



Normes FAO – Etude sur la sécurité semencière

ECHANTIONNAGE ET TAILLE DE L'ECHANTILLON

L'échantillonnage est la sélection d'une partie représentative de la population afin de déterminer les paramètres ou les caractéristiques de l'ensemble de la population. Un **échantillon** est donc la proportion de la population sélectionnée pour l'étude ou l'enquête. Une **unité d'échantillonnage** est l'unité dont l'information est recueillie et qui constitue la base de l'analyse. Dans la recherche par enquête, des éléments sont des personnes ou certains types de personnes. La **méthodologie** ou le plan **d'échantillonnage** fait référence à un ensemble de règles ou procédures précisant comment un échantillon doit être sélectionné. Cela peut se faire de façon aléatoire ou non-aléatoire. Dans l'échantillonnage aléatoire, il est possible d'attribuer des niveaux de confiance quant à la fiabilité de l'échantillon d'un point de vue statistique. Dans l'échantillonnage non aléatoire, ceci n'est pas possible. Dans les études sur la sécurité semencière, les ménages sont sélectionnés par échantillonnage aléatoire, alors que toutes les autres unités d'échantillonnage (individus pour les GD communautaires, commerçants, fonctionnaires, représentants d'ONG, producteurs de semences) ne le sont pas. Dans le cas des unités, la représentativité repose sur un échantillonnage raisonné.

Dans l'échantillonnage aléatoire, l'échantillonnage est souvent effectué de manière à assurer 95% de niveau de confiance dans les estimations tirées de l'échantillon. Le nombre d'unités nécessaires pour atteindre cela augmente à mesure que l'hétérogénéité au sein de l'ensemble de la population augmente. Ainsi, l'échantillon nécessaire pour 95% de confiance dans une population qui est constituée de résidents, des personnes déplacées et de réfugiés sera probablement plus grand que pour une population uniquement constituée de résidents. Dans les situations où il y a hétérogénéité, des échantillons distincts doivent être évalués pour chaque groupe de population, en augmentant ainsi la taille de l'échantillon global.

En règle générale, dans une population de groupes relativement homogène de 10,000 unités ou plus qui sont normalement distribués, les tailles des échantillons nécessaires pour atteindre différents degrés de précision sont données dans le tableau.

Table 1: Degré de fiabilité et taille d'échantillonnage pour une population de plus de 10,000 ménages

Degré de fiabilité	75%	80%	85%	90%	95%
Taille de l'échantillon	133	165	210	270	384

Quand la taille de la population est plus petite que 10,000, il y aura une réduction de la taille de l'échantillon comme le montre le tableau 3.

Table 2: Degré de fiabilité et échantillonnage ajusté de la population

Taille de la population	Degré de fiabilité				
	75%	80%	85%	90%	95%
500	105	124	148	176	217
1,000	117	142	174	213	278
2,000	125	152	190	238	322
4,000	129	159	200	253	350
6,000	130	161	203	258	361
8,000	131	162	205	261	366
10,000	131	162	206	263	370
20,000	132	164	208	266	377
50,000	133	164	209	269	381
100,000	133	165	210	270	383
200,000	133	165	210	270	383
500,000	133	165	210	270	384

Enfin, la taille de l'échantillon final sera un compromis entre les meilleures pratiques théoriques obtenues en appliquant les formules mentionnées ci-dessus, et les ressources disponibles - humaines, financières, logistiques et de temps. Cet arbitrage va influencer le niveau final de la fiabilité à l'emplacement choisi. La décision finale sur la taille de l'échantillon prise pendant la phase préparatoire des enquêtes nécessite souvent que la personne responsable prenne une décision puisque il / elle doit faire un compromis entre ces deux composantes (taille de l'échantillon statistiquement optimale et ressources disponibles).

Une fois que la taille globale de l'échantillon est déterminée au niveau administratif (par exemple au niveau régional / provincial), elle peut être proportionnellement divisée pour les entités de la population du niveau administratif inférieur (par exemple les comtés ou districts). Si les entités géographiques au niveau administratif inférieur ont des tailles de population similaires, alors la taille globale de l'échantillon peut être divisée en parties égales entre les entités sélectionnées. Les différentes tailles d'échantillon auront des répercussions sur le temps nécessaire pour couvrir chaque district comme illustré dans l'encadré 5.

Encadré 1: Exemple sur la sélection de la taille de l'échantillon au niveau du Comté ou du District

- a) La taille de l'échantillon global au niveau du District est 384, et dans la région, trois Comtés ont été sélectionnés pour représenter plusieurs agro-écologies. La taille de chaque sous-échantillon est déterminée par la taille relative de la population des trois Comtés. Si la population des Comtés est à peu près égale, un échantillon de 128 ménages doit être interviewé dans chacun des trois Comtés.
- b) En moyenne, une équipe de cinq personnes sera capable d'interroger 20 ménages, 3-5 commerçants du marché local et 1-3 revendeurs agricoles. Donc, chaque équipe aura besoin d'un minimum de $(126/20)$ 7 de jour de travail pour couvrir chaque Comté.
- c) Si les trois Comtés ont par contre des tailles de population différentes, ex : X=9000; Y= 7000 and Z= 4000 (Total 20000), et la taille de l'échantillon pour chaque district sera:
 - $X = 9/20 \times 384 = 173$
 - $Y = 7/20 \times 384 = 134$
 - $Z = 4/20 \times 384 = 77$

Donc, une équipe de 5 partant de X aura besoin de 9 jours $(173/20)$, Y de 5 $(134/20)$ et Z de 4 $(77/20)$.

Méthode d'échantillonnage au niveau communautaire pour le Questionnaire de Ménage

Après évaluation de la taille, l'échantillonnage au sein du ménage se fait par l'échantillonnage aléatoire. L'échantillonnage aléatoire simple serait la procédure la mieux appropriée dans la mesure où chaque unité d'échantillonnage (ménage) a la même chance d'être sélectionnée sous cette méthode. Cependant, cela peut prendre beaucoup de temps particulièrement dans les régions où la population est dispersée sur une grande région géographique. Dans la plupart des cas, un échantillonnage aléatoire systématique) est préférable pour des raisons pratiques.

Encadré 2: Exemple sur un échantillonnage systématique

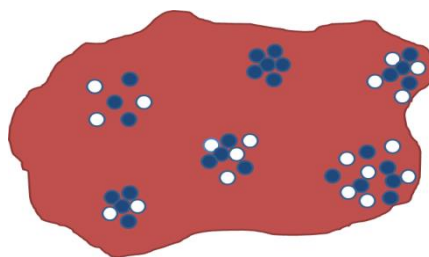
Dans un échantillonnage systématique, un schéma préfixé est normalement utilisé pour ignorer un certain nombre de ménages dans un intervalle. Par exemple, si on estime qu'il y a 20 ménages le long d'un transect donné, et l'agent enquêteur qui suit ce transect est supposé interviewer cinq ménages; il/elle peut déterminer l'échantillonnage de l'intervalle comme suit:

$$SI = \frac{\text{Nombres de ménages (ex: 20) le long du transect}}{\text{Nombre de ménages (5) à interviewer}} = 4$$

L'intervalle d'échantillonnage est donc égal à 4, et l'agent enquêteur peut donc ignorer 4 ménages après chaque échantillon. La distance que l'enquêteur le long du transect doit parcourir devrait être prise en compte. Par conséquent, un transect ne doit pas dépasser les 3 km.

La structure des peuplements et l'hétérogénéité des villages seront déterminées par le type d'échantillonnage systématique. On peut cerner la nature des peuplements au moment où l'équipe arrive dans le village ou à travers un questionnaire rapide administré aux résidents locaux. D'une façon générale, il existe trois types de peuplements :

- a) *Les ménages agglomérés* – cela consiste d'un groupe de 4 à 10 ménages agglomérés dans un espace restreint (diamètre de 20 à 10 m), et qui sont séparés par une distance raisonnable de 200-800 m les uns des autres. Ce schéma est très courant dans les communautés agro-pastorales. Dans ce cas, la méthode la plus appropriée est de sélectionner d'une façon aléatoire 4-8 agglomérations et dans chaque agglomération un échantillon aléatoire de 3 à 5 ménages.



- b) *Les peuplements linéaires* - Ici les ménages s'établissent le long des barrières naturelles ou modernes (faites par les hommes) tels que le pied des montagnes, le lit des rivières ou le long des routes principales. Cependant, la culture des champs de céréales peut être répandue à travers le village. Ici, l'échantillonnage systématique (Encadré 6) est normalement déployé le long du

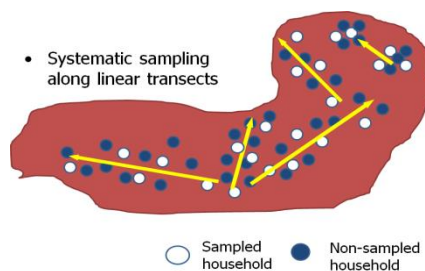
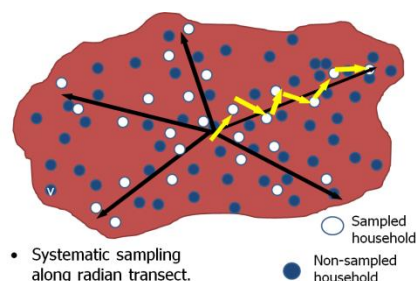


schéma linéaire du peuplement. Quand l'équipe arrive dans le village, ils doivent relever comment est-ce que la distance qui sépare les ménages les uns des autres influence le nombre qui peut être interviewé dans le village/localité. Plus la distance est grande entre deux ménages, plus de temps sera requis pour les agents enquêteurs pour se déplacer d'une habitation à l'autre.

- c) *Les ménages éparpillés* – Ici, il n'existe pas une structure reconnue de peuplements. Les ménages sont éparpillés aléatoirement à travers le village. Ici, comme dans les peuplements linéaires, l'équipe a besoin de relever la distance qui les sépare. La méthode la plus appropriée d'échantillonnage est l'échantillonnage systématique le long d'un diamètre de 3 à 6 transects en fonction du nombre d'agents. Chaque agent peut se déplacer le long d'un transect en relevant des échantillons à des intervalles réguliers.



N.B: Ces schémas d'échantillonnage sont supposés être homogènes. Si le village est divisé en différents groupes, par exemple, les personnes déplacées internes (PDI), celles rapatriées et déplacées, cela nécessitera alors un sous échantillonnage.