

# **Elaboration du plan de sondage EPCV 2008**

## **Elaboration du plan de sondage EPCV 2008 de la Mauritanie**

Nouakchott, 29 mars – 10 avril 2008

Tiral SIDI, consultant

E-mail : [mstiral@yahoo.fr](mailto:mstiral@yahoo.fr)

### **1. Base de sondage**

La Mauritanie est subdivisée en 13 Wilaya réparties elles-mêmes en 53 Moughataa. Des travaux cartographiques censitaires réalisés dans le cadre du recensement général de la population et de l'habitat de 2000 ont fourni une subdivision du territoire national en districts de dénombrement (DR). Les zones de dénombrement sont des unités aréolaires distinctes constituées de manière à contenir chacune, environ un millier d'habitants pour les besoins d'organisation et de répartition des tâches de collecte des données. Le recensement de la population proprement dit de 2000 a fourni l'effectif des ménages ainsi que l'effectif de la population pour chaque ZD. Selon les résultats de ce recensement, au total 2 290 DR ont été constitués et 428 172 ménages ont été recensés (tableau 1).

Les fichiers des DR et des ménages issus de ce RGPH 2000 constituent les bases de sondage de l'EPCV 2008. Le tableau 3 donne la répartition des DR et des ménages selon la Moughataa et le milieu de résidence.

Pour les besoins des tirages de l'échantillon des DR (ou échantillon du premier degré), des DR ont été disqualifiés parce que leurs tailles (en nombre de ménages) ont été jugées trop faibles pour assurer des échantillons de ménages de la taille requise pour ces DR. Ainsi, 28 DR possédant moins de 30 ménages ont été écartés. Cette limite a été choisie en tenant compte du nombre de 20 ménages à tirer sans remise dans chaque DR échantillon, auquel nombre nous avons ajouté une sécurité de 10 ménages. En plus de ces 28 DR, 4 autres DR (1361500, 1361501, 1361502 et 1131432) pour lesquelles il n'existe pas de carte, ont été exclues du tirage. Il en est de même du DR 241403 complètement vide<sup>1</sup> du fait d'une inondation survenue dans ce DR lors de la campagne 2007-2008.

---

<sup>1</sup> Information donnée par l'ONS le 09/04/2008

Tableau 1 : Répartition des DR et des ménages par Moughataa et par milieu de résidence

Wilaya/Moughataa	total DR			ménages RGPH 2000		
	urbain	rural	total	urbain	rural	total
<b>Hodh Echarghi</b>	48	189	237	8 817	40 155	48 972
Amouri	8	45	53	1 612	10 949	12 561
Bassiknou	8	23	31	1 615	4 298	5 913
Diequenni	6	35	41	1 055	6 482	7 537
Nema	13	32	45	2 673	8 689	11 362
Oualata	-	5	5	.	831	831
Temberdra	13	49	62	1 862	8 906	10 768
<b>Hodh El Gharby</b>	18	172	190	3 532	32 852	36 384
Aioune	12	33	45	1 864	6 209	8 073
Koubeni	.	67	67	.	12 022	12 022
Tamechekett	.	21	21	.	4 352	4 352
Tintane	6	51	57	1 668	10 269	11 937
<b>Assaba</b>	43	226	269	7 126	35 035	42 161
Barkeiwei	.	68	68	.	10 703	10 703
Boumdeide	.	10	10	.	1 573	1 573
Guerou	14	22	36	2 244	3 453	5 697
Kenkoussa	.	77	77	.	10 815	10 815
Kiffa	29	49	78	4 883	8 491	13 374
<b>Gorgol</b>	25	182	207	4 544	29 808	34 352
Kaedi	25	44	69	4 544	7 348	11 893
Maghama	.	36	36	.	4 662	4 662
M'bout	.	72	72	.	12 266	12 266
M'onquiel	.	30	30	.	5 532	5 532
<b>Barkna</b>	30	167	197	4 789	31 979	36 768
Aleg	8	34	42	1 572	9 055	10 627
Bababé	.	23	23	.	4 000	4 000
Bohé	13	42	55	1 959	6 989	8 948
Magta-Lahjar	9	41	50	1 258	7 050	8 307
M'bagne	.	27	27	.	4 885	4 885
<b>Trarza</b>	53	183	236	9 677	40 432	50 109
Boutilimitt	13	37	50	2 388	8 025	10 413
Keur-Macen	.	29	29	.	4 904	4 904
Mederdra	.	24	24	.	5 698	5 698
Ouad-Naga	.	21	21	.	5 239	5 239
R'kiz	6	50	56	1 280	10 851	12 132
Rosso	34	22	56	6 009	5 715	11 724
<b>Adrar</b>	27	61	88	4 972	10 492	15 464
Auieft	.	23	23	.	3 107	3 107
Atar	27	27	54	4 972	5 992	10 964
Chinguity	.	7	7	.	812	812
Quadane	.	4	4	.	581	581
<b>Nouadhibou</b>	62	9	71	15 252	1 242	16 493
Nouadhibou	62	9	71	15 252	1 242	16 493
<b>Taoudant</b>	9	66	75	1 882	11 509	13 391
Moudieria	.	38	38	.	6 320	6 320
Tichitt	.	4	4	.	865	865
Tidiikia	9	24	33	1 882	4 324	6 206
<b>Guidimaccha</b>	7	116	123	1 862	17 612	19 473
Oul-Yenie	.	32	32	.	5 489	5 489
Seilibaby	7	84	91	1 862	12 123	13 985
<b>Tiris Zemmour</b>	32	8	40	8 376	1 062	9 438
Bir Mougreine	.	3	3	.	364	364
F'deirik	.	5	5	.	698	698
Zouérate	32	.	32	8 376	.	8 376
<b>Inchiri</b>	5	7	12	1 097	834	1 931
Akiouiet	5	7	12	1 097	834	1 931
<b>Nouakchott</b>	545	-	545	103 234	-	103 234
Tevarett	46	.	46	7 614	.	7 614
Ksar	40	.	40	7 309	.	7 309
Tevrao-Zeina	56	.	56	9 196	.	9 196
Touionine	50	.	50	10 915	.	10 915
Sebkha	73	.	73	12 203	.	12 203
El Mina	103	.	103	19 576	.	19 576
Dar-Naim	52	.	52	11 135	.	11 135
Arafatt	78	.	78	18 278	.	18 278
Rivad	47	.	47	7 007	.	7 007
<b>Mauritanie</b>	<b>904</b>	<b>1 386</b>	<b>2 290</b>	<b>175 160</b>	<b>253 013</b>	<b>428 172</b>

L'élimination des 33 DR a conduit à une base de sondage finale composée de 2257 DR éligibles. Le tirage des DR sera effectué dans cette base. Les DR écartées lors de la redéfinition de la base de sondage auraient pu être regroupées pour constituer d'autres DR viables, mais le grand éloignement de ces ZD les uns des autres et le fait qu'ils ne sont pas toujours contigus a pesé en défaveur de cette option. **Par contre, les 33 DR écartées ainsi que leurs 12571 ménages seront considérés lors de l'estimation du total (ou somme) d'une variable d'étude par l'application d'un coefficient de redressement défini pour chaque Moughataa dans le tableau 3.** Le nombre de ménages de ces DR figurant dans le fichier semblent très sous-estimés, aussi avons-nous considéré la médiane de la taille des DR dans les Moughataa concernées. L'estimation du coefficient de redressement sera faite à partir de la médiane estimée.

Tableau 2 : Coefficient de redressement par Moughataa

Moughataa	coefficient redressement
Amouri	1.01865
Bassikhrou	1
Diequenni	1
Nema	1
Oualata	1
Temberdra	1
Aioune	1
Koubeni	1
Tamechekett	1
Tintane	1.05561
Barkeiwel	1
Boumdeide	1
Guerou	1
Kenkoussa	1
Kiffa	1
Kaedi	1
Maghama	1
M'bout	1
M'onguel	1
Aleg	1
Bababé	1
Boghé	1.01618
Magta-Lahiar	1
M'bagne	1
Boutilimitt	1
Keur-Macen	1
Mederdra	1
Quad-Naga	1
R'kiz	1.01615
Rosso	1.26446
Auieft	1.15663
Atar	1.54954
Chinguity	1
Quadane	1.35262
Nouadhibou	1
Moudleria	1
Tichitt	1
Tidjikia	1
Oul-Yenje	1
Seilibaby	1
Bir Mougreine	1
F'deirik	1
Zouérate	1.30941
Akiouiet	1.07596
Tevarett	1
Ksar	1.02482
Tevrag-Zeina	1.01720
Toujounine	1
Sebkha	1.05557
El Mina	1.06432
Dar-Naim	1.01752
Arafatt	1
Rivad	1

La base de sondage ainsi obtenue d'une grande utilité pour des enquêtes futures sur les conditions de vie des ménages en attendant sa réactualisation au RGPH 2011. Cependant, elle comporte quelques faiblesses dues principalement à son ancienneté. Pour cette raison, l'utilisation de cette base nécessitera une opération d'actualisation préalable qui consistera en un pré-recensement des ménages dans les DR qui auront été tirés au premier degré. Un autre argument qui milite en faveur de cette opération de pré-recensement des ménages vient des mouvements de population qui ont dû intervenir dans les DR depuis le RGPH 2000.

La première étape de l'enquête sur le terrain est consacrée à cette opération de pré-recensement des ménages dans les DR tirés au premier degré. Il en résultera pour chaque DR observé, une liste nominative des chefs de ménage, c'est-à-dire la base de sondage du 2<sup>e</sup> degré de la ZD.

La base de sondage du premier degré a été construite à partir du fichier du RGPH 2000 en suivant les étapes suivantes :

i. Un identifiant a été construit pour chaque district de recensement (DR). Cet identifiant est composé des variables suivantes :

- Wilaya : 1 car
- Moughataa : 1 car
- Commune : 2 car
- Milieu : 1 car
- DR : 3 car

<i>Wilaya</i>	<i>Moughataa</i>	<i>Commune</i>		<i>Milieu</i>	<i>DR</i>		
X	0	0	0	0	0	0	0
	X	0	0	0	0	0	0
		X	X	0	0	0	0
				X	0	0	0
					X	X	X

A partir de ce schéma, l'identifiant des DR a été construit selon la formule suivante :

$$identifDR = Wilaya \times 10000000 + Moughataa \times 1000000 + commune \times 10000 + milieu \times 1000 + DR$$

Une fois l'identifiant créé, la liste des DR est constituée en exécutant à l'aide de SPSS les étapes suivantes :

- ii. Création du fichier ménage en agrégeant le fichier individus par Wilaya, Moughataa, milieu, DR, code ménage (variable à agréger : ménages, individus)

- iii. Création du fichier DR en agrégeant le fichier des ménages par Wilaya, Moughataa, milieu, DR (variable à agréger : ménages, individus). On obtient un fichier qui est la base de sondage des DR avec les variables :
- Wilaya
  - Moughataa
  - Commune
  - Milieu
  - DR
  - nombre de ménages par DR
  - population par DR

Les DR obtenus sont au total 2290. Elles proviennent du RGPH 2000 Mais le tirage ne sera effectué que dans 2257 DR (voir ci-dessus).

NB : Il faut constituer 2 bases de sondage :

- 1 base avec les nomades
- 1 base sans les nomades

En effet, même si le nombre de DR est identique pour ces 2 bases, le nombre de ménages et les populations par DR sont différents pour ceux qui abritent des nomades (128163 d'après RGPH 2000).

## **2. Domaines d'étude**

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.

Deux options ont été analysées: i) Chaque milieu de résidence (urbain/rural) des 53 Moughataa, soit 73 domaines d'études ; ii) Chaque Moughataa des 13 Wilaya, soit 53 domaines d'études. C'est la deuxième option qui a été choisie après des discussions avec les responsables de l'ONS et tenant compte des contraintes budgétaires et de logistique. Ainsi, dans le cas de l'EPCV 2008, les 53 Moughataa sont définis comme des domaines d'étude.

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. Ce critère de stratification n'a pas un objectif de précision locale mais plutôt globale avec toutefois des résultats par milieu au niveau national.

## **3. Type de plan de sondage**

Le plan de sondage sera à deux degrés. Au premier degré les DR seront choisis selon un tirage systématique à probabilités proportionnelles à leur taille en ménages. Au deuxième degré un nombre constant de ménages sera tiré à probabilité égale sans remise.

## 4. Taille de l'échantillon

### 4.1 Taille de l'échantillon des ménages par domaine d'étude (Moughataa)

La taille de l'échantillon nécessaire pour cette enquête est déterminée par ses objectifs. Il a été également tenu compte de quatre critères principaux :

- ✓ la précision que l'on souhaite pour les estimations,
- ✓ les indicateurs prioritaires pour lesquels il est indispensable de réaliser une estimation précise,
- ✓ le niveau de confiance choisi (risque d'erreur de première espèce, souvent fixé à 5 %). Un risque d'erreur plus faible nécessite un échantillon plus important pour une même précision,
- ✓ les moyens dont on dispose pour conduire l'enquête.

Le taux de non réponse et l'effet de sondage sont aussi importants parce que leur influence est nette sur la taille de l'échantillon.

Les objectifs de précision visée sont donc exprimés selon les 53 strates principales d'analyse, soit les Moughataa, ce qui donne lieu à l'allocation première de l'échantillon selon les Moughataa.

La formule de calcul de la taille minimale requise pour un objectif de précision visée est, pour chaque strate principale (Moughataa) :

$$n_{requis} = \frac{Nn^*CV_*^2\left(\hat{\bar{X}}\right)}{N\rho^2 + n^*CV_*^2\left(\hat{\bar{X}}\right)} = \frac{\frac{n^*CV_*^2\left(\hat{\bar{X}}\right)}{\rho^2}}{1 + \frac{n^*CV_*^2\left(\hat{\bar{X}}\right)}{N\rho^2}} \quad (4)$$

$$\text{Soit } n_0 = \frac{n^*CV_*^2\left(\hat{\bar{X}}\right)}{\rho^2}$$

$$\text{Alors } n_{requis} = \frac{n_0}{1 + \frac{n_0}{N}} \quad (\text{correction pour population finie})$$

Où:

- a)  $CV_*^2\left(\hat{\bar{X}}\right)$  = le coefficient de variation, au carré, de l'estimateur de moyenne  $\hat{\bar{X}}$ . Ce coefficient est estimé à partir de l'EPCV 2004, en prenant en compte la structure complexe de l'échantillon de cette dernière enquête.
- b)  $n^*$  est la taille de l'échantillon de l'EPCV 2004, dans la Moughataa considérée.
- c)  $\rho$  = le coefficient de variation choisi comme objectif de précision pour l'enquête EPCV 2008.

L'équation (4) est développée facilement à partir de l'expression de l'effet de structure (DEFF)<sup>2</sup> dans chaque strate, ratio de la variance de l'estimateur incorporant la structure complexe de l'échantillon et de la variance qui aurait été obtenue avec un échantillon simplement aléatoire de même taille, en termes de nombre de ménages.

L'estimation des différents termes requis dans l'équation (4) est faite à partir de l'enquête EPCV 2004 qui a une structure de sondage comparable à la structure de sondage de l'EPCV 2008, ce qui en rend crédites les DEFF estimés sur lesquels s'appuie l'équation (4) ci-dessus. Nous notons cependant que les précisions n'ont pas été calculées pour cette enquête dont les domaines d'étude sont inclus dans ceux de l'EPCV 2008.

Les objectifs de précision par Moughataa, exprimés par le CV visé, ont été fixés selon les principes suivants:

- pour chaque variable, on vise le même CV dans chacune des 53 Moughataa,
- ce CV commun a été fixé à 0,05.

Différentes tailles de l'échantillon ont été estimées en prenant en compte la variable dépenses et avec en plus le souci d'une bonne représentation du milieu de résidence et des GSE dans l'échantillon pour assurer des résultats significatifs au niveau national. La taille obtenue en tenant compte de ces considérations a permis de réajuster la première taille pour finalement déterminer la taille de l'échantillon qui assure des résultats significatifs :

- dans chaque Moughataa,
- dans milieu de résidence au niveau des Wilaya
- dans chaque GSE au niveau national ;

Le GSE ne peut être utilisé pour la stratification du fait que l'information n'est pas disponible pour toutes les unités et même connue, en la croisant avec le milieu de résidence, on obtient 14 strates par Moughataa, ce qui est beaucoup).

On obtient un échantillon global est de 14 040.

Une stratification par Moughataa suivant le milieu de résidence sera prise en compte avec seulement un objectif de précision au niveau global de la Moughataa. L'allocation proportionnelle à la taille en ménages sera utilisée.

#### **4.2 Taille de l'échantillon des UP (DR) par domaine d'étude (Moughataa)**

Après l'estimation de l'échantillon des US(ménages) dans chaque milieu de résidence pour chaque Moughataa, on effectue les étapes suivantes :

1. Traduire le nombre  $N_{uss}$  d'USS (ménages) en nombre  $N_{ups}$  d'UPS (DR) par simple division de  $N_{uss}$  par  $n$  (nombre d'USS à tirer dans chaque UPS) :

$$N_{ups} = \frac{N_{uss}}{n}$$

---

<sup>2</sup> DEFF: Design Effect.

2. Comparer  $N_{ups}$  à  $N_{up}$  (nombre total d'UP dans le domaine d'étude considéré) dans chaque domaine d'étude :

- Si  $N_{ups} \leq N_{up}$ , alors  $N_{ups}^* = N_{ups}$

- Si  $N_{ups} > N_{up}$ , alors  $N_{ups}^* = N_{up}$

3. On ajuste la taille de l'échantillon dans chaque domaine d'étude selon la formule suivante :  $N_{uss}^* = nN_{ups}^*$

NB : Après cet ajustement de la taille, il peut s'avérer parfois utile de suréchantillonner certains domaines d'études en augmentant le nombre USS à tirer dans chaque UPS.



Tableau 3 : Répartition des DR et des ménages tirés par Moughataa et par milieu de résidence

Wilaya/Moughataa	DR			ménages échantillons		
	urbain	rural	total	urbain	rural	total
<b>Hodh Echraqhi</b>	13	47	60	260	940	1 200
Amouri	2	9	11	40	180	220
Bassiknou	3	8	11	60	160	220
Dieguenni	2	7	9	40	140	180
Nema	3	4	7	60	80	140
Oualata		5	5	0	100	100
Temberdra	3	14	17	60	280	340
<b>Hodh El Gharby</b>	18	50	68	360	1 000	1 360
Aioune	12	11	23	240	220	460
Koubeni		9	9	0	180	180
Tamechekett		15	15	0	300	300
Tintane	6	15	21	120	300	420
<b>Assaba</b>	25	39	64	500	780	1 280
Barkeiwel		7	7	0	140	140
Boumdeide		9	9	0	180	180
Guerou	10	6	16	200	120	320
Kenkoussa		9	9	0	180	180
Kiffa	15	8	23	300	160	460
<b>Gorgol</b>	16	45	61	320	900	1 220
Kaedi	16	10	26	320	200	520
Maghama		11	11	0	220	220
M'bout		13	13	0	260	260
M'onquel		11	11	0	220	220
<b>Barkna</b>	29	49	78	580	980	1 560
Aleg	8	8	16	160	160	320
Bababé		9	9	0	180	180
Boghé	12	11	23	240	220	460
Maqta-Lahiar	9	9	18	180	180	360
M'bagne		12	12	0	240	240
<b>Trarza</b>	27	69	96	540	1 380	1 920
Boutilimitt	6	7	13	120	140	260
Keur-Macen		7	7	0	140	140
Mederdra		14	14	0	280	280
Ouad-Naga		15	15	0	300	300
R'kiz	5	15	20	100	300	400
Rosso	16	11	27	320	220	540
<b>Adrar</b>	15	26	41	300	520	820
Auieft		7	7	0	140	140
Atar	15	10	25	300	200	500
Chinquity		5	5	0	100	100
Ouadane		4	4	0	80	80
<b>Nouadhibou</b>	15	6	21	300	120	420
Nouadhibou	15	6	21	300	120	420
<b>Taggant</b>	9	16	25	180	320	500
Moudieria		6	6	0	120	120
Tichitt		4	4	0	80	80
Tidiikia	9	6	15	180	120	300
<b>Guidimagha</b>	7	25	32	140	500	640
Oul-Yenie		10	10	0	200	200
Seilibaby	7	15	22	140	300	440
<b>Tiris Zemmour</b>	12	8	20	240	160	400
Bir Mougreine		3	3	0	60	60
F'derik		5	5	0	100	100
Zouérate	12		12	240	0	240
<b>Inchiri</b>	5	7	12	100	140	240
Akiouiet	5	7	12	100	140	240
<b>Nouakchott</b>	124		124	2 480		2 480
Tevarett	12		12	240		240
Ksar	15		15	300		300
Tevraq-Zeina	16		16	320		320
Touounine	12		12	240		240
Sebkha	11		11	220		220
El Mina	16		16	320		320
Dar-Naim	18		18	360		360
Arafatt	8		8	160		160
Riyad	16		16	320		320
<b>Mauritanie</b>	<b>315</b>	<b>387</b>	<b>702</b>	<b>6 300</b>	<b>7 740</b>	<b>14 040</b>

## 5. Tirage des échantillons

- **UPSS (DR)** : tirage systématique à probabilités proportionnelles à la taille (nombre de ménages de chaque UPS donnés par le RGPH 2000)
- **USS (ménages)** : tirage systématique à probabilité égale d'un nombre de ménages (en général fixe) dans chaque UPSS

### 5.1 Mise en œuvre du tirage des unités primaires (DR)

Les tirages des unités primaires (DR) sont réalisés de façon indépendante d'un domaine d'étude à l'autre. Comme cela est indiqué plus haut, les unités primaires (ou aires 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  $\pi_r$  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

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

Pour un tirage SPPT, la probabilité d'inclusion  $\pi_r$  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

$$\sum_{r=1}^M \pi_r = m$$

On en déduit

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

D'où

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

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

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

Si pour une unité  $r$ ,  $\pi_r \geq 1$ , 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.

## 5.2 Mise en œuvre du tirage des unités secondaires (ménages)

Le tirage systématique à probabilité égale sera utilisé et effectué sur le terrain après le redénombrement des DR échantillons. Les étapes suivantes seront suivies :

**Etape 1 : Numéroté** tous les ménages recensés dans le DR échantillon de 1 à  $N$  ( $N$  étant le nombre total de ménages recensés dans le DR).

**Etape 2 : Calculer** le « pas » de sondage  $k = N / n$ ,  $N$  étant le nombre total de ménages dans le DR échantillon et  $n=20$ , le nombre de ménages à tirer dans le DR.

**Etape 3 : Tirer** un nombre aléatoire (départ aléatoire)  $d$  entier compris entre 1 et  $k$ .

**Etape 4 : L'échantillon des ménages du DR** concernée est formé des ménages identifiés par les numéros les plus proches de  $d, d + k, d + 2.k, \dots, d + (n-1)k$ .

**NB :** *En choisissant les ménages de l'échantillon, il est important de garder les chiffres après la virgule dans l'intervalle d'échantillonnage. La règle est la suivante : quand la partie décimale du nombre de sélection de l'échantillon est inférieure à 0,5, le ménage du numéro inférieur est choisi et, quand la partie décimale du nombre de sélection de l'échantillon est supérieure ou égale à 0,5, c'est le ménage du numéro supérieur qui est choisi.*

## 6. Estimateurs

Les estimateurs généralement utilisés sont :

- Total
- Ratio
- Variance

## 7. Pondération et estimateur du total, moyenne, ratio

Notations

Le domaine d'étude est constitué par une Moughataa.

$M_h$  = Nombre de DR dans le domaine d'étude  $h$

$N_{ih}$  = Nombre de ménages u DR  $i$  du domaine d'étude  $h$

$N'_{ih}$  = Nombre de ménages du DR  $i$  du domaine d'étude  $h$  après la réactualisation du DR

$m_h$  = Nombre de DR échantillons du domaine d'étude  $h$

$n_{ih}$  = Nombre de ménages échantillons du DR  $i$  du domaine d'étude  $h$

Les poids compensent les probabilités inégales de sélection. La méthode pour corriger ces probabilités inégales consiste à appliquer des poids d'échantillonnage aux données de l'enquête pendant l'analyse en multipliant la valeur de l'indicateur par le poids. Le poids convenant pour chaque unité de l'échantillon est simplement l'inverse de la probabilité de sélection de cette unité :

$$w_{jih} = \frac{1}{\pi_{jih}} \quad (1)$$

$w_{jih}$  = poids de sondage pour le ménage  $j$  du DR  $i$  du domaine d'étude  $h$  ; et

$\pi_{jih}$  = probabilité de sélection pour le ménage  $j$  du DR  $i$  du domaine d'étude  $h$

La probabilité de sélection  $\pi_{jih}$  est égale aux produits des probabilités  $p_{1h}$ <sup>3</sup> (probabilité de sélection des DR) et  $p_{2jih}$  (probabilité de sélection des ménages dans les DR) données par :

$$p_{1h} = m_h \frac{N_{ih}}{\sum_{i=1}^{M_h} N_{ih}} \quad (2)$$

La probabilité  $p_{2jih}$  est donné par :

$$p_{2jih} = \frac{n_{ih}}{N'_{ih}} \quad (3)$$

**NB :** Dans la pratique  $p_{1h} = 1$  si  $m_h N_{ih} \geq \sum_{i=1}^{M_h} N_{ih}$  et on recalcule les probabilités d'inclusion de la même manière sur les individus non sélectionnés d'office.

La probabilité totale de sélection du ménage  $j$  est :

---

<sup>3</sup> En toute rigueur, il ne s'agit pas de probabilité de sélection mais de taux de sondage.

$$\pi_{jih} = p_{1h} \times p_{2jih} \quad (4)$$

Soit :

$$\pi_{jih} = m_h \frac{N_{ih}}{\sum_{i=1}^{M_h} N_{ih}} \times \frac{n_{ih}}{N'_{ih}} \quad (5)$$

Le coefficient de pondération  $w_{jih}$  pour l'estimation du total Y d'une variable d'étude dépend de chaque UP échantillonnée et a pour expression :

$$w_{jih} = \frac{\sum_{i=1}^{M_h} N_{ih}}{m_h N_{ih}} \times \frac{N'_{ih}}{n_{ih}} \quad (6)$$

**Remarque :** Dans chaque DR on a tiré  $n_{ih}$  ménages. Après l'enquête, si pour un DR donné, au lieu de  $n_{ih}$  ménages, on a effectivement enquêté  $n'_{ih}$ , il y a lieu de faire une repondération ; le poids de sondage devient :

$$w'_{jih} = \frac{n_{ih}}{n'_{ih}} w_{jih} \quad (7)$$

En prenant en compte également le coefficient de redressement  $CR_h$  (voir tableau 2), on obtient finalement le poids :

$$w''_{jih} = CR_h w'_{jih} = CR_h \frac{n_{ih}}{n'_{ih}} w_{jih} \quad (8)$$

**Estimateur du total dans un domaine d'étude h :**

$$\hat{Y}_h = CR_h \frac{1}{p_{1h}} \sum_{i=1}^{m_h} \frac{N'_{ih}}{n_{ih}} \sum_{j=1}^{n_{ih}} y_{ijh} = \sum_{i=1}^{m_h} \sum_{j=1}^{n_{ih}} CR_h w'_{jih} y_{ijh} \quad (8)$$

Donc tous les ménages du domaine d'étude h du DR i ont pour poids de sondage donné par (7).

**Estimateur de la moyenne dans le domaine d'étude h:**

$$\hat{\bar{Y}}_h = \frac{\hat{Y}_h}{M_h} \quad (9)$$

**Autres résultats**

L'estimation d'un ratio, de l'effectif ou de la proportion des individus d'un sous-ensemble de l'univers sondé se déduit également de tout ce qui précède, l'effectif et la proportion étant des cas particuliers de total et de moyenne respectivement.

Pour l'ensemble du pays, c'est-à-dire pour les 53 domaines d'étude, le total estimé de la variable  $y$  est la somme des totaux estimés pour les différents domaines d'étude. De même, parce que les tirages sont réalisés de façon indépendante d'un domaine d'étude à l'autre, la variance du total estimé pour l'ensemble du pays est la somme des variances estimées des différents totaux estimés des 53 domaines d'étude.

La qualité de l'estimation obtenue pour le total s'apprécie à travers sa variance estimée. L'intervalle de confiance du total estimé ainsi que son coefficient de variation sont déterminés à l'aide de la variance estimée du total. L'intervalle de confiance et le coefficient de variation permettent de qualifier le total estimé.

## 8. Erreurs de sondage

Les formules classiques standard utilisées pour les échantillons aléatoires simples sont inadaptées pour l'estimation des variances des erreurs type du présent plan de sondage car elles sous-estimeraient les variances à des degrés divers.

On peut donner un estimateur de la variance vraie de l'estimateur du total  $Y$  en assimilant le présent plan de sondage à un sondage aléatoire simple avec remise<sup>4</sup> :

$$\hat{V}(\hat{Y}) = \frac{1}{m(m-1)} \sum_{i=1}^m \left( \frac{\hat{Y}_i}{\pi_i} - \hat{Y} \right)^2 \quad (10)$$

On peut également utiliser des méthodes d'estimation de la variance qui rendent compte du plan de sondage utilisé. La méthode Jackknife de calcul de la variance et la méthode Taylor de linéarisation peuvent être utilisées.

### **Méthode Jackknife**

L'estimation de la variance en utilisant la méthode Jackknife nécessite de constituer des sous-échantillons de l'échantillon complet en éliminant, chaque fois, de façon aléatoire une grappe (un DR) d'un domaine ou d'une strate. Une pseudo-estimation est alors calculée à partir des DR conservés, qui sont pondérés pour compenser l'unité éliminée. Par conséquent, pour une strate donnée contenant  $k$  grappes,  $k$  estimations dédoublées sont calculées en éliminant, chaque fois, une grappe et en augmentant le poids des  $(k - 1)$  grappes restantes par un facteur  $k / (k - 1)$ . Cette procédure est répétée pour chaque grappe.

Pour une strate ou un domaine donné, l'estimation de la variance d'un taux  $r$  est calculée de la façon suivante :

$$\text{var}(r) = \frac{1}{k(k-1)} \sum_{i=1}^k (r_i - r)^2 \quad (11)$$

---

<sup>4</sup> P. Ardilly, les techniques de sondage

Où

$k$  est le nombre de grappes dans la strate ou le domaine d'études,

$r$  est l'estimation pondérée calculée pour l'échantillon complet des grappes de la strate,

$r_i = kr - (k-1)r_{ri}$ , où  $r_{ri}$  est l'estimation répondérée, calculée pour le sous-échantillon de  $k-1$  grappes.

Pour obtenir une estimation de la variance à un niveau plus élevé, disons au niveau national, la procédure est répétée pour toutes les strates, avec  $k$  redéfini pour correspondre au nombre total de grappes (par opposition au nombre de grappes dans la strate).

### **Méthode de linéarisation de Taylor**

La méthode Taylor de linéarisation, organise les unités de sondage au premier degré (DR ou grappes) en strates implicites contenant, au moins, deux de ces unités. Pour obtenir des strates implicites homogènes, on apparie les unités adjacentes dans l'ordre de sélection.

Pour un taux combiné,  $r = y/x$ , de deux ensembles de l'échantillon  $y$  et  $x$ , où  $y$  est l'ensemble pondéré de l'échantillon pour la variable  $y$  (par exemple, la dépense dans une Moughataa donnée) et  $x$  l'ensemble pondéré de l'échantillon pour le sous-groupe formant le dénominateur de l'indicateur (par exemple, l'ensemble des ménages de ladite Moughataa), la variance de  $r$  est estimée par :

$$\text{var}(r) = \frac{1-f}{x^2} \sum_{h=1}^H \left[ \frac{m_h}{m_h-1} \left( \sum_{i=1}^{m_h} z_{hi}^2 - \frac{z_h^2}{m_h} \right) \right] \quad (12)$$

où ,

$f$  est le taux global de sondage, généralement négligé à moins qu'il ne soit supérieur à 0,05,

$x$  est l'ensemble pondéré de l'échantillon pour le nombre de cas dans le sous-groupe,

$m_h$  est le nombre de grappes (DR) dans la strate  $h$ ,

$H$  est le nombre total de strates,

$z_{hi} = y_{hi} - rx_{hi}$  , où  $y_{hi}$  et  $x_{hi}$  sont, respectivement, les sommes pondérées des variables  $y$  et  $x$ , dans le DR  $i$  de la strate  $h$