

DEPARTEMENT AMENAGEMENT  
LICENCE AMENAGEMENT S4  
Matière les techniques d'enquête  
**COURS N3 L'ECHANTILLONNAGE**

Une enquête systématique sur une population entière est très coûteuse en temps et en moyen. Le recours à un échantillon représentatif est la seule solution pour approcher la réalité de cette même population.

Avant de parler de l'échantillonnage proprement dit il est intéressant ; de définir la population objet de l'enquête.

### **1 / LA POPULATION**

La population parente, population mère est : l'ensemble des éléments sur lequel porte l'étude. La population est définie par les caractéristiques des individus qui intéressent l'enquête.

**Exemple1** : Etude sur les besoins en matière de crèche.

Pour faire partie de la population d'une enquête. Il faut être mère de famille résidant dans telle ou telle commune, ayant au moins un enfant de moins de deux ans dans le mois de l'enquête

Il est important que l'on puisse déterminer dans une étude si un individu fait partie de la population d'étude ou pas.

### **2 / ECHANTILLON**

L'échantillon ou le sondage, c'est le fait de choisir une partie, pour représenter la population parente (mère). L'échantillon sert à extrapoler les résultats sur l'ensemble de la population. A condition que le sondage reflète réellement la population mère.

L'échantillonnage est une opération délicate qu'il faut prendre avec beaucoup de prudence et de rigueur méthodologique.

L'échantillonnage non probabiliste est un moyen de sélectionner des unités d'une population à l'aide d'une méthode subjective (c.-à-d. non aléatoire). Il n'est pas nécessaire d'avoir une base de sondage complète pour l'échantillonnage non probabiliste qui est donc un moyen rapide, facile et bon marché d'obtenir des données

Par exemple, il est courant que l'intervieweur décide subjectivement qui doit être échantillonné. Étant donné que l'intervieweur sélectionnera probablement les membres de la population les plus amicaux ou faciles d'accès, une partie importante de la population n'aura aucune chance d'être sélectionnée

<b>Échantillonnage non probabiliste</b>	
<b>Avantages</b>	<b>inconvénients</b>
Rapide et pratique. Il suffit de sortir terrain et poser des questions à la première dizaine de personnes qui se présentent. Bon marché quelques heures seulement suffisent pour ce type d'enquête Les échantillons ne sont pas géographiquement dispersés donc pas de frais de déplacement. Ne nécessite pas une base de sondage Très utile pour les études de recherche et d'élaboration d'enquête	les échantillons non probabilistes comportent un biais de sélection Il faut donc avoir des hypothèses solides sur la représentativité de l'échantillon pour formuler des inférences sur la population. Il est impossible de faire des estimations fiables et des estimations de l'erreur d'échantillonnage

## 2-1/ les risques d'un échantillon

### 2-1-1/ généralisation abusive

Généraliser à partir d'un échantillon très limité risque de conduire à des conclusions hasardeuses

### 2-1-2/ un échantillon non représentatif

Le nombre ou la taille de l'échantillon n'est pas la seule bonne règle à observer dans un échantillon. Mais plutôt la règle de choix ou la méthode de choix de l'échantillon est aussi importante.

- La méthode de choix de l'échantillon doit assurer à toute unité de la population de figurer dans l'échantillon.

Exemple réaliser une enquête auprès des jeunes de 18 à 24 ans d'une ville universitaire, l'échantillon à prendre ne doit pas se faire uniquement sur le campus universitaire sous prétexte que c'est là qu'ils sont les plus nombreux. Les jeunes de 18 à 24 ans figurent aussi dans la population de la ville.

- La taille de l'échantillon est plus ou moins précise selon sa taille.

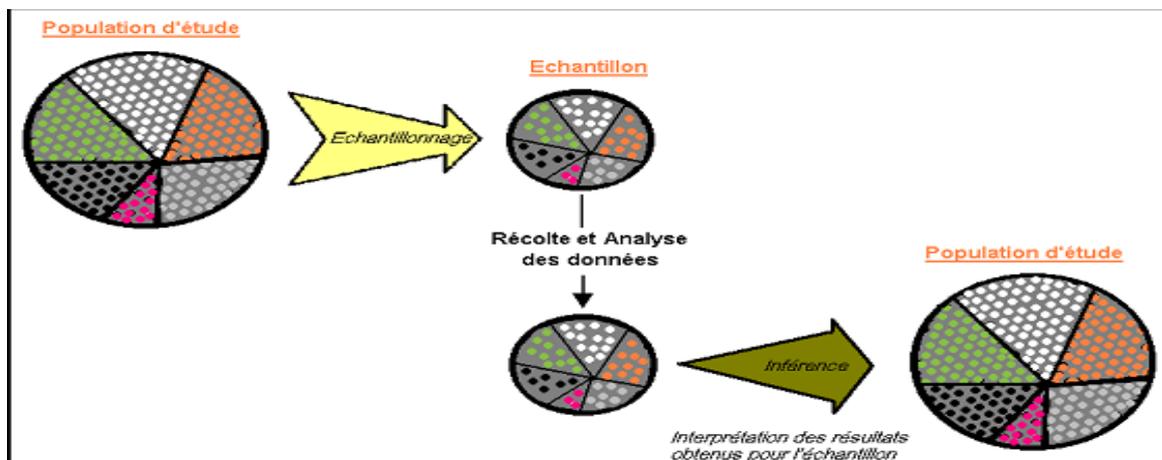


Fig. n°1 représentation schématique d'un échantillon

## 3 / METHODOLOGIE DE L'ECHANTILLONNAGE

Avant de procéder à l'échantillonnage il y a lieu de préciser :

### 3-1 / OBJECTIF DE L'ENQUETE

**Justification**, exemple / Déterminer la population nécessitant une aide

- Avant l'aide : déterminer les caractéristiques des populations vulnérables.  
Répondre aux questions « où ? », « qui ? », « combien ? ».

- Objectif et modalité de l'enquête

**Décrire, vérifier, comprendre et projeter :**

Avoir recours à des données issues d'une enquête de terrain ne signifie pas s'intéresser à l'échantillon lui-même, mais à ce qu'il est possible d'apprendre à partir de l'enquête et à la façon dont on peut appliquer cette information à l'ensemble de la population.

### 3-2 / Modalités d'enquête

Il existe deux types d'enquêtes, les enquêtes rapides et les enquêtes auprès d'un échantillon représentatif

**3-2-1 / Enquête rapide** Temps disponible : 24h à 3 semaines

Interlocuteurs : groupes de personnes, informateurs clés, échantillon arbitraires.

Informations récoltées : qualitatives

**3-2-2 / Enquête par sondage** Temps disponible : > 3 semaines

Interlocuteurs : population d'étude ou échantillon représentatif de celle-ci

Informations récoltées : quantitatives et qualitatives

Les enquêtes par sondage sont nécessaires lorsque nous voulons analyser plus finement les comportements d'une population face à un événement.

### 3-2 / DEFINITION DE LA POPULATION

Une population peut être générale ou très restreinte, passant d'une définition large (ex : tous les pêcheurs d'un pays ; l'ensemble des fermes de la région étudiée ; la population totale d'un camp de réfugiés) à une définition parfois très précise (ex : les chômeurs d'une région particulière enclavée ; les paysans d'une région dont le foyer est composé de plus de 6 personnes, et dont le revenu total est inférieur à 8000 da ; les enfants nécessiteux dans les écoles primaires).

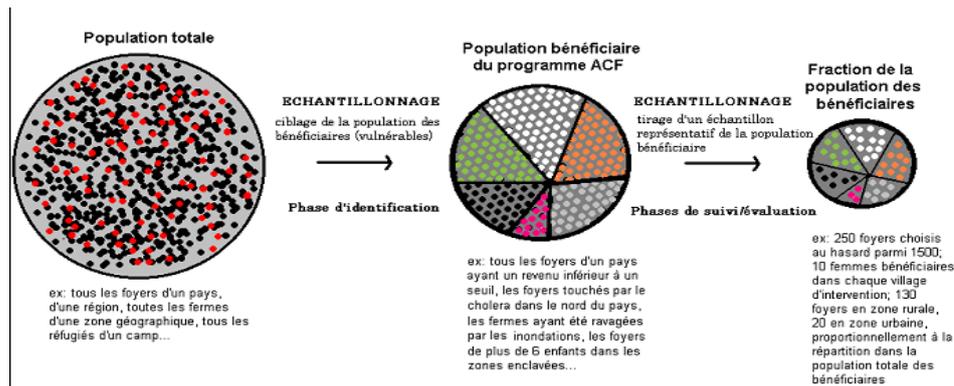


Fig. n°2 Définition de la population et échantillon

Ex : la population d'étude.

Dans certains cas, il s'agit d'individus particuliers, comme les enfants ou les femmes.

Ex : les enfants de moins de 5 ans d'une population atteinte de malnutrition ; le projet veut injecter de l'argent dans une société pour promouvoir les activités non-agricoles génératrices de revenu (la population cible sera les femmes en charge des activités de petit commerce).

#### 4 / Echantillonnage

##### 4-1 / Taille de la population

Il est important dans le processus d'échantillonnage de connaître la taille de la population d'origine car cette grandeur permet de calculer la taille et le degré de précision de l'échantillon désiré et effectuer correctement l'interprétation des résultats lorsque l'on applique à la population totale les conclusions faites sur l'échantillon d'étude (phase d'inférence).

La population d'étudiant UBMA 15000 étudiants.

échantillons total  $E_t = Np/(1+Np)xe^2$

$$n = \frac{N}{1 + N \times e^2}$$

Où

N = taille de la population

E = niveau de précision

L'arbre de décision, se présente comme suit, il se trace à l'aide des variables d'intérêt retenues.

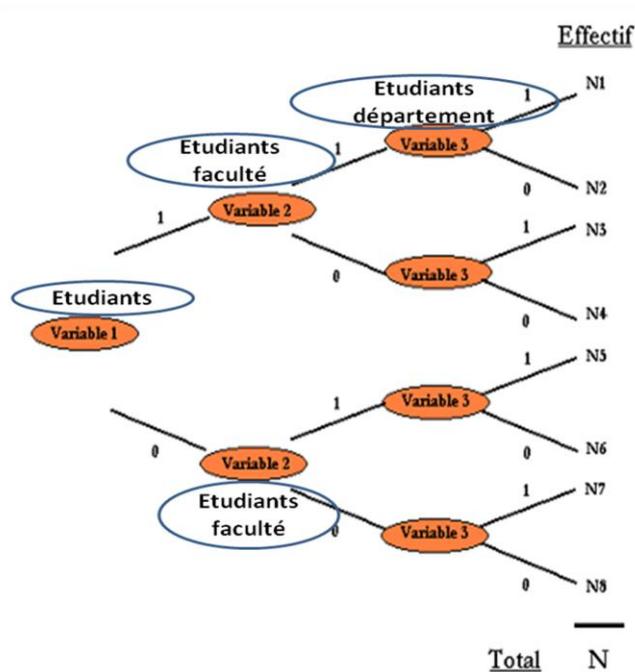


Fig. N°3 arbre de décision stratification d'une population en vue de réaliser un échantillon

Cette étape est utile pour récolter les effectifs associés selon les critères discriminants retenus, dans la mesure du possible.

Chaque variable discriminante est binaire ; Si un seuil de convergence a été retenu, alors elle prend la valeur 0, lorsqu'elle est en dessous du seuil retenu et 1 lorsque la valeur est au delà de la variable seuil (ex : le seuil du revenu minimum est de 18000 da/personne, **pour l'arbre de décision associé** : Revenu=0 si le revenu < 18000/personne &-Revenu=1 si le revenu > 18000/personne).

N.B. : L'absence d'effectifs servira pour argumenter le choix d'une méthode d'échantillonnage non probabiliste (sans liste quantifiable d'unités le tirage aléatoire est impossible).

TYPE DE FERME		CRITERE 1		CRITERE2		CRITERE3		STRATE	POIDS STRATE	NEW STRATE	POIDS STRATE			
F1	1640	1	1410	1	1181	1	984	1	24,60	1	24,60			
						0	197	2	4,93	2	4,93			
				0	229	1	0	3	0,00	3	5,725			
						0	229	3	5,73					
	0	230	1	197	1	33	4	0,83	4	5,75				
					0	164	5	4,10						
			0	33	1	0	6	0,00						
					0	33	6	0,83						
F12	1240	1	682	1	558	1	434	7	10,85	5	17,05			
						0	124	8	3,10					
				0	124	1	0	9	0,00					
						0	124	9	3,10					
		0	558	1	372	10	3,10	6	9,30					
				0	186	11	6,20	7	4,65					
	0	186	1	0	12	0,00	7	4,65						
			0	186	12	4,65								
			1	896	1	336			1	112	13	2,80	8	8,40
									0	224	14	5,60		
			0	560	1	0			15	0,00	9	14,00		
					0	560			15	14,00				
0	224	1	67	1	11	16	0,28	10	5,60					
				0	56	17	1,40							
		0	157	1	0	18	0,00							
				0	157	18	3,93							

4000

4000

4000

4000

100

Tableau n°1 calculé d'un échantillon dans une population stratifiée

Quand les strates ont un poids de la strate inférieur à 5% pour être représentée et si l'intérêt de l'étude ne porte pas spécifiquement sur ces strates particulières, il convient alors de les regrouper pour former des sous-groupes moins détaillées. Une strate trop petite (<5%) ne peut faire l'objet d'interprétations à grande échelle statistiquement correcte. Plus l'échantillon est petit, moins le nombre de strates doit être élevé.

En règle générale, il vaut mieux se restreindre à des échantillons comportant au plus 10 strates différentes.

Comment obtenir un échantillon représentatif :

#### 4-2 / La Taille de l'échantillon

L'enquête, menée sur un échantillon, doit vous permettre de généraliser les résultats à l'ensemble de la population ciblée. La taille de l'échantillon est bien sûr déterminante : Plus votre échantillon est important, plus la généralisation est fiable. Toutefois, les gains de fiabilité ne sont pas proportionnels à l'augmentation de la taille de l'échantillon. Ainsi, en interrogeant 2.000 personnes vous n'aurez pas des résultats deux fois plus fiables qu'en n'en interrogeant que 1.000.

La notion de fiabilité d'échantillon est matérialisée par un seuil de confiance et une marge d'erreur. Ainsi, un échantillon défini à un seuil de confiance de 95% et avec une marge d'erreur de 3% vous permettra d'extrapoler chaque résultat issu de votre enquête, avec 5% de risques de vous tromper de + ou - 3%. Ainsi, si vous obtenez un taux de satisfaction de 67% sur un produit ou service, vous pourrez affirmer que le taux de satisfaction

réel sur l'ensemble de vos clients a 95% de chances de se situer entre 64 (67-3) et 70 (67+3) %.

#### 4-3 / Niveau de précision

Le niveau de précision, encore appelé erreur d'échantillonnage, estime l'intervalle de confiance dans lequel on va situer la valeur réelle de la population. La valeur prise par la population sera comprise en deçà et au-delà de la valeur estimée pour l'échantillon, selon le niveau de précision voulu. Ce dernier est exprimé en points de pourcentage (ex : +/-5%). Si la valeur estimée est un pourcentage alors la valeur réelle est comprise entre « la valeur estimée- le niveau de précision » et « la valeur estimée + le niveau de précision ». (ex : si le gain de l'Indice de Poids/Taille des enfants malnutris de l'échantillon est de 20% avec un degré de précision de +/- 5%, alors on peut dire que ce gain se situe entre 15% et 25% pour la population totale des enfants malnutris). Si la valeur estimée est un nombre, la largeur de l'intervalle se calcule en multipliant la valeur estimée par le niveau de précision adopté ; la valeur réelle de la population est alors [ «valeur estimée – largeur de l'intervalle »; « valeur estimée + largeur de l'intervalle »]. (ex : la moyenne des revenus pour la population des fermiers enquêtés est de 85 UM. La moyenne pour la population totale se situera dans l'intervalle suivant  $[85-(85 \times 5\%) ; 85+(85 \times 5\%)] = [81 ; 89]$  avec un degré de précision de +/-5%). Plus le degré de précision est élevé, plus l'intervalle sera étendu (ex : niveau de précision : +/-6,6% -> l'intervalle de confiance sera [79 ;91]) :

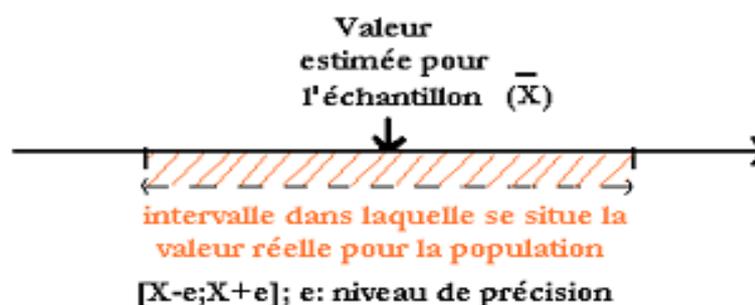


Fig. N3 schéma représentant le niveau de précision

L'erreur d'échantillonnage d'un estimateur est mesurée par sa variance d'échantillonnage déterminée comme fluctuation de sa moyenne calculée en tenant compte de tous les échantillons possibles tirés du plan d'échantillonnage. Un estimateur ayant une variance d'échantillonnage minimale est considéré *précis*. La précision augmente quand la variance d'échantillonnage diminue. Il faut noter qu'un estimateur peut être précis et biaisé. L'*exactitude* tient compte à la fois de la variance et du biais ; un estimateur exact jouit d'une bonne précision et est peu entaché de biais.

#### 4-4 / Niveau de confiance

Le niveau de confiance (ou marge d'erreur) permet d'indiquer le pourcentage de chances que l'échantillon sélectionné est représentatif de la population étudiée.

Imaginons que l'on puisse répéter la mesure d'une valeur (ex : moyenne du revenu des foyers) auprès d'un grand nombre d'échantillons de la population étudiée. La moyenne des valeurs obtenues (ex : 8400; 8500; 8400; 8600;... ; 8700; 8300; 8400; 8400;... ; 8300; 8100; 8700;... ; 8600; 8200; 7900 ;... ; 8500; 8500; 8600; 8500; 8900; 8400;

moyenne=8500da) sera égale à la moyenne réelle de la population étudiée (ex : revenu moyen=8500da).

Les valeurs obtenues pour ces échantillons suivent une distribution normale autour de la moyenne réelle. Certaines sont proches de la valeur réelle (ex : 8400; 8500; 8600), d'autres sont plus éloignées (ex : 7900; 8100; 8900...). La déviation standard (ou écart type  $\sigma$ , noté  $\sigma$ ) mesure la largeur de la distribution (dispersion des valeurs obtenues autour de la moyenne).

Dans le cas d'une distribution normale, la théorie montre que 95% des valeurs obtenues gravitent autour de la valeur réelle de la population avec une différence de moins de deux écart-types. En d'autres termes, un niveau de confiance de 95% assure que, parmi 100 échantillons tirés aléatoirement, 95 donnent une valeur estimée égale à la valeur réelle de la population totale (selon un certain niveau de précision). Au delà (dans 5% des cas) les valeurs dépassent la moyenne de plus de deux écart-types. Elles sont considérées comme trop éloignées de la moyenne réelle et les échantillons correspondants ne sont pas représentatifs de la population d'étude (ex : on retrouve ces 5% de cas extrêmes dans les « queues » (zone orange) de la cloche du graphique ci-dessus).

L'écart type n'étant pas connu lorsque l'on tire l'échantillon, il est d'usage de retenir un niveau de confiance égal à 95%, caractérisant toute distribution normale. La probabilité (**t**) associée, permet de déterminer l'écart correspondant pour la distribution de la loi de Student (proche de la loi normale lorsque le nombre d'individus dans l'échantillon est >200). Ainsi, lorsque le niveau de confiance est de 95% et la taille de l'échantillon assez grande, **t=1.96**. C'est cette valeur qui sera utilisée dans les formules permettant de calculer la taille optimum de l'échantillon.

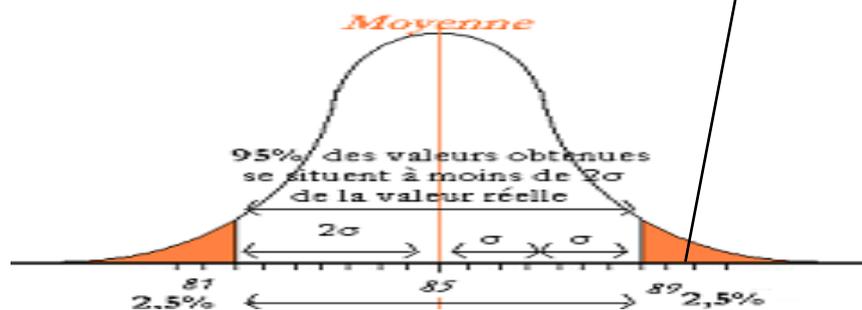


Fig. N4 schéma représentant le niveau de confiance

#### 4-5 / Degré de variabilité

Ce critère **détermine la ressemblance (degré d'homogénéité) des individus de la population** selon leurs caractéristiques communes. Moins les individus d'une population se ressemblent, plus l'échantillon doit être grand pour atteindre un même degré de précision. Inversement, plus la population est homogène, plus petit sera l'échantillon. **Une proportion de 50% indique une plus forte variabilité que 20% ou 80%.**

Cette proportion est suspectée, mais rarement quantifiable d'avance, il est donc d'usage d'utiliser la variabilité maximale (P=0.5) pour éviter les risques d'erreurs. Les deux notions étant intimement liées, on tiendra compte de cette variabilité au moment de décider du niveau de précision requis.

Ainsi lorsque l'on sait que les mesures recueillies seront très proches les unes des autres on veillera à retenir un degré de précision plus fin (ex : +/-3%), notamment lorsque notre étude vise à comparer deux types de populations très homogènes. En revanche, lorsque l'objectif de l'étude est simplement de décrire les comportements de populations que l'on sait à priori différenciées, le degré de précision sera moins fin; e= +/-14%).

#### 5/ Calcule d'un échantillon Comparaison entre deux formules

$$N = N / (1 + Np)xe^2 \quad (1) \quad N = ((38033 / ((1 + (38033) * (0.05 * 0.05)))) = 395.84$$

$$n = t^2 N / t^2 + (2e)^2 (N - 1) \quad (2)$$

$$= \underbrace{((1,96 * 1,96) * (38033))}_{t \quad N} / \underbrace{((1,96 * 1,96) + (2 * 0,05) * (2 * 0,05) * (38033 - 1))}_{t \quad e \quad N} = 350.04$$

**N** : Taille de la population-mère (le nombre de locaux du type d'activité)

**n** : Taille de l'échantillon représentatif de l'activité

**t** : Coefficient de marge **pour un niveau de confiance donné (pour notre cas, la valeur qui correspond au niveau de confiance de 95 % est  $t=1.96$ )**

**e** : Marge d'erreur (5% soit 0.05)

## **6 / Les types de tirage d'échantillon**

### **6-1 / Tirage probabiliste ou aléatoire**

Pour employer cette technique il faut disposer d'un fichier ou d'un répertoire exhaustif de la population objet de l'enquête. On tire au hasard dans la liste des personnes à interrogés.

Si la liste est longue, on procède à un tirage systématique d'une unité à intervalle régulier (toutes les 50, 100, selon le nombre de personnes à interroger) exemple : la taille de l'échantillon est de 400 sur une population totale de 5000, le taux de sondage est de  $10000/400 = 25$  source soit un sujet tous les 25 personnes après avoir choisi au hasard un point de départ entre 1 et 25.

**A** / Un échantillon systématique est équivalent à un échantillon au hasard simple si la première personne est tirée au hasard.

**B** / Dans certaines situations la population est stratifiée c .a. d, la population est classée en strate. Un tirage aléatoire est réalisé dans chaque strate.

Comment choisir les strates ? Pour établir des strates il faut :

Prendre en compte un caractère dont la valeur est connue avant l'enquête (projet d'avenir pour une population d'étudiant) on peut classer cette population selon les filières d'étude donnée connue par les fiches d'inscription.

Le critère de stratification soit en relation avec l'objet de l'étude.

**C** / Echantillon par groupe, dans ce cas de figure l'unité n'est pas un élément, mais un groupe d'élément qui ont entre eux un lien naturel (groupe d'écolier groupe d'ouvrier d'une même usine etc.). La base de sondage est à établir ainsi que la liste de ces groupes. Le tirage aléatoire sélectionne des grappes dans lesquelles tous les éléments seront interrogés.

Un cas particulier de sondage par grappe est le sondage aréolaire (ou sondage par zone). Les unités soumises au tirage aléatoire sont des aires géographiques. Il est nécessaire de fixer aux zones des limites claires et facilement repérables sur le terrain.

- **D** / Tirage à plusieurs degrés Il s'agit d'une succession de tirage par grappe, le hasard intervient dans la désignation de grappe de plus en plus petites, incluse dans celles du niveau précédent. Le tirage du dernier niveau ou (degrés) permet la désignation des éléments constituant l'échantillon exemple : un premier tirage (1 degrés) des îlots dans une ville ( sur la liste des îlots) ; les îlots tirés au sort constituent des grappes d'immeubles dont on peut établir la liste et parmi lesquelles un tirage au second degrés est effectué, les immeubles tirés au sort sont des grappes de ménages ( dont on peut encore dresser la liste) parmi lesquelles un échantillon aléatoire est sélectionné. Un dernier tirage au sort permet de désigner la personne à interroger dans chaque famille.

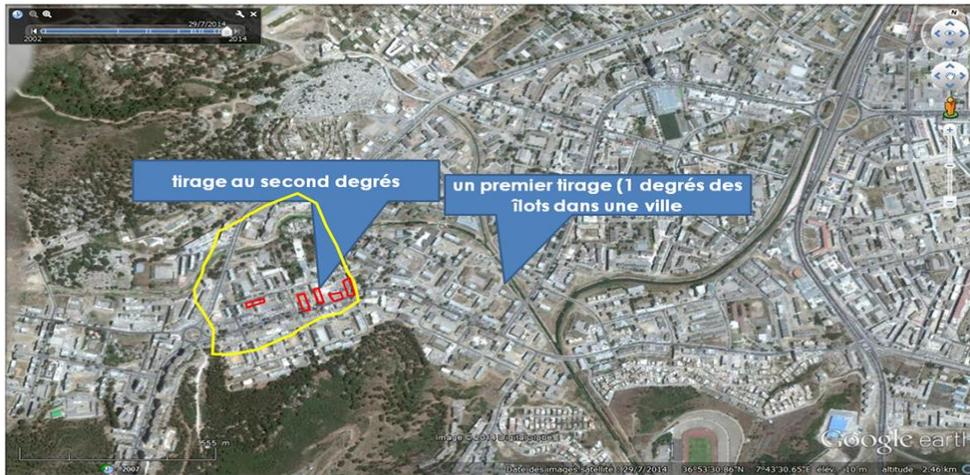


Fig. N5 schéma représentant le tirage par grappe

## 6-2 / LES ECHANTILLONS EMPIRIQUES

### A / échantillon par quotas ou échantillon proportionnel

L'objectif dans ce cas de figure est de construire un échantillon qui ressemble à la population mère, à partir de données statistiques sur la structure de cette population en quelque sorte un modèle réduit de cette population.

- Le principe de base de cette méthode est qu'un échantillon identique à la population concernant la répartition de certains caractères (variables contrôlées) sera peu différent concernant la répartition des autres variables (variables non contrôlées) les variables contrôlées sont par exemple le sexe, l'âge, la taille de la commune de résidence, la profession du chef de ménage, la profession du répondant, le niveau du répondant. Mais vous pouvez également intégrer des critères comportementaux ou d'opinion, dont vous connaissez la proportion dans la population mère.

Au delà de l'objectif de représentativité statistique de l'échantillon, les quotas peuvent également être utilisés comme critères de sélection des personnes à interroger en liaison avec l'objet de l'étude. Ainsi, vous pouvez chercher à interroger 50% de clients et 50% de non clients, non pas parce que c'est le reflet de la réalité mais simplement parce que vous souhaitez comparer les opinions de vos clients avec ceux des non-clients. Dans ce cas, vous êtes en présence de deux échantillons distincts dont la représentativité est à mesurer séparément.

Dans tous les cas, ne retenez qu'un nombre restreint de quotas. Au delà de 2 ou 3 quotas, vous compliquerez la tâche de vos enquêteurs, surtout pour la gestion des "queues de quotas".

De même optez pour des quotas simples où l'enquêteur cherchera à remplir séparément les différents critères plutôt que pour les quotas croisés, même si ces derniers traduisent plus fidèlement la structure de la population globale.

Taille de la population mère	Taille souhaitable de l'échantillon <sup>1</sup>
10 000 000	1 067
100 000	1 056
5 000	880
1 000	516
500	341
200	169
100	92

<sup>1</sup>. Nombre de personnes à interroger pour qu'il y ait 95% de chances que les caractéristiques réelles de la population mère ne soient pas différentes de plus ou moins 3% de celles observées dans l'échantillon interrogé

## 7 / La réalisation du questionnaire

- Augmenter le taux de réponses

Se présentez et expliquez brièvement l'objectif de enquête contribue à instaurer une confiance entre l'enquêteur et la personne à interroger. Cette confiance participe dans l'augmentation du taux de réponse.

- Choisir le lieu d'administration du questionnaire.

Le lieu d'administration du questionnaire est primordial lorsque l'on souhaite obtenir un échantillon représentatif. **Placez-vous dans un lieu qui a des chances de connaître une certaine mixité, en termes d'origine sociale, d'âge, de sexe, etc.** Cela allégera votre travail pour respecter la méthode des quotas. Interrogez les personnes en tête-à-tête plutôt que dans un espace public.