

**Institut National de la Statistique
et des Etudes Economiques et Démographiques
(INSEED)**

**Conception et mise en place d'un échantillon maître
bâti à partir de la base RGPH 2010**

Document technique préparé
par
Julien Amegandjin, consultant

Juin 2015

Sommaire

1. Introduction	2
2. La base RGPH 2010	3
3. Définition de l'échantillon maître	
3.1 Taille de l'échantillon maître	4
3.2 Répartition de l'échantillon maître entre les domaines d'étude	5
3.3 Méthode d'échantillonnage	8
3.4 Evaluation de la base RGPH 2010	8
3.5 Répartition de l'échantillon maître entre les strates	9
3.6 Tirage de l'échantillon maître	10
3.7 Probabilités d'inclusion des ZD de l'échantillon maître	10
4. Activités d'aménagement des ZD de l'échantillon maître	12
5. Mode d'emploi de l'échantillon maître	12

Documents annexes

Annexe 1 : Extrait du tableau des probabilités d'inclusion des ZD de l'EM (2 pages sur 40)

Annexe 2 : Extrait de la liste des ZD de l'échantillon maître (15 champs sur 19)

1. Introduction

Comme beaucoup d'autres pays africains, le Togo doit répondre à un large éventail de besoins d'information dans les domaines des statistiques démographiques, sociales et économiques au profit des décideurs politiques, des opérateurs économiques, des ONG et autres utilisateurs de statistiques. L'INSEED, l'organe central national statistique du Togo, est donc amené à réaliser un grand nombre d'enquêtes auprès des ménages pour compléter les données du RGPH (Recensement Général de la Population et de l'Habitat) qui est mis en œuvre une fois tous les dix ans.

Font partie de la panoplie des enquêtes conduites par l'INSEED :

- les enquêtes sur les conditions de vie des ménages ;
- l'enquête sur l'emploi et le secteur informel ;
- l'enquête démographique et de santé (EDS) ;
- l'enquête sur la pauvreté (QUIBB) ;
- ou l'enquête par grappes à indicateur multiple (MICS), etc.

La réalisation de ces enquêtes requiert une base de sondage exhaustive et à jour pour fournir des échantillons représentatifs des différentes composantes de la population. La base de sondage utilisée est celle fournie par le dernier recensement général de la population et de l'habitat (RGPH). Le recensement étant organisé tous les dix ans, il arrive que les données de la base de sondage deviennent de plus en plus obsolètes au fur et à mesure que l'on s'éloigne de la date de ce recensement. En effet, de nombreuses zones de dénombrement (ZD) qui constituent les unités primaires de la base de sondage voient évoluer leurs caractéristiques, à savoir :

- l'étendue ;
- la liste des concessions et des logements ;
- la liste des ménages ;
- la liste des individus.

L'INSEED et les autres services sectoriels de statistique doivent donc conduire pour chaque enquête, des opérations de mise à jour de la base de sondage qui portent essentiellement sur les unités primaires sélectionnées. De telles opérations exigent des moyens humains et financiers importants de même que beaucoup de temps.

L'un des inconvénients de ce procédé d'actualisation de la base est qu'il faut en répéter le coût pour chaque enquête sans tirer aucun avantage des différents travaux de mise à jour précédents. Pour réduire le coût de la mise à jour de la base de sondage sur une période de temps donnée et aussi pour augmenter l'efficacité de cette mise à jour, l'INSEED envisage de mettre en place un échantillon maître à partir de la base RGPH 2010.

► *Un échantillon maître est un échantillon à partir duquel des sous-échantillons peuvent être sélectionnés pour répondre aux besoins de plusieurs enquêtes ou cycles d'enquêtes (Nations Unies, 1986).*

L'échantillon maître est un échantillon de grande taille, représentatif des régions et des différentes couches sociales du pays, établi à partir d'un recensement, et en vue de servir de base de sondage au programme des enquêtes sociales et économiques d'une période intercensitaire.

Un échantillon maître de conception simple et assez courante est composé d'unités primaires de sondage qui sont des zones de dénombrement. L'échantillon est utilisé pour un processus de sélection à deux degrés où les unités de sondage du deuxième degré sont des ménages ou des unités d'habitation.

L'échantillon maître est mis à jour de façon continue pendant toute la période intercensitaire, au moins une fois l'an, de façon à être utilisable à tout moment. Il s'ensuit que les différentes enquêtes du programme peuvent être organisées rapidement et plus facilement grâce à l'existence d'une base de sondage prête à l'emploi. En outre, le coût de la sélection des échantillons s'en trouve réduit.

En dépit de ces avantages, l'échantillon maître possède quelques limites dont la principale est le chevauchement des échantillons des différentes enquêtes et son possible délaissement pour défaut de programme d'enquêtes conséquent.

2. La base RGPH 2010

La base de sondage dont l'échantillon maître en construction sera une représentation fidèle et d'un emploi plus simple est constituée par les données du recensement général de la population et de l'habitat (RGPH) de 2010. Cette base de données est composée de 6657 zones de dénombrement (ZD), 1 206 396 ménages et 5 843 327 habitants. Le tableau 1 ci-après présente la structure de cette base en termes de distribution des ZD, des ménages et de la population selon la région (ou domaine d'étude).

Tableau 1 : Structure de la base RGPH 2010

Région	Effectif des ZD	Proportion	Effectif des ménages	Proportion	Effectif de la population	Proportion
Maritime	1 258	0,19	231 993	0,19	1 064 950	0,18
Plateaux	1 515	0,23	276 740	0,23	1 309 310	0,22
Centrale	681	0,10	106 084	0,09	589 258	0,10
Kara	830	0,12	141 678	0,12	731 884	0,13
Savanes	816	0,12	113 861	0,09	783 071	0,13
Lomé	1 557	0,23	336 040	0,28	1 364 854	0,23
Ensemble	6 657	1,00	1 206 396	1,00	5 843 327	1,00

On observe que la distribution des ZD et celle de la population telles qu'elles figurent respectivement dans les colonnes 3 et 7 sont proches, ce qui dénote que les consignes ont été bien suivies dans la délimitation des ZD en fonction d'une taille de population exigée. La distribution fournie par les ménages est la plus disparate des trois distributions. La distribution de la base selon la population paraît la moins disparate et c'est elle qui sera le plus souvent retenue pour représenter la taille relative des domaines d'étude.

Le tableau 2 présente la structure de la base RGPH 2010 selon le domaine d'étude et le milieu de résidence. Il donne, en particulier, les tailles relatives des strates urbaine et rurale de chaque domaine d'étude. L'on observe pour l'ensemble du pays que le milieu urbain et le milieu rural se composent respectivement de 37% et 63% de la population, et respectivement de 35% et 65% des zones de dénombrement.

Tableau 2 : Structure de la base RGPH 2010 selon le domaine d'étude et le milieu de résidence

Région	Milieu	ZD		Ménages		Population	
		Effectif	Proportion	Effectif	Proportion	Effectif	Proportion
Maritime	Urbain	127	0,10	31 911	0,14	125 670	0,12
	Rural	1 131	0,90	200 082	0,86	939 280	0,88
	Sous-total	1 258	1,00	231 993	1,00	1 064 950	1,00
Plateaux	Urbain	257	0,17	63 740	0,23	256 620	0,20
	Rural	1 258	0,83	213 000	0,77	1 052 690	0,80
	Sous-total	1 515	1,00	276 740	1,00	1 309 310	1,00
Centrale	Urbain	125	0,18	28 508	0,27	144 258	0,24
	Rural	556	0,82	77 576	0,73	445 000	0,76
	Sous-total	681	1,00	106 084	1,00	589 258	1,00
Kara	Urbain	169	0,20	42 758	0,30	175 512	0,24
	Rural	661	0,80	98 920	0,70	556 372	0,76
	Sous-total	830	1,00	141 678	1,00	731 884	1,00
Savanes	Urbain	109	0,13	21 619	0,19	110 142	0,14
	Rural	707	0,87	92 242	0,81	672 929	0,86
	Sous-total	816	1,00	113 861	1,00	783 071	1,00
Lomé	Lomé Commune	870	0,56	195 406	0,58	761 993	0,56
	Golfe Urbain	687	0,44	140 634	0,42	602 861	0,44
	Sous-total	1 557	1,00	336 040	1,00	1 364 854	1,00
Ensemble	Urbain	2 344	0,35	524 576	0,43	2 177 056	0,37
	Rural	4 313	0,65	681 820	0,57	3 666 271	0,63
	Total	6 657	1,00	1 206 396	1,00	5 843 327	1,00

3. Définition de l'échantillon maître

Pour l'échantillon maître à construire, nous en retenons la forme la plus simple et la plus courante.

► *Il est conçu comme un échantillon d'unités primaires (ou zones de dénombrement) provenant de la base RGPH 2010 pour servir de base de sondage dans un processus de sondage à plusieurs degrés.*

L'échantillon maître devra être représentatif des six grandes entités géographiques (les cinq régions et la commune de Lomé) ainsi que du milieu de résidence.

Il devra posséder une taille et une répartition appropriées afin de pouvoir servir de base de sondage à toutes les enquêtes du programme d'enquêtes en place et lieu de la base RGPH 2010.

3.1 Taille de l'échantillon maître

L'échantillon maître en construction doit permettre la conception et la mise en œuvre des plans de sondage de toutes les enquêtes d'un programme décennal d'enquêtes couvrant la période intercensitaire 2010-2020. Au nombre des enquêtes du programme figure l'Enquête Agricole Annuelle qui possède une spécificité, celle de ne couvrir que le milieu rural. Plus précisément, l'Enquête Agricole Annuelle porte sur les ménages agricoles comme unités du second degré tandis que les ZD rurales en constituent les unités primaires.

L'échantillon maître devant également servir l'Enquête Agricole Annuelle, il doit comporter un sous-échantillon de ZD rurales d'une taille suffisante pour fournir l'échantillon à enquêter chaque année. Actuellement l'Enquête Agricole Annuelle utilise un échantillon de 500 unités primaires ou ZD rurales. Il pourrait être intéressant de retenir que la composante des ZD rurales dans l'échantillon maître ait une taille qui soit le double de la taille de l'échantillon de l'Enquête Agricole Annuelle, à savoir 1000 ZD rurales. Le tableau 3 présente les détails de cette option dans la ligne 5 ainsi que ceux de cinq autres options.

Tableau 3 : Caractéristiques de quelques options de taille d'EM

Taille de l'échantillon maître en nombre de ZD	Proportion représentée dans la base RGPH 2010	Effectif des ZD urbaines dans l'EM	Effectif des ZD rurales dans l'EM
500	0,07	175	325
1 000	0,15	350	650
1 200	0,18	420	780
1 270	0,19	445	826
1 600	0,24	560	1 040
2 000	0,30	700	1 300

L'Enquête Agricole Annuelle semble être la seule enquête du programme avec ce type de contrainte. La taille de l'échantillon maître sera donc définie en tenant compte de cette spécificité de l'Enquête Agricole Annuelle. L'on pourrait donc retenir l'option de la ligne 5 et construire un échantillon maître de 1600 ZD avec environ 1040 ZD rurales et 560 ZD urbaines. Pour cette option, l'échantillon maître représente 24% de la base originelle.

S'il arrive qu'une enquête du programme porte seulement sur des ménages urbains, la contrainte apparue avec l'Enquête Agricole Annuelle serait dans ce cas beaucoup difficile à gérer. Mais, l'option de la ligne 5 du tableau 3 pourrait fournir un échantillon de taille acceptable avec ses 560 ZD urbaines.

3.2 Répartition de l'échantillon maître entre les domaines d'étude

Quelle méthode d'allocation utiliser pour répartir les ZD de l'échantillon maître entre les six domaines d'étude ? Il existe plusieurs méthodes pour réaliser une telle répartition. Ce sont, parmi les plus courantes :

- l'allocation proportionnelle à la taille du domaine d'étude (en termes d'effectif des ménages ou de la population) ;
- l'allocation égalitaire ou allocation à parts égales entre les domaines d'étude ;
- l'allocation proportionnelle à la racine carrée de la taille du domaine d'étude ;
- l'allocation de Kish.

Examinons ces différentes méthodes en considérant les notations suivantes :

- n désigne la taille globale de l'échantillon maître, c'est-à-dire l'effectif des ZD qui le composent ;
- n_d est la taille du sous-échantillon alloué au domaine d , c'est-à-dire l'effectif des ZD définissant ce sous-échantillon ;
- N_d est l'effectif des ménages (ou de la population) du domaine d dans la base RGPH 2010 ;

- N est l'effectif global des ménages (ou de la population) de l'univers sondé dans la base RGPH 2010 ;
- W_d est la taille relative du domaine d , soit $W_d = N_d/N$;
- I désigne l'indice d'allocation de Kish ;
- D désigne le nombre de domaines d'étude de l'univers.

L'allocation de l'échantillon maître selon les différentes méthodes s'exprime par les équations ci-après.

a) Allocation proportionnelle à la taille W_d du domaine d

$$n_d = n \frac{N_d}{N} = n W_d \quad (1)$$

b) Allocation égalitaire

$$n_d = \frac{n}{D} \quad (2)$$

c) Allocation proportionnelle à la racine carrée de la taille N_d

$$n_d = n \frac{\sqrt{N_d}}{\sum_j \sqrt{N_j}} \quad (3)$$

d) Allocation de Kish

$$n_d = n \frac{\sqrt{D^{-2} + I W_d^2}}{\sum_j \sqrt{D^{-2} + I W_j^2}} \quad (4)$$

L'allocation de Kish au domaine d est définie proportionnelle à la grandeur K_d définie par la relation

$$K_d = \sqrt{D^{-2} + I W_d^2} \quad (5)$$

Cette méthode est souvent appliquée avec $I = 1$, valeur pour laquelle la méthode possède l'avantage de fournir des précisions d'estimation très proches pour les estimations au niveau national et les estimations au niveau des domaines d'étude.

L'allocation de Kish se confond à l'allocation égalitaire si l'indice I prend la valeur 0 ; elle se confond à l'allocation proportionnelle à la taille du domaine quand l'indice I tend vers l'infini.

Dans la construction de nombreux échantillons maîtres, il est fait appel à l'allocation proportionnelle aux tailles des domaines d'étude exprimées en effectif des ménages ou en effectif de la population. Pour aussi bonne qu'elle soit, parce qu'elle assure de bonnes estimations au niveau national, cette stratégie n'alloue qu'une faible proportion de l'échantillon aux domaines de petite taille. Il s'ensuit que les estimations séparées voulues pour de tels domaines de petite taille possèdent des précisions limitées.

L'allocation à égalité de l'échantillon maître entre les domaines d'étude permet de meilleures estimations pour les différents domaines avec toutefois de grands écarts entre les taux de sondage quand les domaines d'étude sont de tailles très différentes.

L'allocation proportionnelle à la racine carrée de la taille du domaine d'étude et l'allocation de Kish sont toutes les deux des compromis entre l'allocation proportionnelle à la taille du domaine et l'allocation égalitaire. Toutes les deux ont pour but de relever les parts des petits domaines et de réduire celles des grands domaines dans la répartition de l'échantillon maître.

Tableau 4 : Données de la répartition d'un échantillon maître de 1600 ZD selon les différentes méthodes d'allocation

Domaine	N _d	$\sqrt{N_d}$	W _d	K _d	Allocation égalitaire	Allocation proportionnelle à N _d	Allocation proportionnelle à la raine carrée de N _d	Allocation proportionnelle à K _d (allocation de Kish)
					n _d	n _d	n _d	n _d
Maritime	1 064 950	1 032,0	0,18	0,25	266,7	291,6	282,1	276,3
Plateaux	1 309 310	1 144,3	0,22	0,28	266,7	358,5	312,8	312,4
Centrale	589 258	767,6	0,10	0,19	266,7	161,3	209,9	217,9
Kara	731 884	855,5	0,13	0,21	266,7	200,4	233,9	233,2
Savanes	783 071	884,9	0,13	0,21	266,7	214,4	241,9	239,2
Lomé	1 364 854	1 168,3	0,23	0,29	266,7	373,7	319,4	321,0
Ensemble	5 843 327	5 852,5	1,00	1,43	1 600	1 600	1 600	1 600

Le tableau 4 ci-dessus présente les données de la répartition d'un échantillon maître de 1600 ZD pour les quatre méthodes d'allocation, la taille N_d du domaine d'étant l'effectif de la population. L'on observe que les répartitions des deux dernières colonnes sont moins disparates que celles fournies par l'allocation proportionnelle à la taille du domaine.

► *En définitive, nous retenons la répartition fournie par la méthode d'allocation de Kish figurant dans la dernière colonne du tableau 4.*

3.3 Méthode d'échantillonnage

Le plan d'échantillonnage conduisant à l'élaboration de l'échantillon maître repose sur les paramètres suivants :

- la base RGPH 2010 est la base de sondage de laquelle sera prélevé l'échantillon maître ;
- la base de sondage compte six domaines d'étude : les cinq régions et la commune de Lomé ;
- un traitement spécial est apporté à la Région Maritime et à la commune de Lomé pour rendre compte d'une réalité du terrain que les mesures administratives tardent à prendre : la partie urbaine de la préfecture du Golfe dans la Région Maritime est en réalité, la périphérie de Lomé et est considérée ici comme une strate urbaine de Lomé en plus de Lomé-Commune qui est aussi une strate urbaine du domaine d'étude défini par Lomé ;
- il est fait appel à un sondage à un degré pour déterminer l'échantillon maître ;
- les unités de sondage sont les unités primaires constituées par les ZD (zones de dénombrement) ;
- les unités primaires ou zones de dénombrement sont stratifiées selon le milieu de résidence : chaque domaine d'étude est subdivisé en deux strates (une strate urbaine et une strate rurale) à l'exception de Lomé qui compte deux strates urbaines ;
- les zones de dénombrement composant l'échantillon maître sont sélectionnées selon le mode de tirage systématique avec probabilités proportionnelles aux tailles des unités ;
- la répartition de l'échantillon maître entre les strates d'un même domaine d'étude se fera proportionnellement aux tailles des strates en termes d'effectif de ZD.

3.4 Evaluation de la base RGPH 2010

Une évaluation de la base RGPH 2010 a montré qu'elle n'est pas utilisable en l'état dans la construction de l'échantillon maître. En effet, la base comporte plusieurs ZD de très petite taille et plusieurs autres de grande taille. Par rapport à ces deux catégories de ZD, des aménagements de la base ont été retenus, certains en amont de l'échantillonnage devant conduire à la construction de l'échantillon maître, et d'autres en aval du tirage de l'échantillon maître pour en faciliter l'utilisation future comme base de sondage.

L'évaluation de la base RGPH 2010 a révélé l'existence de 70 ZD de moins de 50 ménages. Le rattachement de chacune de ces ZD à d'autres ZD a été recommandé afin que la base ne compte désormais plus que des ZD de 50 ménages ou plus.

Ce travail d'aménagement de la base vis-à-vis des ZD de très petite taille a été réalisé par la Division de la Cartographie au moment¹ de l'élaboration du présent document technique. Il a conduit à une base aménagée dont les caractéristiques sont présentées dans les tableaux 1 et 2.

¹ Ce travail d'aménagement de la base a été réalisé entre le 3 et le 12 juin 2015.

Concernant les ZD de grande taille on observe que certaines comptent jusqu'à 800 ménages. Celles qui comptent 200 ménages ou plus représentent 33% de la base RGPH 2010. Il est recommandé de procéder, en aval du tirage de l'échantillon maître, à une segmentation systématique de ces ZD de 200 ménages ou plus. Selon sa taille ou son étendue géographique, une ZD de 200 ménages ou plus comportera 2 à 5 segments. La procédure de segmentation a pour but de faciliter les travaux de mise à jour de l'échantillon maître ainsi que les travaux de mise en œuvre des différents plans de sondage à venir.

3.5 Répartition de l'échantillon maître entre les strates

La répartition de l'échantillon maître selon les six domaines d'étude est celle fournie par le tableau 4 à la dernière colonne (allocation de Kish).

Il est fait appel à un sondage stratifié représentatif pour répartir le sous-échantillon des ZD d'un domaine d'étude entre ses deux strates. Le plan de sondage porte sur 12 strates définies en colonne 3 du tableau 5. La taille relative des strates au sein de chaque domaine est celle définie par la répartition des ZD dans la base RGPH 2010 et qui figure à la colonne 5 du tableau 5.

Le tableau 5 présente la répartition de l'échantillon maître selon la strate ainsi que le taux de sondage respectivement dans les colonnes 8 et 9. On observe que le taux de sondage au niveau national est de 0,24, ce qui signifie que l'échantillon maître représente environ le quart de la base RGPH 2010. Au niveau des domaines d'étude, le taux de sondage varie entre 0,21 pour la Région des Plateaux et Lomé et 0,32 pour la Région Centrale.

Tableau 5 : Répartition de l'échantillon maître selon la strate

Domaine	Milieu	Strate	Répartition des ZD dans la base		Effectif des ZD à tirer dans		Effectif ajusté des ZD à tirer par strate	Taux de sondage
			Effectif	Proportion	le domaine	la strate		
Maritime	Urbain	Strate11	127	0,10		27,86	28	0,22
	Rural	Strate12	1 131	0,90		248,14	248	0,22
	Sous-total		1 258	1,00	276	276	276	0,22
Plateaux	Urbain	Strate21	257	0,17		53,10	53	0,21
	Rural	Strate22	1 258	0,83		259,90	260	0,21
	Sous-total		1 515	1,00	313	313	313	0,21
Centrale	Urbain	Strate31	125	0,18		40,01	40	0,32
	Rural	Strate32	556	0,82		177,99	178	0,32
	Sous-total		681	1,00	218	218	218	0,32
Kara	Urbain	Strate41	169	0,20		47,44	47	0,28
	Rural	Strate42	661	0,80		185,56	186	0,28
	Sous-total		830	1,00	233	233	233	0,28
Savanes	Urbain	Strate51	109	0,13		31,93	32	0,29
	Rural	Strate52	707	0,87		207,07	207	0,29
	Sous-total		816	1,00	239	239	239	0,29
Lomé	Lomé Centre	Strate61	870	0,56		179,36	179	0,21
	Lomé Périphérie	Strate62	687	0,44		141,64	142	0,21
	Sous-total		1 557	1,00	321	321	321	0,21
Ensemble	Urbain		2 344	0,35		521	521	0,22
	Rural		4 313	0,65		1 079	1 079	0,25
	Total		6 657	1,00	1 600	1 600	1 600	0,24

3.6 Tirage de l'échantillon maître

La sélection des 1600 ZD de l'échantillon maître est réalisée par tirage systématique avec probabilités proportionnelles aux tailles des ZD sur la base de la répartition définie par le tableau 5 en colonne 8. Il a été fait appel au logiciel² TIRAGE 2.1 pour la mise en œuvre de la sélection de l'échantillon.

Le tirage de l'échantillon a été réalisé strate par strate pour les 12 strates de l'univers sondé. On a pris soin de vérifier, si en fonction de la répartition retenue pour l'échantillon, certaines strates possèdent des ZD atypiques, c'est-à-dire des ZD dont la probabilité d'inclusion est supérieure à 1 et donc susceptibles d'être tirées plus d'une fois. On a observé l'existence de 5 ZD atypiques dans cinq des strates de grande taille. Le tableau 6 présente les 5 ZD atypiques et leurs caractéristiques.

Tableau 6 : ZD atypiques rencontrées lors du tirage de l'EM

Strate	IDZD	MENAGES	Probabilité d'inclusion
Strate 12	101015052	810	1,003988
Strate 22	206010052	840	1,025352
Strate 32	303050152	459	1,053187
Strate 42	401010102	586	1,101860
Strate 52	504050102	589	1,321773

Les 5 ZD atypiques rencontrées sont prises d'office et considérées comme tirées avec la probabilité 1, c'est-à-dire obtenues de façon certaine. Dans chaque strate concernée, la ZD atypique est soustraite et l'on procède ensuite à la sélection du reste des ZD à tirer.

La liste des 1600 ZD tirées est fournie dans un fichier Excel séparé du présent document technique et dénommé "EchantillonMaître_Documenté". Un extrait de deux pages de ce fichier est présenté en annexe 2.

3.7 Probabilités d'inclusion des ZD de l'échantillon maître

Les ZD constitutives de l'échantillon maître sont issues de la base RGPH 2010 par tirage aléatoire. Elles ont, de ce fait, avec cette base originelle, un lien de parenté qu'il conviendra de retrouver facilement pour toute ZD de l'échantillon maître chaque fois qu'elle sera considérée. Deux éléments définissent le lien d'une ZD de l'échantillon maître avec la base originelle. Ce sont :

- l'identifiant de la ZD dans la base originelle ;
- et la probabilité d'inclusion (ou probabilité de sélection) de la ZD.

L'identifiant IDZD d'une ZD de la base RGPH 2010 est un code numérique obtenu par concaténation des codes de la région, de la préfecture, du canton, de la ZD et du milieu de résidence. Le code de la ZD au sein de l'identifiant IDZD indique le positionnement de cette ZD à l'intérieur du canton. Il s'ensuit que la seule connaissance de l'identifiant IDZD d'une ZD suffit pour donner la position géographique de la ZD dans l'univers représenté par la base RGPH.

² Le logiciel TIRAGE 2.1 est développé par Julien Amegandjin avec une première version produite en 2005, et la dernière version en 2009. Il peut être téléchargé du site d'AFRIATAT (www.afristat.org).

La probabilité d'inclusion d'une ZD de l'échantillon maître, c'est-à-dire sa probabilité d'appartenance à l'échantillon maître, est le deuxième élément du lien de la ZD avec la base mère. La probabilité d'inclusion d'une ZD de l'échantillon maître résume dans une certaine mesure, la manière dont la ZD est engendrée par la base originelle par tirage aléatoire.

La suite du présent paragraphe est consacrée à la méthode de calcul des probabilités d'inclusion des ZD constitutives de l'échantillon maître. A cet effet, les notations ci-après seront utilisées :

- h désigne la strate dans un domaine d'étude ;
- H est le nombre des strates du domaine d'étude ;
- la strate h est composée de M_h unités primaires ou ZD d'étiquettes $1, 2, \dots, M_h$;
- m_h est l'effectif des unités primaires ou ZD tirées dans la strate h ;
- l'unité primaire i de la strate h sera notée UP_{hi} ;
- N_{hi} désigne la taille de l'unité primaire UP_{hi} ;
- la taille N_{hi} est pour la base de sondage utilisée, l'effectif des ménages de l'unité primaire UP_{hi} ;
- N_h désigne la somme des tailles N_{hi} des unités primaires UP_{hi} et est défini par la relation

$$N_h = \sum_{i=1}^{M_h} N_{hi} \quad (6)$$

- P_{hi} désigne la probabilité d'inclusion de l'unité primaire UP_{hi} dans l'échantillon maître.

On démontre³ que la probabilité d'inclusion P_{hi} a pour expression

$$P_{hi} = m_h \frac{N_{hi}}{N_h} \quad (7)$$

La relation (7) sera utilisée pour toute ZD i de la strate h si cette dernière ne compte pas de ZD atypiques.

Il a été observé que 5 strates possèdent des ZD atypiques de probabilité d'inclusion supérieure à 1. Pour toute ZD atypique de l'une de ces strates, la probabilité d'inclusion P_{hi} a pour expression

$$P_{hi} = 1 \quad (8)$$

pour être sélectionnée d'office.

Pour toutes les autres ZD d'une strate qui compte au moins une ZD atypique, la probabilité d'inclusion P_{hi} aura pour expression

$$P_{hi} = (m_h - n_h) \frac{N_{hi}}{(N_h - S_h)} \quad (9)$$

où n_h et S_h sont respectivement le nombre de ZD atypiques et le cumul des tailles des ZD atypiques de la strate h . L'on notera que la relation (9) se ramène à la relation (7) pour $n_h = 0$ et $S_h = 0$, c'est-à-dire pour toute strate h ne possédant pas de ZD atypique.

³ Voir le manuel "Pratique des Sondages" (page 32) de J. Amegandjin, Editions AFRISTAT, Bamako, Mars 2014

Pour chacune des ZD de l'échantillon maître, la probabilité d'inclusion P_{hi} sera calculée au moyen de celle des trois relations qui s'applique.

L'ensemble des probabilités d'inclusion calculées constituera une des caractéristiques de définition de l'échantillon maître. Elles figureront aux côtés de l'identifiant IDZD et d'un deuxième identifiant NUM_EM pour définir l'échantillon maître, NUM_EM n'étant autre que le numéro d'ordre séquentiel des ZD de l'échantillon maître.

4. Activités d'aménagement des ZD de l'échantillon maître

L'échantillon maître est une nouvelle base de sondage extraite de la base RGPH 2010. Afin de pouvoir servir à l'élaboration des plans de sondage des futures enquêtes à mener dans le pays, elle doit être organisée. Des activités d'aménagement de l'échantillon maître sont donc nécessaires dès sa création. Au nombre de ces activités, l'on retiendra, en particulier :

- l'actualisation des cartes des ZD constitutives ;
- la création de segments dans les ZD de grande taille ;
- l'actualisation de la liste des concessions ;
- l'actualisation de la liste des ménages ;
- l'actualisation de la liste des ménages agricoles dans les ZD rurales.

L'issue de ces travaux d'aménagement de l'échantillon maître devra déboucher sur une base de sondage bien structurée avec prise en compte des trois unités de sondage, à savoir :

- les unités primaires ou ZD (les 1600 ZD qui définissent l'échantillon maître) ;
- les unités secondaires qui sont les segments créés ;
- les unités tertiaires qui sont les ménages dénombrés dans les segments.

La mise à jour périodique de l'échantillon maître devrait intervenir une fois l'an et consister en l'actualisation des cartes et des listes des concessions et des ménages dans les segments.

5. Mode d'emploi de l'échantillon maître

Selon son format, l'échantillon maître va donner lieu tout naturellement à des plans de sondage à trois degrés, les unités primaires étant les ZD, les unités secondaires, les segments, et les unités tertiaires, les ménages.

La notion de segment sera généralisée de sorte que même les ZD non segmentées soient considérées comme des ZD à un segment. Au deuxième degré, on sélectionnera un et un seul segment, et au troisième degré, les ménages seront tirés du segment échantillonné.

Ce mode d'emploi est conforme au standard classique des plans de sondage en vigueur pour les enquêtes courantes de notre région telles que l'EDS, le QUIBB, le MICS ou l'enquête sur l'emploi et le secteur informel.

Annexe 1 : Extrait du tableau des probabilités d'inclusion des ZD de l'EM (2 pages sur 40)

Numéro d'ordre de la ZD dans l'EM	Identifiant de la ZD dans la base	Numéro d'ordre de la ZD dans la base	STRATE	Milieu de résidence	Effectif des ZD tirées dans la strate	Taille de la ZD (Effectif des ménages)	Taille de la strate (Effectif des ménages)	Nombre de ZD atypiques de la strate	Cumul des tailles des ZD atypiques de la strate	Probabilité d'inclusion de la ZD
NUM_EM	IDZD	NUM_BASE			m_h	N_{hi}	N_h	n_h	S_h	$P_{hi} = (m_h - n_h) * N_{hi} / (N_h - S_h)$
1	102010031	1650	Strate11	1	28	199	31 911	-	-	0,1746
2	102010081	1655	Strate11	1	28	186	31 911	-	-	0,1632
3	102010131	1660	Strate11	1	28	256	31 911	-	-	0,2246
4	102010171	1664	Strate11	1	28	209	31 911	-	-	0,1834
5	102010221	1669	Strate11	1	28	347	31 911	-	-	0,3045
6	102010261	1673	Strate11	1	28	180	31 911	-	-	0,1579
7	103010031	1837	Strate11	1	28	573	31 911	-	-	0,5028
8	104010021	1942	Strate11	1	28	179	31 911	-	-	0,1571
9	104010051	1948	Strate11	1	28	797	31 911	-	-	0,6993
10	104010101	1956	Strate11	1	28	232	31 911	-	-	0,2036
11	104010131	1961	Strate11	1	28	869	31 911	-	-	0,7625
12	105010031	2159	Strate11	1	28	282	31 911	-	-	0,2474
13	105010071	2164	Strate11	1	28	265	31 911	-	-	0,2325
14	105010121	2169	Strate11	1	28	217	31 911	-	-	0,1904
15	105010201	2176	Strate11	1	28	241	31 911	-	-	0,2115
16	106005011	2363	Strate11	1	28	337	31 911	-	-	0,2957
17	106005061	2373	Strate11	1	28	231	31 911	-	-	0,2027
18	106005101	2377	Strate11	1	28	395	31 911	-	-	0,3466
19	106005141	2381	Strate11	1	28	220	31 911	-	-	0,1930
20	106005191	2386	Strate11	1	28	270	31 911	-	-	0,2369
21	106005241	2391	Strate11	1	28	198	31 911	-	-	0,1737
22	106005291	2396	Strate11	1	28	234	31 911	-	-	0,2053
23	106005341	2401	Strate11	1	28	318	31 911	-	-	0,2790
24	106005401	2407	Strate11	1	28	192	31 911	-	-	0,1685
25	106005461	2413	Strate11	1	28	202	31 911	-	-	0,1772
26	106005511	2418	Strate11	1	28	188	31 911	-	-	0,1650
27	106005561	2423	Strate11	1	28	254	31 911	-	-	0,2229
28	107010011	2706	Strate11	1	28	399	31 911	-	-	0,3501

Numéro d'ordre de la ZD dans l'EM	Identifiant de la ZD dans la base	Numéro d'ordre de la ZD dans la base	STRATE	Milieu de résidence	Effectif des ZD tirées dans la strate	Taille de la ZD (Effectif des ménages)	Taille de la strate (Effectif des ménages)	Nombre de ZD atypiques de la strate	Cumul des tailles des ZD atypiques de la strate	Probabilité d'inclusion de la ZD
NUM_EM	IDZD	NUM_BASE			m_h	N_{hi}	N_h	n_h	S_h	$P_{hi} = (m_h - n_h) * N_{hi} / (N_h - S_h)$
29	101015032	973	Strate12	2	248	387	200 082	1	810	0,4797
30	101015092	977	Strate12	2	248	230	200 082	1	810	0,2851
31	101032012	1082	Strate12	2	248	169	200 082	1	810	0,2095
32	101032052	1086	Strate12	2	248	175	200 082	1	810	0,2169
33	101032102	1091	Strate12	2	248	212	200 082	1	810	0,2628
34	101032132	1094	Strate12	2	248	402	200 082	1	810	0,4983
35	101032162	1097	Strate12	2	248	480	200 082	1	810	0,5950
36	101032222	1101	Strate12	2	248	212	200 082	1	810	0,2628
37	101032262	1105	Strate12	2	248	219	200 082	1	810	0,2715
38	101032312	1110	Strate12	2	248	492	200 082	1	810	0,6098
39	101040022	1113	Strate12	2	248	142	200 082	1	810	0,1760
40	101040072	1116	Strate12	2	248	267	200 082	1	810	0,3309
41	101040122	1120	Strate12	2	248	207	200 082	1	810	0,2566
42	101040152	1123	Strate12	2	248	272	200 082	1	810	0,3371
43	101040212	1128	Strate12	2	248	129	200 082	1	810	0,1599
44	101040262	1133	Strate12	2	248	333	200 082	1	810	0,4128
45	101040282	1135	Strate12	2	248	246	200 082	1	810	0,3049
46	101040312	1138	Strate12	2	248	101	200 082	1	810	0,1252
47	101052052	1174	Strate12	2	248	181	200 082	1	810	0,2244
48	101052082	1177	Strate12	2	248	274	200 082	1	810	0,3396
49	101052132	1182	Strate12	2	248	163	200 082	1	810	0,2020
50	101052192	1188	Strate12	2	248	172	200 082	1	810	0,2132
51	102020012	1674	Strate12	2	248	226	200 082	1	810	0,2801
52	102020042	1677	Strate12	2	248	158	200 082	1	810	0,1958
53	102020082	1681	Strate12	2	248	212	200 082	1	810	0,2628
54	102020122	1685	Strate12	2	248	251	200 082	1	810	0,3111
55	102020162	1689	Strate12	2	248	218	200 082	1	810	0,2702
56	102020202	1693	Strate12	2	248	102	200 082	1	810	0,1264
57	102020242	1697	Strate12	2	248	371	200 082	1	810	0,4599
58	102020272	1700	Strate12	2	248	382	200 082	1	810	0,4735
59	102020322	1705	Strate12	2	248	182	200 082	1	810	0,2256
60	102020372	1710	Strate12	2	248	251	200 082	1	810	0,3111
61	102020402	1713	Strate12	2	248	227	200 082	1	810	0,2814
62	102020442	1717	Strate12	2	248	203	200 082	1	810	0,2516

Annexe 2 : Extrait de la liste des ZD composant l'échantillon maître (avec 15 champs sur 19)

NUM_EM	IDZD	NUM_BASE	STRATE	RANG_TIR	MILIEU	MENAGES	POPULATION	CODE_REG	NOM_REG	CODE_PRE	PREFECTUR	CODE_CAN	VILLE_CANTON	CODE_ZD
1	102010031	1650	Strate11	1	1	199	762	1	MARITIME	2	LACS	10	ANEHO	3
2	102010081	1655	Strate11	2	1	186	685	1	MARITIME	2	LACS	10	ANEHO	8
3	102010131	1660	Strate11	3	1	256	821	1	MARITIME	2	LACS	10	ANEHO	13
4	102010171	1664	Strate11	4	1	209	771	1	MARITIME	2	LACS	10	ANEHO	17
5	102010221	1669	Strate11	5	1	347	1773	1	MARITIME	2	LACS	10	ANEHO	22
6	102010261	1673	Strate11	6	1	180	859	1	MARITIME	2	LACS	10	ANEHO	26
7	103010031	1837	Strate11	7	1	573	2356	1	MARITIME	3	BAS-MONO	10	AFAGNAGAN	3
8	104010021	1942	Strate11	8	1	179	617	1	MARITIME	4	VO	10	VOGAN	2
9	104010051	1948	Strate11	9	1	797	3633	1	MARITIME	4	VO	10	VOGAN	5
10	104010101	1956	Strate11	10	1	232	1081	1	MARITIME	4	VO	10	VOGAN	10
11	104010131	1961	Strate11	11	1	869	3246	1	MARITIME	4	VO	10	VOGAN	13
12	105010031	2159	Strate11	12	1	282	1198	1	MARITIME	5	YOTO	10	TABLIGBO	3
13	105010071	2164	Strate11	13	1	265	1111	1	MARITIME	5	YOTO	10	TABLIGBO	7
14	105010121	2169	Strate11	14	1	217	1035	1	MARITIME	5	YOTO	10	TABLIGBO	12
15	105010201	2176	Strate11	15	1	241	922	1	MARITIME	5	YOTO	10	TABLIGBO	20
16	106005011	2363	Strate11	16	1	337	1297	1	MARITIME	6	ZIO	5	TSEVIE	1
17	106005061	2373	Strate11	17	1	231	886	1	MARITIME	6	ZIO	5	TSEVIE	6
18	106005101	2377	Strate11	18	1	395	1494	1	MARITIME	6	ZIO	5	TSEVIE	10
19	106005141	2381	Strate11	19	1	220	946	1	MARITIME	6	ZIO	5	TSEVIE	14
20	106005191	2386	Strate11	20	1	270	1055	1	MARITIME	6	ZIO	5	TSEVIE	19
21	106005241	2391	Strate11	21	1	198	824	1	MARITIME	6	ZIO	5	TSEVIE	24
22	106005291	2396	Strate11	22	1	234	896	1	MARITIME	6	ZIO	5	TSEVIE	29
23	106005341	2401	Strate11	23	1	318	1196	1	MARITIME	6	ZIO	5	TSEVIE	34
24	106005401	2407	Strate11	24	1	192	690	1	MARITIME	6	ZIO	5	TSEVIE	40
25	106005461	2413	Strate11	25	1	202	715	1	MARITIME	6	ZIO	5	TSEVIE	46
26	106005511	2418	Strate11	26	1	188	690	1	MARITIME	6	ZIO	5	TSEVIE	51
27	106005561	2423	Strate11	27	1	254	995	1	MARITIME	6	ZIO	5	TSEVIE	56

NUM_EM	IDZD	NUM_BASE	STRATE	RANG_TIR	MILIEU	MENAGES	POPULATION	CODE_REG	NOM_REG	CODE_PRE	PREFECTUR	CODE_CAN	VILLE_CANTON	CODE_ZD
28	107010011	2706	Strate11	28	1	399	1378	1	MARITIME	7	AVE	10	KEVE	1
29	101015032	973	Strate12	1	2	387	1881	1	MARITIME	1	GOLFE	15	AFLAO-SAGBAD	3
30	101015092	977	Strate12	2	2	230	1086	1	MARITIME	1	GOLFE	15	AFLAO-SAGBAD	9
31	101032012	1082	Strate12	3	2	169	822	1	MARITIME	1	GOLFE	32	SANGUERA RUF	1
32	101032052	1086	Strate12	4	2	175	843	1	MARITIME	1	GOLFE	32	SANGUERA RUF	5
33	101032102	1091	Strate12	5	2	212	849	1	MARITIME	1	GOLFE	32	SANGUERA RUF	10
34	101032132	1094	Strate12	6	2	402	1857	1	MARITIME	1	GOLFE	32	SANGUERA RUF	13
35	101032162	1097	Strate12	7	2	480	1865	1	MARITIME	1	GOLFE	32	SANGUERA RUF	16
36	101032222	1101	Strate12	8	2	212	899	1	MARITIME	1	GOLFE	32	SANGUERA RUF	22
37	101032262	1105	Strate12	9	2	219	957	1	MARITIME	1	GOLFE	32	SANGUERA RUF	26
38	101032312	1110	Strate12	10	2	492	2344	1	MARITIME	1	GOLFE	32	SANGUERA RUF	31
39	101040022	1113	Strate12	11	2	142	643	1	MARITIME	1	GOLFE	40	LEGBASSITO	2
40	101040072	1116	Strate12	12	2	267	1341	1	MARITIME	1	GOLFE	40	LEGBASSITO	7
41	101040122	1120	Strate12	13	2	207	960	1	MARITIME	1	GOLFE	40	LEGBASSITO	12
42	101040152	1123	Strate12	14	2	272	1197	1	MARITIME	1	GOLFE	40	LEGBASSITO	15
43	101040212	1128	Strate12	15	2	129	542	1	MARITIME	1	GOLFE	40	LEGBASSITO	21
44	101040262	1133	Strate12	16	2	333	1580	1	MARITIME	1	GOLFE	40	LEGBASSITO	26
45	101040282	1135	Strate12	17	2	246	985	1	MARITIME	1	GOLFE	40	LEGBASSITO	28
46	101040312	1138	Strate12	18	2	101	525	1	MARITIME	1	GOLFE	40	LEGBASSITO	31
47	101052052	1174	Strate12	19	2	181	858	1	MARITIME	1	GOLFE	52	TOGGLEKOPE R	5
48	101052082	1177	Strate12	20	2	274	1064	1	MARITIME	1	GOLFE	52	TOGGLEKOPE R	8
49	101052132	1182	Strate12	21	2	163	871	1	MARITIME	1	GOLFE	52	TOGGLEKOPE R	13
50	101052192	1188	Strate12	22	2	172	938	1	MARITIME	1	GOLFE	52	TOGGLEKOPE R	19
51	102020012	1674	Strate12	23	2	226	1177	1	MARITIME	2	LACS	20	AGBODRAFO	1
52	102020042	1677	Strate12	24	2	158	708	1	MARITIME	2	LACS	20	AGBODRAFO	4
53	102020082	1681	Strate12	25	2	212	898	1	MARITIME	2	LACS	20	AGBODRAFO	8
54	102020122	1685	Strate12	26	2	251	1116	1	MARITIME	2	LACS	20	AGBODRAFO	12
55	102020162	1689	Strate12	27	2	218	1012	1	MARITIME	2	LACS	20	AGBODRAFO	16
56	102020202	1693	Strate12	28	2	102	475	1	MARITIME	2	LACS	20	AGBODRAFO	20
57	102020242	1697	Strate12	29	2	371	1651	1	MARITIME	2	LACS	20	AGBODRAFO	24