	1
110507	1	1	GOLFE	5	AFLAO	7	1
110508	1	1	GOLFE	5	AFLAO	8	1
110509	1	1	GOLFE	5	AFLAO	9	1
110510	1	1	GOLFE	5	AFLAO	10	1
110511	1	1	GOLFE	5	AFLAO	11	1
110512	1	1	GOLFE	5	AFLAO	12	1
110513	1	1	GOLFE	5	AFLAO	13	1
110514	1	1	GOLFE	5	AFLAO	14	1
111501	1	1	GOLFE	15	TOGBOLE	1	1
111502	1	1	GOLFE	15	TOGBOLE	2	1
111503	1	1	GOLFE	15	TOGBOLE	3	1
111504	1	1	GOLFE	15	TOGBOLE	4	1
111505	1	1	GOLFE	15	TOGBOLE	5	1
111506	1	1	GOLFE	15	TOGBOLE	6	1
112001	1	1	GOLFE	20	AGOE	1	1
112002	1	1	GOLFE	20	AGOE	2	1
112003	1	1	GOLFE	20	AGOE	3	1
112004	1	1	GOLFE	20	AGOE	4	1
112005	1	1	GOLFE	20	AGOE	5	1
112006	1	1	GOLFE	20	AGOE	6	1
112007	1	1	GOLFE	20	AGOE	7	1
112008	1	1	GOLFE	20	AGOE	8	1
112009	1	1	GOLFE	20	AGOE	9	1
112010	1	1	GOLFE	20	AGOE	10	1
112011	1	1	GOLFE	20	AGOE	11	1
112012	1	1	GOLFE	20	AGOE	12	1
112013	1	1	GOLFE	20	AGOE	13	1
112501	1	1	GOLFE	25	AMOUTIVE	1	1
113001	1	1	GOLFE	30	BAGUIDA	1	1
113002	1	1	GOLFE	30	BAGUIDA	2	1
*							

Tableau 2 : ZD rurales de la préfecture du Golfe observées comme doublons

IDENTITE	CODE_REG	CODE_PREF	PREFECTURE	CODE_CANT	VILLECANTON	CODE_ZD	TYPE_RES
110501	1	1	GOLFE	5	AFLAO	1	2
110502	1	1	GOLFE	5	AFLAO	2	2
110503	1	1	GOLFE	5	AFLAO	3	2
110504	1	1	GOLFE	5	AFLAO	4	2
110505	1	1	GOLFE	5	AFLAO	5	2
110506	1	1	GOLFE	5	AFLAO	6	2
110507	1	1	GOLFE	5	AFLAO	7	2
110508	1	1	GOLFE	5	AFLAO	8	2
110509	1	1	GOLFE	5	AFLAO	9	2
110510	1	1	GOLFE	5	AFLAO	10	2
110511	1	1	GOLFE	5	AFLAO	11	2
110512	1	1	GOLFE	5	AFLAO	12	2
110513	1	1	GOLFE	5	AFLAO	13	2
110514	1	1	GOLFE	5	AFLAO	14	2
111501	1	1	GOLFE	15	TOGBOLE	1	2
111502	1	1	GOLFE	15	TOGBOLE	2	2
111503	1	1	GOLFE	15	TOGBOLE	3	2
111504	1	1	GOLFE	15	TOGBOLE	4	2
111505	1	1	GOLFE	15	TOGBOLE	5	2
111506	1	1	GOLFE	15	TOGBOLE	6	2
112001	1	1	GOLFE	20	AGOE	1	2
112002	1	1	GOLFE	20	AGOE	2	2
112003	1	1	GOLFE	20	AGOE	3	2
112004	1	1	GOLFE	20	AGOE	4	2
112005	1	1	GOLFE	20	AGOE	5	2
112006	1	1	GOLFE	20	AGOE	6	2
112007	1	1	GOLFE	20	AGOE	7	2
112008	1	1	GOLFE	20	AGOE	8	2
112009	1	1	GOLFE	20	AGOE	9	2
112010	1	1	GOLFE	20	AGOE	10	2
112011	1	1	GOLFE	20	AGOE	11	2
112012	1	1	GOLFE	20	AGOE	12	2
112013	1	1	GOLFE	20	AGOE	13	2
112501	1	1	GOLFE	25	AMOUTIVE	1	2
113001	1	1	GOLFE	30	BAGUIDA	1	2
113002	1	1	GOLFE	30	BAGUIDA	2	2
*							

Enr : 36 sur 36

REPUBLIQUE TOGOLAISE
DIRECTION GENERALE
DE LA STATISTIQUE ET DE LA COMPTABILITE NATIONALE

ENQUETE QUIBB DE 2006

Mise en œuvre du tirage de l'échantillon des ménages

Document préparé par Julien Amegandjin

Avril 2006

Sommaire

1. Introduction	2
2. Redénombrement des ménages des ZD échantillonnées	2
3. Identification des ZD et des ménages	2
4. Mise en œuvre du tirage des ménages	4
5. Remplacement des ménages absents ou opposant un refus à l'enquête	5

Liste des fichiers Access fournis

- BaseDeSondage_Ménages
- Echantillon_ZD
- Echantillon_Ménages
- Echantillon_MénagesAvecCoordonnées

1. Introduction

Le présent document rend compte de la mise en œuvre du tirage de l'échantillon des ménages. Il fait suite à un premier document datant d'octobre 2005 et intitulé "Plan de sondage et résultat du tirage de l'échantillon des unités primaires".

On rappelle que les tirages des échantillons sont mis en œuvre de façon indépendante d'une strate à l'autre. Les unités primaires (ou zones de dénombrement) sont tirées selon le mode du tirage systématique avec probabilités proportionnelles aux tailles des unités, la taille étant définie par l'effectif de la population de la ZD. Au total, un échantillon de 300 ZD a été sélectionné au niveau national.

Pour les tirages du deuxième degré, c'est-à-dire les tirages des unités secondaires (ou ménages), il est fait appel à un sondage aléatoire simple sans remise. Les ménages sont donc tirés avec probabilités égales et sans remise.

Un nombre constant de ménages est prélevé dans chaque ZD-unité primaire de l'échantillon du premier degré. Ce nombre est choisi égal à 25, ce qui conduira à la sélection d'un échantillon national de 7500 ménages.

2. Redénombrement des ménages des ZD échantillonnées

Les travaux cartographiques censitaires à partir desquels la base de sondage des zones de dénombrement a été constituée datant d'une dizaine d'années, il a été nécessaire de procéder en 2006 à l'actualisation de la liste des ménages des ZD échantillonnées préalablement au tirage du deuxième degré. Cette actualisation de la liste des ménages est axée sur la localisation et l'identification des ménages de chaque ZD observée.

Les listes des ménages redénommés dans les ZD échantillonnées définissent la base de sondage du deuxième degré du sondage à partir de laquelle le tirage des ménages sera effectué. Les ménages redénommés pour les 300 ZD sont présentés dans un fichier dénommé BaseDeSondage_Ménages qui comporte essentiellement des éléments d'identification des ménages dans la ZD, dont :

- le numéro d'ordre de la concession ;
- le numéro d'ordre du ménage ;
- les nom et prénoms du chef de ménage ;
- le quartier ou le nom de la rue.

3. Identification des ZD et des ménages

La méthode d'identification retenue pour une ZD de l'échantillon des ZD fait intervenir :

- le code du domaine d'étude ;
- le code de la strate à l'intérieur du domaine d'étude ;
- le numéro d'ordre de la ZD dans l'échantillon des 300 ZD.

Une concaténation des trois éléments ci-dessus fournit l'identité de la ZD comme un nombre de 5 chiffres.

Pour identifier un ménage dans une ZD donnée, on se référera à l'identité de la ZD et au numéro d'ordre du ménage dans la ZD d'après le redénombrement des ménages. Une concaténation de ces deux éléments fournit l'identité d'un ménage comme un nombre de 8 chiffres.

Les tableaux 1 et 2 ci-après présentent des exemples de construction de l'identité d'une ZD échantillonnée et de l'identité d'un ménage d'une telle ZD.

Dans le fichier BaseDeSondage_Ménages, les identités des ZD et des ménages sont construites de cette manière et sont dénommées respectivement ID_ZD et ID_MEN.

Tableau 1 : Exemples de code de ZD de l'échantillon des ZD

Nom du domaine d'étude	Nom de la strate	Code du domaine d'étude	Code de la strate	Numéro d'ordre de la ZD	Identité de la ZD
Lomé	Lomé	1	1	43	11043
Lomé	Lomé	1	1	44	11044
Lomé	Lomé	1	1	45	11045
Lomé	Lomé	1	1	46	11046
Lomé	Lomé	1	1	47	11047
Lomé	Lomé	1	1	48	11048
Lomé	Lomé Périphérique	1	2	49	12049
Lomé	Lomé Périphérique	1	2	50	12050
Lomé	Lomé Périphérique	1	2	51	12051
Lomé	Lomé Périphérique	1	2	52	12052
Lomé	Lomé Périphérique	1	2	53	12053
Lomé	Lomé Périphérique	1	2	54	12054
Lomé	Lomé Périphérique	1	2	55	12055
Région des Savanes	Milieu rural	5	2	295	52295
Région des Savanes	Milieu rural	5	2	296	52296
Région des Savanes	Milieu rural	5	2	297	52297
Région des Savanes	Milieu rural	5	2	298	52298
Région des Savanes	Milieu rural	5	2	299	52299
Région des Savanes	Milieu rural	5	2	300	52300

Tableau 2 : Exemples de code de ménage

Nom du domaine d'étude	Nom de la strate	Code du domaine d'étude	Code de la strate	Numéro d'ordre de la ZD	Identité de la ZD	Numéro d'ordre du ménage	Identité du ménage
Lomé	Lomé	1	1	44	11044	12	11044012
					11044	39	11044039
					11044	117	11044117
					11044	129	11044129
					11044	246	11044246
Région des Plateaux	Milieu rural	3	2	127	32127	4	32127004
					32127	58	32127058
					32127	156	32127156
					32127	212	32127212
					32127	278	32127278
Région Centrale	Milieu urbain	4	1	180	41180	8	41180008
					41180	15	41180015
					41180	33	41180033
					41180	158	41180158
Région des Savanes	Milieu rural	6	2	291	62291	28	62291028
					62291	66	62291066
					62291	190	62291190
					62291	200	62291200

3. Mise en œuvre du tirage des ménages

Première étape du tirage

La première étape du tirage des ménages va consister à créer un fichier à partir de l'échantillon des 300 ZD tirées. Le nouveau fichier dénommé "Echantillon_ZD" comportera les variables suivantes :

- code du domaine d'étude ;
- code de la strate (ou le numéro d'ordre de la strate dans le domaine d'étude) ;
- le numéro d'ordre de la ZD (variant de 1 à 300) ;
- l'identité de la ZD (qui est une concaténation des 3 variables ci-dessus) ;
- la taille de la ZD (ou l'effectif des ménages de la ZD après l'opération de redénombrement des ménages).

Le fichier "Echantillon_ZD" sera modifié pour que sa structure soit conforme avec celle de la table "SourceTirages2eD_SPPT" de la base de données Data.mdb associée au logiciel TIRAGE 1.0 et utilisée pour le tirage des unités secondaires (ou ménages).

L'identité de la ZD et la taille de la ZD seront nommées "Identité_UP" et "Taille_UP" respectivement afin d'être reconnues par le logiciel de tirage. En outre, il sera créé dans le fichier "Echantillon_ZD", une nouvelle variable (ou champ) dénommée "NombreUS_àTirer". Ce champ prendra la valeur 25 pour chacune des 300 ZD du fichier.

Deuxième étape du tirage

On procède ensuite au tirage de l'échantillon des ménages en se servant du module du logiciel TIRAGE1.0 dénommé "Tirages PESR de l'échantillon du 2^e degré avec un nombre constant d'US par UP" associé au module "Table-source des tirages du 2^e degré à la suite de tirages SPPT au 1^{er} degré" dans le menu "Tirages dans un sondage à 2 degrés".

Le tirage des 25 ménages dans une ZD qui en compte 188, consiste à tirer avec probabilités égales et sans remise, 25 nombres compris entre 1 et 188 inclus. Les 25 nombres tirés sont les numéros d'ordre des ménages tirés dans la ZD.

Les numéros d'ordre des 7500 ménages sélectionnés dans les 300 ZD sont envoyés dans la table "RésultatTirages2eD_SPPT" de la base de données Data.mdb.

Troisième étape du tirage

A partir de la table "RésultatTirages2eD_SPPT", on construit le fichier Echantillon_Ménages en adjoignant aux données existantes, un nouveau champ (ID_MEN) qui fournit l'identité du ménage par concaténation de l'identité de la ZD et du numéro d'ordre du ménage. Chacun des 7500 ménages de ce fichier Echantillon_Ménages comporte comme éléments d'identification seulement :

- le rang du tirage du ménage dans la ZD ;
- le numéro d'ordre de ménage dans la ZD ;
- l'identité de la ZD (ID_ZD) ;
- l'identité du ménage (ID_MEN).

Il ne comporte pas les coordonnées géographiques du ménage permettant à l'enquêteur de le retrouver sur le terrain. L'étape suivante du tirage va donc consister à associer à chaque ménage tiré, ses coordonnées géographiques.

Quatrième étape du tirage

La quatrième étape consiste à mettre en relation le fichier Echantillon_Ménages avec le fichier BaseDeSondage_Ménages de manière à apparier les 7500 ménages tirés avec leurs correspondants dans la base de sondage. Cet appariement permet d'obtenir pour chaque ménage tiré, ses coordonnées géographiques complètes.

Le fichier BaseDeSondage_Ménages comporte 65630 enregistrements et possède un volume admissible pour permettre la mise en relation des fichiers BaseDeSondage_Ménages et Echantillon_Ménages en Access sans que le fichier BaseDeSondage_Ménages soit nécessairement scindé au préalable en sous-fichiers par strate ou par domaine d'étude.

L'opération d'appariement des ménages tirés avec leurs correspondants de la base de sondage fournit un fichier dénommé Echnatillon_MénagesAvecCoordonnées. Ce dernier fichier contient toutes les informations pour retrouver un ménage échantillonné sur le terrain.

Pour des questions de commodité, il sera édité autant d'états de ce fichier pour une ZD qu'il y aura d'enquêteurs affectés à la ZD. En outre, il sera édité un état du fichier BaseDeSondage_Ménages pour chaque ZD pour les besoins des contrôleurs chargés du contrôle de la collecte dans les ZD.

5. Remplacement des ménages absents ou opposant un refus à l'enquête

Quand peut-on remplacer un ménage de la liste des 25 ménages à interviewer dans une ZD ? Cela est possible seulement dans les cas de ménages absents ou de ménages opposant un refus à l'enquête. Mais, seul le contrôleur d'enquête chargé de la ZD peut décider si un ménage doit être remplacé pour cause d'absence ou de refus de répondre à l'interview, et lorsqu'il se sera assuré que tous les recours possibles ont été utilisés en vain pour rencontrer le chef de ménage absent ou pour faire changer d'avis au chef de ménage réticent.

Pour remplacer un ménage absent ou qui refuse de répondre à l'enquête on pourrait tirer un certain de ménages qu'on met en réserve pour être utilisés comme ménages de remplacement. Ainsi, on pourrait tirer 32 ménages au lieu de 25 de chaque ZD et mettre en réserve 7 ménages. Cette procédure est entachée d'erreur parce que le tirage de 32 ménages avec probabilités égales et sans remise dans une ZD conduit à une probabilité de sélection d'un ménage de la ZD différente de celle correspondant au tirage de 25 ménages dans la même ZD.

Il est suggéré la solution de remplacement suivante : si le ménage absent ou qui oppose un refus porte le numéro n , il sera remplacé par le ménage de numéro $n+3$. Si le ménage de numéro $n+3$ est absent ou oppose un refus, on ira interviewer le ménage de numéro $n+6$, et ainsi de suite, on considèrera les ménages de numéros $n+9$, $n+12$, etc. Le nombre 3 a été obtenu de façon aléatoire.