



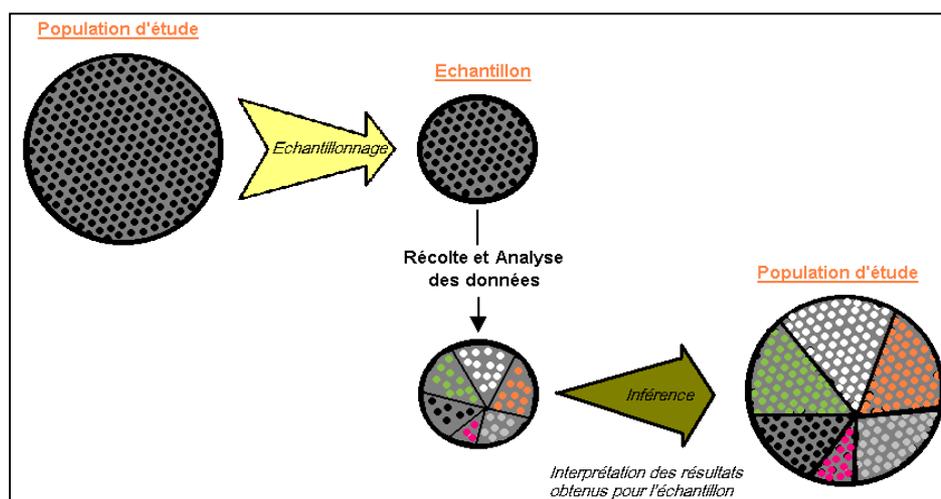
INTRODUCTION

Pour évaluer les programmes de Sécurité Alimentaire, et en premier lieu pour cibler les populations vulnérables auprès desquelles venir en aide, il convient de mener différentes enquêtes sur le terrain, auprès des populations, pour récolter des informations spécifiques (production, consommation, dépenses, ...) au niveau national, régional, local et familial.

Un recensement de l'ensemble de la population concernée est la plupart du temps irréalisable. Bien souvent la population est trop vaste pour être enquêtée dans sa totalité (ex : *camps de 100 000 réfugiés*), s'ajoutent aussi les moyens limités (ex : *nombre d'enquêteurs restreints, zones enclavées*), et le manque de temps (*le temps consacré aux enquêtes chez ACF varie entre 24h et 4 semaines*).

Dans de tels cas, les enquêtes de sondage permettent d'obtenir des informations sur une population en n'interrogeant qu'une fraction de celle-ci, l'échantillon. Plus celui-ci ressemble à la population d'étude, plus il est possible d'extrapoler à cette population les résultats obtenus pour l'échantillon.

Le graphique suivant illustre le principe employé pour ce faire :



Pour effectuer des enquêtes par sondage appréciées pour leurs données détaillées, leur précision et leur représentativité, il convient de sélectionner soigneusement l'échantillon d'étude. Cette étape est primordiale pour être en mesure d'interpréter les résultats de l'étude sur l'échantillon au niveau de la population totale. Il s'agit tout d'abord d'opter pour la technique d'échantillonnage la plus adaptée compte tenu des **objectifs majeurs de l'étude menée**, des **caractéristiques de la population d'étude** (taille, groupes différenciés), ainsi que des **contraintes rencontrées sur le terrain**. Au delà de cette notion de « représentativité », le principe de l'échantillonnage implique que **tous les individus ou « unités » de la population considérée doivent avoir, au mieux, la même probabilité de faire partie de l'échantillon choisi**, du moins, une probabilité connue.

Le présent module décrit précisément la méthodologie à adopter et propose des outils répondant à de nombreuses questions liées au thème de l'échantillonnage, tant au niveau pratique que technique.





METHODOLOGIE DE L'ECHANTILLONNAGE

1ère étape: OBJECTIFS DE L'ENQUETE

Justification, objectifs et modalités de l'enquête

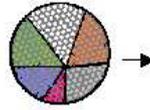
2ème étape: DEFINITION DE LA POPULATION

1. La population d'étude

Unité d'étude, taille de la population, la population ciblée

2. Critères d'échantillonnage

Variables d'intérêt, arbre de décision, strates retenues, tableau d'échantillonnage



| Typologie de population | Zone d'Economie Alimentaire | | |
|-------------------------|-----------------------------|----|----|
| | ET 2004-05 | 21 | 24 |
| 11000-000-000 | 20000 | | |
| 12000-000-000 | 20000 | | |
| 13000-000-000 | 20000 | | |
| 14000-000-000 | 20000 | | |
| 15000-000-000 | 20000 | | |
| Total | 100000 | | |

3ème étape: TAILLE DE L'ECHANTILLON

1. Critères statistiques

Niveau de précision - Niveau de confiance - Degré de variabilité

2. Calcul de la taille de l'échantillon

Petite population/études similaires, tables statistiques, formules mathématiques

3. Contraintes de terrain

Accès, temps, nombre d'enquêteurs, nombre de questionnaires remplis

4. Autres considérations

Enquêtes rapides, objectifs de l'enquête, analyse comparative, tirage en grappes

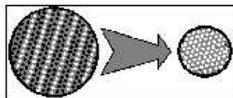
| | |
|------------------------------|------------|
| Nb. Enquêteurs | 4 |
| Nb. d'enquêtes/questionnaire | 6 |
| Nb. jours disponibles | 20 |
| Total | 400 |

$$n = \frac{Z^2 \cdot p \cdot q}{e^2}$$

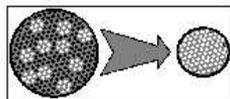
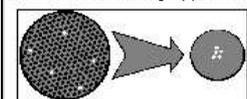
4ème étape: LA METHODE D'ECHANTILLONNAGE

NON ALEATOIRE

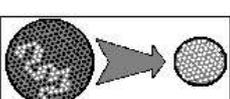
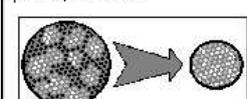
1. Echantillonnage Aléatoire Systématique



2. Echantillonnage Raisonné

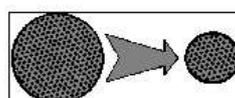


2. Echantillonnage Cumulatif



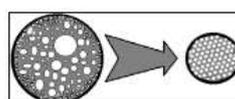
ALEATOIRE

1. Echantillonnage Aléatoire Simple



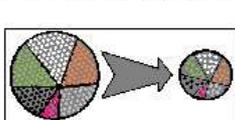
- TIRAGE**
1. Table de nombre aléatoire
 2. Technique des p'tits papiers
 3. Technique des pas
 4. Fonction ALEA (Excel)

2. Echantillonnage Pondéré



TIRAGE
technique

3. Echantillonnage Aléatoire Stratifié Proportionnel

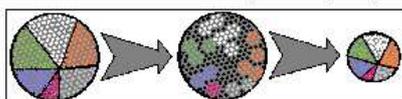


TIRAGE
Plan d'échantillonnage

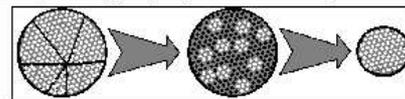
METHODES MIXTES

1. Echantillonnages par grappes

Echantillon Aléatoire Stratifié Proportionnel puis par Grappes

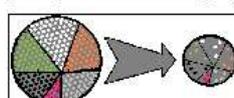


Echantillon par Grappes puis par Echantillonnage Aléatoire Simple



2. Echantillonnage Aléatoire Stratifié Non-Proportionnel

Sur-représentation d'un groupe



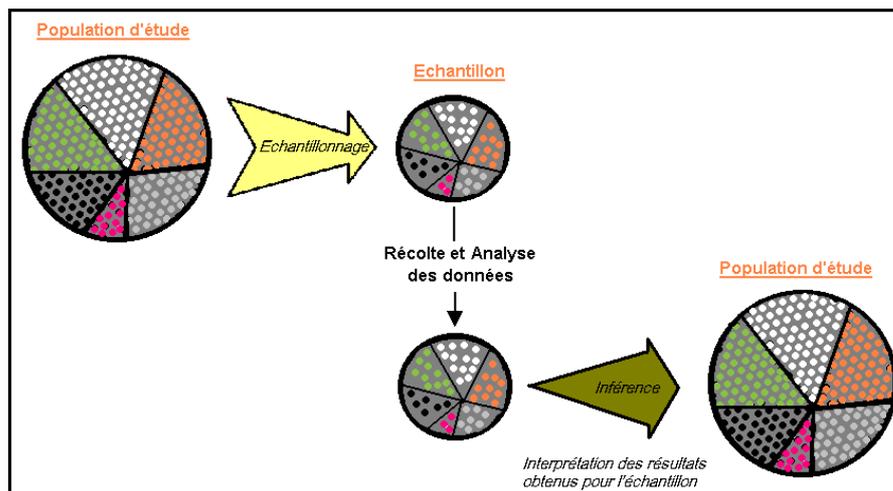


QU'EST-CE QU'UN ECHANTILLON REPRESENTATIF ?

La notion d'échantillon étant associée à la fiabilité des résultats obtenus, celui-ci doit posséder les mêmes caractéristiques que la population que l'on souhaite étudier, c'est-à-dire permettre d'**estimer avec une marge d'erreur acceptable les caractéristiques de la population** qui nous intéressent à partir des résultats de l'analyse de celles de l'échantillon. On parle alors d'« **échantillon représentatif** » de la population d'étude. Il va de soi que la **précision** de cette estimation nommée « inférence » dépendra de la **diversité et de la ressemblance** de la population composant l'échantillon proportionnellement à la population totale au niveau géographique, social, démographique, « agro-économique », etc...

Il conviendra donc de déterminer au préalable les caractéristiques¹ essentielles de la population, selon les intérêts de l'évaluation menée, pour s'assurer d'obtenir un échantillon dit « représentatif ».

La figure suivante schématise le principe pour l'obtention d'un échantillon représentatif. Sur le schéma introductif (p.2), l'échantillon était simplement une réduction de la population, tirée au hasard. Désormais, la population est scindée en amont en plusieurs sous-groupes, selon les caractéristiques intéressantes pour les besoins de l'étude, et l'échantillon respecte cette répartition de la population par sous-groupes.



Pour mieux comprendre l'intérêt de choisir un échantillon représentatif, penchons nous sur les deux exemples suivants. Le premier présente les trois principales techniques d'échantillonnage utilisées sur le terrain, le second se focalisera sur la plus-value opérationnelle de la technique dite de stratification.

Exemple 1 : Nous désirons déterminer la production moyenne de riz dans les fermes d'une région ravagée par une inondation, faisant l'objet d'une action de distribution de kit Seeds&Tools initiée par ACF, à partir d'un échantillon de 10 fermes. La production moyenne pour la population totale des 86 fermes, est de 0.75 tonnes par hectares (T/HA). Il est aussi connu que les fermes du Sud produisent plus de riz que celles du Nord.

❶ Dans un souci d'équité, sachant que les exploitations du sud produisent en général plus de riz que celles du nord, nous choisissons de construire un **échantillon contenant autant de fermes du Sud que de fermes du Nord**, soit 5 exploitations de chacune de ces régions au hasard:

| Production des fermes du Sud | Production des fermes du Nord |
|------------------------------|-------------------------------|
| 2,5 | 2,2 |
| 4,2 | 0,84 |
| 0,79 | 0,45 |
| 2,6 | 0,24 |
| 0,7 | 0,45 |





La production moyenne de cet échantillon de 10 fermes est de **1.5 T/HA**, soit 0.75 T/HA (i.e. +100%) de plus que la valeur exacte ! Cette technique intuitive ne semble donc pas être la plus appropriée.

Nous savons que l'ensemble des 86 fermes de l'étude est réparti géographiquement comme suit : 69 au Nord , 17 au Sud.

Sachant que les fermes du Sud produisent plus que celles du Nord, l'échantillon a été biaisé en faveur des fermes du Sud. En effet chaque ferme du Sud avait 5 chances sur 17 (soit une probabilité de $5/17=0.27$) de faire partie de l'échantillon, tandis que chaque ferme du Nord avait 5 chances sur 69 (soit une probabilité de $5/69=0.07$) d'être choisie. Il n'est donc pas surprenant d'obtenir une production moyenne trop élevée à partir de cet échantillon.

2 La technique appropriée pour que chaque ferme ait la même probabilité ($p=10/86=0.12$) d'appartenir à l'échantillon est de choisir 10 fermes au hasard, sans considération géographique (Nord/Sud).

Un tel tirage au hasard a donné les mesures de production suivantes (en T/HA):

Tirage 1 : 1.3, 2.9, 1.5, 2.2, 0.6, 0.36, 0.7, 0.26, 0.36, 0.98

La **moyenne pour cet échantillon est de 1.12 T/HA** ; elle est plus proche de la valeur réelle (+0.37) que pour le premier échantillon. Néanmoins, la moyenne aurait pu être totalement différente (moyenne plus élevée, moyenne plus proche de la valeur réelle, moyenne plus faible), si l'échantillon tiré avait été différent (cf trois tirages suivants) :

Tirage 2 : 1.3, 2.9, 1.5, 2.2, 0.6, 0.36, 0.7, 4.2, 0.36, 0.98 -> moyenne : 1.51

Tirage 3 : 0.36, 0.5, 1.5, 0.19, 2.3, 0.23, 0.5, 0.56, 0.27, 0.98 -> moyenne : 0.74

Tirage 4 : 0.36, 0.5, 1.5, 0.19, 0.3, 0.23, 0.5, 0.56, 0.27, 0.98 -> moyenne : 0.54

Les échantillons obtenus par tirage aléatoire simple étant très petits, le hasard peut donner des estimations très différentes les uns des autres et parfois très éloignées du résultats réel, même lorsqu'une seule valeur diffère. On parlera de variance élevée (cf chapitre sur l'échantillonnage aléatoire simple). Ce phénomène de variabilité est beaucoup moins important pour des échantillons de grande taille.

3 Une autre technique, donnant toujours à chaque ferme la même probabilité d'être choisie, est de **s'assurer que l'échantillon soit fidèle à la répartition en catégories observée dans la population totale**, relativement à la production de riz dans notre cas précis. La localisation géographique et la production de riz des fermes étant intimement liées (production plus élevée au Sud), nous allons effectuer un tirage aléatoire dans chacune des deux sous populations des fermes (Sud et Nord) comme suit :

- 20% des fermes sont au Sud : nous choisissons $20\%*10$, soit 2 fermes parmi les 17 fermes du Sud
- 80% des fermes sont au Nord : nous choisissons $80\%*10$, soit 8 fermes parmi les 69 fermes du Nord

Cette technique dite de stratification, respecte le principe que **chaque ferme a la même probabilité d'être choisie pour faire partie de l'échantillon** ($p=2/17=0.12$ au Sud et $p=8/69=0.12$ au Nord) :

| Production des fermes du Sud | Production des fermes du Nord |
|------------------------------|-------------------------------|
| 4,2 | 1,2 |
| 0,81 | 0,36 |
| | 0,6 |
| | 0,2 |
| | 0,15 |
| | 0,84 |
| | 0,3 |
| | 0,24 |

¹ La nature de ces caractéristiques sera définie dans la partie « 1^{ère} étape : POPULATION D'ETUDE »





La moyenne pour cet échantillon est de 0.89 T/HA ; c'est la valeur la plus proche de la valeur réelle (+0.14) comparativement aux deux premiers échantillons tirés auparavant. Cette amélioration s'explique

par le fait que la stratification **réduit le risque d'erreur** d'échantillonnage, i.e. elle empêche les combinaisons « extrêmes » qui peuvent être obtenues par un tirage aléatoire simple (trop de fermes du Sud ou trop de fermes du Nord qui biaisent le résultat lors du calcul de la moyenne, car les fermes du Sud produisent plus que celles du Nord).

Le second exemple qui suit illustre à nouveau le **bénéfice d'utiliser la stratification**, notamment lorsque la taille de l'échantillon est petite:

Exemple 2 : Imaginons maintenant que nous cherchons à estimer la moyenne de production de riz dans 4 fermes distinctes, 2 étant situées dans le Sud, deux autres dans le Nord. Des contraintes de temps nous empêchent de nous rendre dans ces 4 fermes, nous devons donc nous contenter de deux d'entre elles. Voici les informations dont nous disposons sur ces 4 fermes :

| Production (T/HA) | Région Sud | | Région Nord | | Production moyenne (T/HA) |
|-------------------|------------|---------|-------------|---------|---------------------------|
| | Ferme 1 | Ferme 2 | Ferme 3 | Ferme 4 | |
| | 2 | 2,3 | 0,4 | 0,5 | 1,3 |

Quelle est la technique d'échantillonnage à adopter compte tenu des contraintes de l'enquête ?

❶ La méthode de **tirage aléatoire simple** donne la même probabilité à chaque ferme d'être choisie parmi les quatre de notre population d'étude ($p=2/4=1/2$). Les combinaisons possibles, au nombre de 6, sont les suivantes :

| Production (T/HA) | Région Sud | | Région Nord | | Production moyenne (T/HA) |
|---|------------|---------|-------------|---------|---------------------------|
| | Ferme 1 | Ferme 2 | Ferme 3 | Ferme 4 | |
| | 2 | 2,3 | 0,4 | 0,5 | 1,3 |
| Tirages aléatoires simples possibles | | | | | |
| Tirage 1 | X | X | | | 2,15 |
| Tirage 2 | X | | X | | 1,2 |
| Tirage 3 | X | | | X | 1,25 |
| Tirage 4 | | X | X | | 1,35 |
| Tirage 5 | | X | | X | 1,4 |
| Tirage 6 | | | X | X | 0,45 |
| Moyenne des estimations | | | | | 1,3 |

La moyenne des 6 tirages aléatoires simples donne la vraie valeur de la production moyenne (1.3T/HA). On remarque ici que les **moyennes associées aux 6 tirages sont très dispersées** (moyenne allant de 0.45 à 2.15). De plus, les **productions dans chacune des deux régions retenues ici sont très différentes entre elles** (production moyenne élevée au Sud=2.15, production moyenne faible au Nord=0.45), mais **très homogènes en leur sein**.

Si l'on se rend dans deux fermes d'une même région, la production moyenne associée sera très éloignée de la production moyenne réelle, compte tenu du fait que les comportements en terme de production diffèrent énormément d'une région à l'autre. Ces combinaisons « extrêmes » sont au nombre de 2 (Tirages 1 et 6) , i.e. il y a une chance sur trois (2/6) d'obtenir une estimation biaisée de la moyenne.

L'échantillon aléatoire simple peut donc se révéler **peu précis** quand il existe des **groupes fortement différenciés** dans la population. Cela est d'autant plus vrai lorsque la **taille** de l'échantillon est **petite**.





② **La méthode de stratification**, tenant compte de l'effet « région » (en terme de production), donne également à chaque ferme du Sud ($p=1/2$) et chaque ferme du Nord ($p=1/2$) la même probabilité d'être choisie. Les 4 combinaisons possibles sont les suivantes :

| Production (T/HA) | Région Sud | | Région Nord | | Production moyenne (T/HA) |
|-------------------------------------|------------|---------|-------------|---------|---------------------------|
| | Ferme 1 | Ferme 2 | Ferme 3 | Ferme 4 | |
| | 2 | 2,3 | 0,4 | 0,5 | 1,3 |
| Tirages stratifiés possibles | | | | | |
| Tirage 2 | X | | X | | 1,2 |
| Tirage 3 | X | | | X | 1,25 |
| Tirage 4 | | X | X | | 1,35 |
| Tirage 5 | | X | | X | 1,4 |
| Moyenne des estimations | | | | | 1,3 |

Ici encore, la moyenne des 4 tirages donne la vraie valeur de la production moyenne (1.3T/HA), mais désormais, les **moyennes associées aux 4 tirages sont très rapprochées** (moyenne allant de 1.2 à 1.4). Cela est lié au fait que l'on effectue un tirage aléatoire dans chacune des catégories retenues (une ferme dans le Sud, une autre dans le Nord), empêchant ainsi la possibilité de combinaisons « extrêmes » (deux fermes d'une même région).

Cette méthode est d'autant plus efficace que la population est composée de **sous-groupes très hétérogènes entre eux** et très **homogènes en leur sein**, toute la difficulté résidant dans **le choix des variables d'intérêt²** entrant dans la construction de ces sous-groupes.

Notons que cette technique est à privilégier, surtout lorsque la **taille** de l'échantillon est **petite**, mais elle nécessite deux conditions majeurs :

- l'objectif de l'étude est bien établi (ici, calcul de la production de riz),
- vous disposez de variables d'intérêt chiffrées (proportion de la répartition géographique Nord/Sud) qui sont suffisamment discriminantes relativement à l'objet de l'étude (les fermes du Sud produisent plus que les fermes du Nord).

Le choix de l'échantillon influe sur la qualité des résultats finaux du travail d'enquête. Si votre méthode d'échantillonnage est biaisée ou votre échantillon est trop restreint, vos estimations imputées à la population totale risquent d'être elles aussi biaisées et ainsi mener à des interprétations erronées et de mauvais ciblage. Pour minimiser ces risques vous aurez besoin d'outils statistiques pour déterminer la représentativité de vos données et la fiabilité de l'information résultant de vos études.

La suite du document vous fournit ces outils et décrit la méthodologie à adopter pour l'obtention de l'échantillon optimum compte tenu des contraintes de terrain rencontrées, de la prise en compte des objectifs, à la définition de la population d'étude et de la taille de l'échantillon, en passant par le choix de la méthode d'échantillonnage adaptée, pour finir par le type de tirage aléatoire pour lequel opter.

Les questions essentielles devant être posées lors de ce processus sont :

- quelle est la population ciblée pour cette enquête et pourquoi ces données sont-elles collectées ?
- quels sont les intérêts de l'étude menée ?
- quel est l'échantillon représentatif choisi ?

A noter que les premières étapes nécessaires lors du processus d'échantillonnage sont également indispensables lors de la phase d'élaboration des questionnaires d'enquête (cf **module 2**) ; les effectuer avec rigueur permettra donc de gagner du temps (et de l'énergie !) dans le processus d'enquête.

Pour faciliter la compréhension de la méthodologie d'échantillonnage avancée, nous allons nous appuyer sur le cas pratique suivant que l'on traitera au fur et à mesure des étapes abordées.

² cf partie suivante, « 1^{ère} étape : POPULATION D'ETUDE »





CAS PRATIQUE

La zone d'action ACF considérée est composée de **deux villages (A et B)** à vocation pastorale (élevage de chameaux, vaches) et semi-pastorale (culture du sorgho et du maïs). Le **bétail et la production animale sont les principales sources de revenu**, dont le niveau est conditionné par les deux principales ressources naturelles que sont le pâturage et l'eau. **La culture de sorgho et de maïs, irriguée par les eaux de pluie, permet des revenus complémentaires.** Dans cette zone, le bétail et les conditions de vie sont **affectés par des variations climatiques saisonnières** qui, complées avec des **conditions politiques et démographiques difficiles**, ont des conséquences humanitaires sérieuses et durables sur la population de fermiers, installant une situation d'urgence depuis quelques années : ils modifient leurs habitudes de migrations qui deviennent anarchiques, créant parfois des conflits, et amenuisent les ressources et revenus.

Une étude de la situation a permis de dégager les **4 problèmes majeurs** auxquels la population de fermiers est confrontée :

- des variations considérables des ressources naturelles, saisonnières et spatiales (vagues de périodes de grandes sécheresses empêchant le pâturage de se renouveler, augmentant la vulnérabilité du bétail face à la maladie et empêchant la culture de sorgho et de maïs)
- un contexte politique instable (conflits armés qui durent depuis 20 ans et affectent les conditions de vie, ayant obligé jadis, certaines familles à se retrancher dans des camps de réfugiés)
- une augmentation de la densité de la population (causée par le retour des familles de réfugiés et l'assistance accrue en terme d'aide alimentaire)
- des capacités limitées pour faire face à ces changements (Pour faire face aux contraintes naturelles, les bergers sont forcés de se déplacer avec leurs troupeaux à la recherche de pâturage et d'eau ; seule la culture de sorgho et de maïs permet des revenus supplémentaires, mais les deux systèmes agricoles utilisés sont soit hasardeux compte tenu de la tendance à la sécheresse (irrigation par les eaux de pluie) soit peu rentables compte tenu des contraintes techniques et environnementales (pas de bon systèmes d'irrigation des eaux issues des inondations)).

Parmi les **18000 fermiers cultivant les céréales le long de la rivière**, trois sous-catégories ont été répertoriées :

Foyers agropastoraux de longue date: fermage, culture et élevage depuis longtemps

Anciens réfugiés : Essayent de combiner fermage et élevage

Pasteurs destitués : fermage pour retrouver des ressources suffisantes de bétail pour faire face à la sécheresse

Beaucoup de contraintes empêchent la population agricole d'optimiser ses ressources et l'adaptation aux incidents climatiques récurrents :

- Un service de santé animale efficace et peu coûteux, mais inexploité (les pasteurs continuent à soigner leur bétail tous seuls)
- Un faible pouvoir commercial des fermiers lié à la forte variation des prix des récoltes qui oblige les fermiers à adhérer au système de crédit (ils empruntent de l'argent lorsque le sorgho commence à pousser, le prix de la nourriture étant haut à cette période, et doivent rembourser au moment des récoltes, lorsque le prix de la nourriture est au plus bas. Les plus vulnérables sont endettés sur plusieurs années.
- la culture irriguée par les eaux des inondations fait face à des contraintes techniques et d'organisation (travail intensif, semences et insectes, invasion de mauvaises herbes)

Les recommandations exprimées après l'étude préalable sont synthétisées dans la Matrice du Cadre Logique (Logical Framework Matrix). Ainsi, pour permettre d'améliorer la résistance à la sécheresse de populations pastorales et agropastorales dans les deux zones, les **deux objectifs** assignés au projet sont les suivants :

1- Soutien des foyers en augmentant la valeur des céréales produites par le biais d'un **service de Banque de Céréales** (sans déranger ou concurrencer le traditionnel système de crédit qui fait partie des mécanismes d'adaptation). Pour cela ACF s'assurera que l'action soit menée à petite échelle et cible les plus vulnérables.

2- Travail d'irrigation des terres par le biais d'une aide de **Cash For Work** qui aura un double impact : développer un système d'irrigation bénéficiant à la communauté et aider pécuniairement les plus vulnérables, en échange de main d'œuvre. Ces derniers, endettés, pourront ainsi rembourser leurs crédits.





1^{ère} étape : OBJECTIFS DE L'ENQUETE

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.

La plupart des programmes d'aide ACF font l'objet d'enquêtes de terrain pour récolter des informations soit pour déterminer la population bénéficiaire, soit pour suivre l'impact de l'aide apportée. Dans les deux cas, et avant même de déployer une opération de terrain grande nature, il est important de bien cerner les objectifs du projet pour mener à bien ces enquêtes complémentaires.

La plupart du temps, des documents issus d'une première évaluation de la situation et une **programmation** établie sont disponibles et recèlent d'informations importantes sur les raisons du déploiement de l'aide, sa nature et ses objectifs, les critères de vulnérabilité, la définition et la taille des populations ciblées, etc...

Le responsable sur le terrain, qui est le plus apte à juger des conditions sur place, doit donc s'approprier les résultats de cette première évaluation et capitaliser les informations contextuelles disponibles pour amorcer sa réflexion sur la **justification, les objectifs et les modalités de l'enquête à mettre en œuvre.**

Justification de l'enquête

La décision de mener une enquête intervient lorsqu'il y a **nécessité de récolter des informations supplémentaires non disponibles**. Mettre en œuvre une enquête implique une mobilisation de temps et de ressources humaines et financières importantes. Il faut donc s'assurer que l'information recherchée apporte une réelle valeur ajoutée pour l'avancement du projet. Trois types de raisons sont avancées pour justifier le besoin d'une enquête de terrain dans le cadre des programmes ACF:

- **Avant l'aide** : déterminer les caractéristiques des populations vulnérables lorsque la situation de crise alimentaire est observée. Ce type d'enquête répond essentiellement aux questions « où ? », « qui ? », « combien ? ».

Ex : zones dans lesquels une inondation, une sécheresse ou un embargo ont été signalés, mais les répercussions alimentaires ne sont pas connues ; il est visible que les enfants d'une zone sont malnutris, mais l'âge et le poids critiques ne sont pas connus ; identifier les besoins et la taille des terres des fermiers dont les semences ont été ravagées pour la distribution de kits Seeds&Tools...

- **Juste après la distribution** : vérifier que l'aide bénéficie effectivement aux populations ciblées.

Ex : les bénéficiaires d'une distribution de nourriture sont bien les enfants ciblés et non des adultes ; les compléments de vitamines n'ont pas été revendus au marché ; les kits de semence n'ont pas été volés ou échangés contre d'autres articles...

- **Après l'aide** : étudier l'impact de l'aide apportée et le qualifier/quantifier. Ce sont les résultats de l'action menée, donc indisponibles dans des documents ou base de données anciennes.

Ex : évolution des caractéristiques de santé des enfants bénéficiaires d'une distribution de nourriture ; utilisation correcte et régulière des compléments de vitamines et évolution de la santé des populations bénéficiaires ; utilisation optimale des kits de semence, quantité de récoltes et de revenus correspondants ...

Cas pratique : Deux phases d'enquête et d'échantillonnage distinctes seront nécessaires :

1 : *Identification*. Il s'agit de détecter les **caractéristiques des fermiers les plus vulnérables** qui bénéficieront de l'aide.

2 : *Suivi/évaluation*. Il s'agit de **surveiller l'impact de l'aide apportée** aux fermiers bénéficiaires et s'assurer de l'atteinte des objectifs du projet.





Objectifs de l'enquête

L'enquête doit permettre de **compléter les données secondaires déjà disponibles** (et exploitées) avec de nouvelles données (primaires), en récoltant des informations utiles à la mise en place du programme en cours, auprès d'un échantillon de la population. Plusieurs types de programmes sont développés par le département de Sécurité Alimentaire d'ACF (cf documents de capitalisation et PAD pour plus de détails). Ils ont l'objectif principal d'améliorer les conditions alimentaires des populations vulnérables et des objectifs propres à chacun d'entre eux, très diversifiés.

Ex : délivrer la quantité de semence adéquate aux catégories d'agriculteurs dont les récoltes ont été dévastées par une inondation ; effectuer une distribution de vitamines auprès d'enfants malnutris ; distribuer des kits de pêche aux pêcheurs les plus sinistrés par un tsunami pour permettre la relance de l'activité dans la région...

Cas pratique : Les objectifs particuliers du projet sont :

** d'améliorer la résistance à la sécheresse de populations pastorales et agropastorales dans les deux villages d'étude A et B. L'action est ciblée sur les pasteurs les plus vulnérables cultivant les céréales le long de la rivière. L'indicateur le plus pertinent pour déterminer le niveau de vulnérabilité semble être la taille du cheptel qui, lorsqu'elle est petite, oblige les fermiers à augmenter leurs activités de culture pour subvenir à leurs besoins. Les actions à déployer sont :*

*- installer un service de Banque de Céréales qui permettra d'augmenter la valeur marchande des céréales
- proposer un système d'aide Cash for Work pour effectuer des travaux d'irrigation pour les cultures tout en permettant aux plus démunis de gagner de l'argent qui leur permettra de rembourser leurs dettes et d'améliorer leur condition alimentaire.*

** de vérifier qu'il s'agit d'une action efficace pour augmenter les revenus des fermiers (une étude similaire a permis d'estimer la hausse de revenu liée à ce type d'aide à environ +30%).*

** de proposer d'étendre ce type d'aide à d'autres populations similaires si elle s'avère efficace*

Les objectifs de l'enquête doivent **répondre aux objectifs du programme**. Selon l'état d'avancement du projet, les objectifs de l'enquête sont différents si l'on effectue le ciblage des populations vulnérables et/ou le suivi des actions d'aide en cours :

- Avant l'aide : « **définir** » : (a) les caractéristiques particulières qui définissent la population vulnérable, (b) des sous-groupes spécifiques dans cette population cible qui feront l'objet d'aides diversifiées.

Ex : trouver les facteurs communs aux populations pâtissant le plus des effets d'une inondation, d'une sécheresse ou d'un embargo en terme de sécurité alimentaire, mais les répercussions alimentaires ne sont pas connues ; il est visible que les enfants d'une zone sont malnutris, mais l'âge et le poids critiques ne sont pas connus ;

- Juste après la distribution : « **vérifier** » : savoir si les quantités ont été respectées selon le plan établi et si l'aide n'est pas détournée ou utilisée à mauvais escient.

Ex : distribution de nourriture : les bénéficiaires sont bien les enfants ciblés et non des adultes ; les compléments de vitamines n'ont pas été revendus au marché ; les kits de semence n'ont pas été volés ou échangés contre d'autres articles...

- Après l'aide : « **comprendre/projeter** » : étudier l'impact de l'aide apportée, auprès d'un échantillon, la qualifier/quantifier, pour l'extrapoler à la population et être en mesure de rectifier ou implémenter les actions menées, selon

Ex : évolution des caractéristiques de santé des enfants bénéficiaires d'une distribution de nourriture ; utilisation correcte et régulière des compléments de vitamines et évolution de la santé des populations bénéficiaires ; utilisation optimale des kits de semence, quantité de récoltes et de revenus correspondants ...

Cas pratique : Les objectifs des enquêtes sont donc de :

① : Identification. Définir la population vulnérable. *Le Cette enquête permettra donc de :*

*- s'assurer que les critères sont pertinents et déterminer un seuil pour discriminer les plus vulnérables
- déterminer le nombre exact de vulnérables et les identifier selon les critères de sélection validés*

② : Suivi/évaluation. *Effectuer le suivi de l'aide apportée. L'enquête menée doit permettre de :*

*- s'assurer de la bonne implémentation de l'aide apportée et la rectifier sinon
- savoir si l'augmentation de 30% des revenus escomptée est effective une fois l'aide déployée
- savoir si la combinaison CW+BC est plus efficace que l'action unique de CW*





Modalités d'enquête

Le type d'enquête mise en place dépend en grande partie des contraintes rencontrées sur place. Les outils permettant de déterminer ces modalités d'exécution sont délivrés en PAD et sont résumés dans le module introductif de ce guide (*cf. Introduction, partie 2, page 5 et tableau en annexe*). Il existe deux types d'enquêtes, les enquêtes rapides et les enquêtes auprès d'un échantillon représentatif (assessments).

Enquête rapide

Temps disponible : 24h à 3 semaines

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

Informations récoltées : qualitatives

Lorsque nous sommes dans une situation d'urgence, c'est le seul moyen d'obtenir rapidement de l'information supplémentaire avant de lancer l'action d'aide compte tenu des contraintes observées (manque de temps, de moyens, zones inaccessibles...). C'est à l'issue de ces entretiens que l'on prendra les décisions concernant l'action à mener. Il est donc essentiel de bien organiser ces entrevues pour récolter les informations les plus pertinentes (*cf module 3 : Opérations de terrain, partie XX*).

Lorsque l'objectif est de décrire une population ou une situation, on fait parfois appel à une pré-enquête rapide pour améliorer la qualité de l'enquête de terrain à venir. Il s'agit là de définir des caractéristiques qui différencient les populations vulnérables des autres populations, et des caractéristiques qui différencient les vulnérables entre eux, en terme de comportement en sécurité alimentaire. A la fin des entretiens, l'enquêteur est capable de catégoriser la population en quelques sous-groupes distincts (jamais plus de dix) différenciés par deux ou trois indicateurs clés (cf partie suivante : zonage/typologie). Ce travail de synthèse sera très utile pour la suite et facilitera le choix de l'échantillon à enquêter.

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 (*ex : mécanismes d'adaptation développées au sein d'une population de pêcheurs dont les pirogues ont été détruites par un ouragan ->crédit pour reprendre l'activité ? petits commerces ? envoi des enfants dans la famille ? autres ?*), ou étudier l'impact d'une action déployée par ACF (*ex : les récoltes des fermiers ayant reçu des kits Seeds&Tools permettent-elles d'améliorer leur condition alimentaire (consommation personnelle et/ou vente) ? ; la distribution de compléments vitaminés au sein des familles les plus vulnérables permet-elle une amélioration de leur condition alimentaire (regain de poids, moins de maladies) ?*). Ce processus d'enquête par sondage est beaucoup plus lourd que l'enquête rapide, mais les données récoltées sont plus nombreuses, plus précises et informatisées ce qui permet des analyses statistiques poussées. Il convient donc de réfléchir au niveau de précision des informations que l'on désire recueillir, pour ne pas se lancer inutilement dans une enquête de grand envergure.

Cas pratique : *Le type d'enquêtes menées dans notre cas seront :*

① : Identification : **Enquête rapide** auprès d'informateurs clés et de groupes de personnes pour s'informer sur les effectifs de bétail (combien pour les plus vulnérables, combien pour les autres,...), sur la force de travail disponible pour les actions de CW et d'autre informations utiles pour la mise en place de l'aide.

② : Suivi/évaluation : **Enquête par sondage** auprès d'un échantillon représentatif des bénéficiaires du programme par le biais de questionnaire. Ce type d'enquête nécessite la mise en place du processus entier d'enquête de terrain décrit dans le présent guide.

Dans le cas d'un sondage, lorsque la population est trop grande (>1000), il est d'usage d'enquêter un échantillon représentatif de celle-ci. Il doit être le plus représentatif possible de la population ciblée pour garantir l'obtention d'analyses et d'interprétations valables. L'étape suivante permet d'étudier les caractéristiques la population d'étude et ainsi établir un échantillon à l'image de son « modèle ».



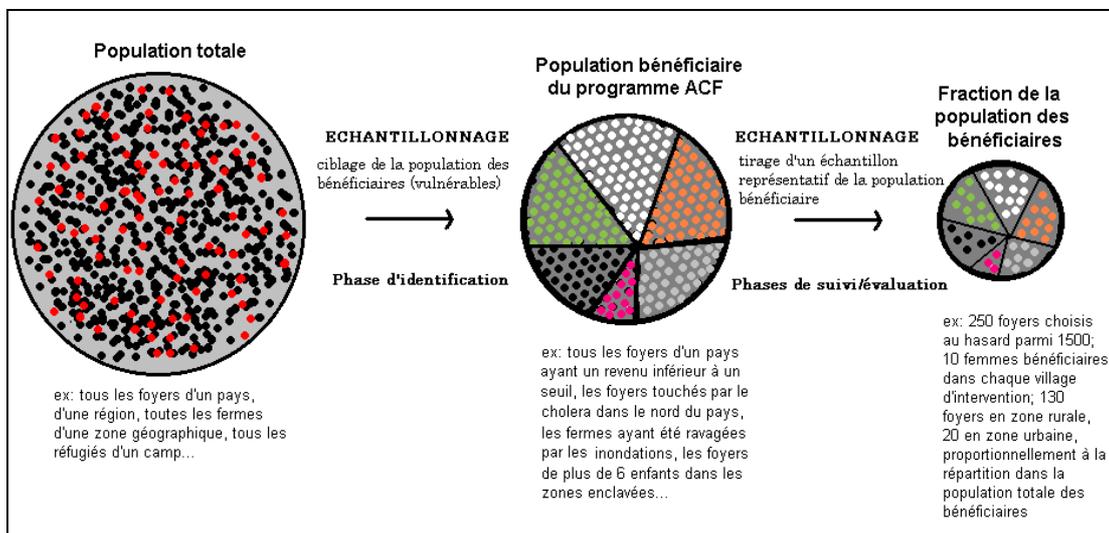


2^{ème} étape : DEFINITION DE LA POPULATION

La **connaissance de la population de référence** est **primordiale** pour établir le plan d'échantillonnage optimal, garant de la qualité des analyses menées.

La population d'étude

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 pêcheurs d'une région particulière dévastée par un cyclone dont les pirogues ont été endommagées; les fermiers de la région dont le foyer est composé de plus de 6 personnes, et dont le revenu total est inférieur à 75UM ; les enfants malnutris enregistrés dans un centre de soin ACF du camp de réfugiés*).



Unité d'étude

Selon le degré d'avancement du projet et l'intérêt de celui-ci, la population d'étude peut être différente. Dans certains cas, nous nous intéresserons à des types d'individus particuliers, comme les enfants ou les femmes. *Ex : l'aide cible les enfants de moins de 5 ans d'une population atteinte de malnutrition (les enfants plus fétiches sont plus sujets aux maladies) ; 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).*

Toutefois, dans la majeure partie des actions initiées par ACF, surtout pour les programmes de Sécurité Alimentaire, on s'intéresse le plus souvent à l'entité **foyer** (qui peut être assimilé à la **famille**, ou encore à l'**exploitation agricole**). Celui-ci représente une unité alimentaire plus facile à observer, car les revenus et la nourriture sont disponibles pour l'ensemble des personnes le constituant et sont ensuite distribués plus ou moins également en son sein. La norme internationale retenue par le PAM est de **5 individus par foyers**. (*ex : lorsque l'on parle d'une population de 1000 foyers bénéficiaires, il s'agit in fine de 5000 individus auxquels l'on fournit de l'aide*).

Cas pratique : Ici le fermier est assimilé à son foyer, car on considère que le fruit des travaux de fermes (élevage, culture, travaux d'irrigation) est la principale ressource permettant de subvenir aux besoins du foyer concerné.

- ① : **Identification** : Les fermiers des villages A et B (ceux qui élèvent du bétail et cultivent le sorgho au bord de la rivière).
- ② : **Suivi/évaluation** : L'ensemble des fermiers (foyers) des villages A et B bénéficiaires de l'aide ACF.





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 (cf **partie suivante**) 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).

Plus la mise en place d'un projet ACF est avancée, plus la taille des populations considérées s'amenuise. Ainsi on commence par étudier une population globale, puis on se cantonne à l'étude d'une population particulière pour trouver les plus vulnérables, pour finir par l'analyse d'un échantillon de la population des bénéficiaires. (cf **graphique p.XX Introduction, partie 2: « étapes successives de récolte d'information »**) :

Au moment de la programmation, la population de départ est souvent l'ensemble de la population nationale, ou une (des) région(s) où l'on a connaissance de problèmes graves en sécurité alimentaire (ex : *population d'un pays où règnent des conflits armés – 7M d'habitants*).

Cas pratique : *l'ensemble de la population de la région qui subit la sécheresse depuis plusieurs années (3M de personnes).*

La détection de problèmes alimentaires permet ensuite de se focaliser sur quelques zones géographiques particulières (ou parlera de zones d'intervention prioritaires). Le projet ACF s'installera là où les besoins sont les plus importants, en fonction de la capacité d'intervention (critères de sécurité, existence ou non d'autres projets similaires, niveau de couverture...). La population devient alors l'ensemble de la population de la zone ACF (ex : *population totale d'un camp de réfugiés d'un pays en conflit – 150 000 personnes*).

Cas pratique : *Population d'étude pour la phase ① : Identification : l'ensemble des 18 000 fermiers des villages A et B couverts par la base ACF. Il s'agit des villages accessibles (les conflits empêchent de pouvoir agir partout) où l'insécurité alimentaire était la plus grave dans la région subissant la sécheresse, et où réside une communauté de pasteurs ayant du mal à développer des stratégies d'adaptation (les bêtes sont affaiblies (plus de pâturage) et les cultures secondaires mal irriguées). Ainsi les pasteurs sont obligés d'avoir recours à un système de crédit pour subvenir à leurs besoins alimentaires, et n'arrivent pas à rembourser leurs dettes compte tenu des périodes de sécheresse répétées.*

Une fois la base ACF installée, la phase d'identification permet de détecter les populations les plus vulnérables (au niveau géographique et selon des critères bien définis). Elles constituent la nouvelle population de référence (ex : *les enfants malnutris enregistrés dans les centres de santé d'un camp de réfugiés d'un pays en conflit – 50 000 enfants*).

Cas pratique : *La phase d'identification a permis de déterminer un seuil critique pour cibler les plus vulnérables parmi les fermiers qui élèvent du bétail et cultivent le sorgho au bord de la rivière :*

- ceux dont l'effectif du **troupeau est inférieur à 4 bêtes** sont incapables de faire face à une sécheresse pour maintenir leur troupeau rentable. Ils sont au nombre de 4000.
- parmi ces 4000 fermiers, **une partie peut fournir de la main d'œuvre** pour construire un système d'irrigation plus efficace qui bénéficiera à toute la communauté. Ce travail complémentaire permettra aux plus démunis de rembourser leurs dettes.

La population ciblée

C'est auprès de cette population des plus vulnérables que sera déployée l'aide, la population d'étude devient donc l'ensemble des bénéficiaires du projet ACF. Parfois, par manque de moyens logistiques ou financiers, il sera impossible de venir en aide à toutes les personnes dans le besoin. Dans ce cas précis, la population des bénéficiaires sera donc plus petite que la population définie comme vulnérable. (ex : *les enfants malnutris enregistrés dans les centres de santé où n'agissent pas d'autres ONG d'un camp de réfugiés d'un pays en conflit – 30 000 enfants*).





Cas pratique : *Taille de la population des bénéficiaires = 4000 fermiers (20000 individus), dont :*
 - 2500 qui bénéficieront uniquement du service de Banque de Céréales, permettant de vendre leur sorgho à un prix plus élevé que celui pratiqué sur le marché
 - 1500 fermiers bénéficieront d'une aide couplée BC+CFW ;

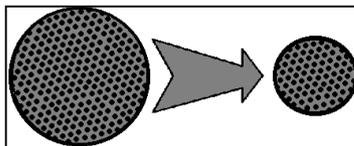
Critères d'échantillonnage

Une fois l'aide déployée, un système de suivi est mis en place. La population des bénéficiaires est souvent trop vaste pour être enquêtée dans sa totalité (*ex : 30 000 enfants*). Une partie seulement de cette population, **l'échantillon**, fera l'objet d'une enquête après la distribution de l'aide (*ex : 500 enfants parmi les 30 000 bénéficiaires*). L'échantillon doit avoir les mêmes caractéristiques que la population d'étude pour que l'on soit en mesure d'inférer à la population d'étude les résultats obtenus pour l'échantillon.

Cas pratique : 2 : *Suivi/évaluation : Un échantillon représentatif des 4000 fermiers bénéficiaires sera choisi.*

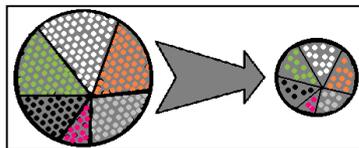
Pour choisir l'échantillon d'étude, on peut effectuer un tirage totalement aléatoire, sans aucune autre considération que les critères de vulnérabilité qui définissent la population ciblée (*cf graphique 1*).

Graphique 1 : tirage aléatoire simple



Néanmoins, il existe des procédés préalables au tirage, telle que la **stratification** qui, en assurant représentativité de l'échantillon, donne une meilleure précision des sondages aléatoires. La recherche des populations vulnérables pour un projet de sécurité alimentaire est l'application directe de la stratification (**Food Economy Zones et Typologies :cf partie XX Introduction**). Cette technique permet de diviser la population en un nombre de groupes homogènes (strates) définis selon des variables d'intérêt (*ex : terres irriguées/non irriguées, revenus existants/inexistants, famille nombreuse/un seul enfant, activités de commerce supplémentaire oui/non, 1 /plusieurs repas par jour...*). Le tirage de l'échantillon sera effectué indépendamment, groupe par groupe, afin de diminuer les fluctuations d'échantillonnage. Lorsque une répartition chiffrée des différentes strates est disponible (exacte ou approximative, *ex : 30% mangent plusieurs repas par jour, 70% un seul*), il est alors possible de tirer un échantillon stratifié proportionnel assurant une meilleure représentativité de la population d'étude que celle obtenue par tirage aléatoire simple (*cf graphique 2*) :

Graphique 2 : tirage stratifié proportionnel



Variables d'intérêt

Pour les besoins de la phase de suivi d'un projet d'aide en sécurité alimentaire, il est d'usage d'enquêter un échantillon de la population des bénéficiaires. Une stratification de cette population de bénéficiaires permet de dégager les sous-groupes (strates) susceptibles de réagir différemment face à l'aide apportée. Le secret d'un bon échantillonnage réside dans le choix des variables, en lien avec les objectifs de l'étude, qui vont différencier au mieux les strates d'études. Une analyse approfondie des informations disponibles, notamment par le biais d'entretiens de groupes auprès de la population, permet de dégager les strates de la population étudiée (*ex : Chez les enfants malnutris de moins de 5 ans, les variables susceptibles d'influencer le résultat des compléments vitaminés apportés sont: 1. Structure familiale (deux parents, monoparentale, ... ; 2. Nombre d'enfants dans la famille (moins de 4, plus de 4) ; 3. Première visite dans le Centre de Nutrition thérapeutique (oui/non)...*).

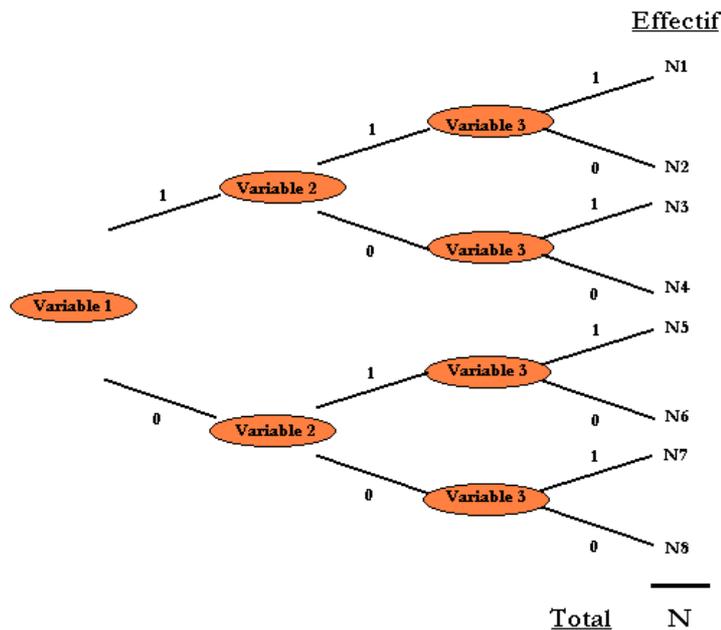




Cas pratique : 2 : *Suivi/évaluation* : les trois types de fermiers bénéficiant de l'aide sont soit de « longue date » (vivant de culture et d'élevage depuis longtemps), soit des « anciens réfugiés » (combinant culture et élevage depuis leur retour) soit des « pasteurs destitués » (complètent leurs activités d'élevage par de la culture depuis les vagues de sécheresse). Cette hiérarchie ne permet néanmoins pas de classer automatiquement les uns et les autres selon leur degré d'adaptation aux sécheresses des dernières années. A ces trois types de fermiers, on retient trois variables différenciant ces fermiers en terme de vulnérabilité :

1. **Taille du bétail** (ils ont tous moins de 4 bêtes, mais on différencie ceux qui en ont 0 et ceux qui en ont entre 1 et 3)
2. **Durée d'autonomie alimentaire rendue possible par les récoltes** (soit elle est supérieure à 6 mois, et les foyers peuvent subsister aux sécheresses, soit elle est inférieure à 6 mois, et les foyers doivent avoir recours au crédit).
3. **Endettement** (cette variable permet d'estimer les bénéfices entraînés par la vente de récoltes : selon le **seuil** arrêté à **75 UM**, les fermiers sont capables ou non de rembourser leurs crédit, et donc de faire face à la sécheresse d'une année sur l'autre

Arbre de décision



L'arbre de décision, qui se présente comme suit, se trace à l'aide des variables d'intérêt retenues. Cette étape est utile pour recueillir 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 critique du revenu des foyers est de 75 UM/personne, pour l'arbre de décision associé : Revenu=0 si le revenu < 75 UM/personne & Revenu=1 si le revenu > 75 UM/personne).

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

Cas pratique : L'arbre de décision pour notre cas est à développer pour les trois catégories des fermiers bénéficiaires (« longue durée », « anciens réfugiés », « pasteurs destitués »), chacune ayant ses propres mécanismes d'adaptation. Les trois variables d'intérêt retenues sont représentées par les trois logos suivants :



: **Bétail.** Prend la valeur 0 si le fermier n'a aucune bête ; prend la valeur 1 s'il en possède entre 1 et 3.



: **Récolte permettant une autonomie alimentaire.** Prend la valeur 1 si l'autonomie > 6 mois ; 0 sinon

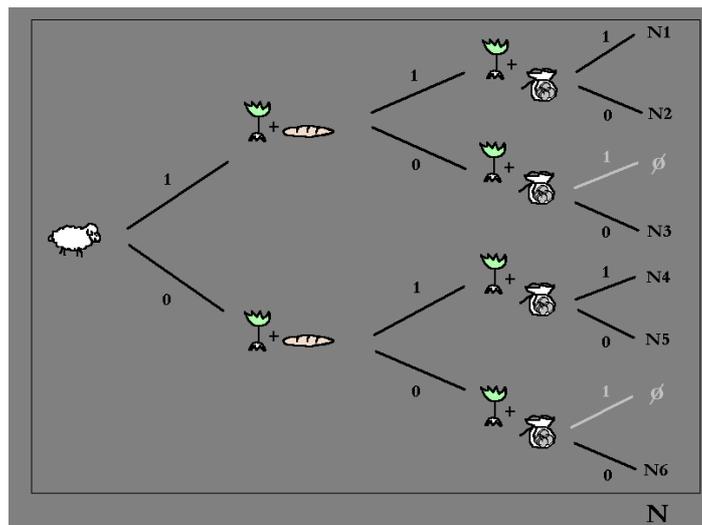


: **Récolte permettant une autonomie alimentaire et des bénéfices issus de la vente.** Prend la valeur 1 si oui ; 0 sinon. La combinaison bénéficie sans autonomie alimentaire > 6 mois n'est pas possible (∅ dans l'arbre)





Cas pratique : l'arbre de décision est le suivant :



L'arborescence est retranscrite dans un tableau Excel que l'on cherche à remplir par les effectifs observés dans la population, catégorie par catégorie. L'effectifs total des colonnes est toujours égal à l'effectif total (N) de la population.

| Variable 1 | | Variable 2 | | Variable 3 | |
|------------|----|------------|----|------------|----|
| 0 | xx | 0 | xx | 0 | xx |
| | | 1 | xx | 1 | xx |
| 1 | xx | 0 | xx | 0 | xx |
| | | 1 | xx | 1 | xx |
| Total | N | Total | N | Total | N |



Souvent il est difficile d'obtenir les effectifs exacts ; il convient alors de les estimer à partir des répartitions approximatives (ex : une population de 500 enfants malnutris compte environ 30% d'enfants atteints de diarrhée (nous ne disposons pas des effectifs exacts)). La variable Maladie est discriminante pour différencier les enfants. Cette variable prendra la valeur 1 pour 30%*500=150 enfants ; elle prendra la valeur 0 pour le reste, i.e. 350 enfants (70%). Ensuite on sait que parmi ces enfants malades, 80% fréquentent le centre de santé pour la première fois. Cette information permet de compléter le tableau au niveau supérieur (variable 1^{ère} visite CNT). On appliquera ce taux de 80% non pas à la totalité des 500 enfants, mais aux 150 enfants ayant été classés dans la catégories Malade :

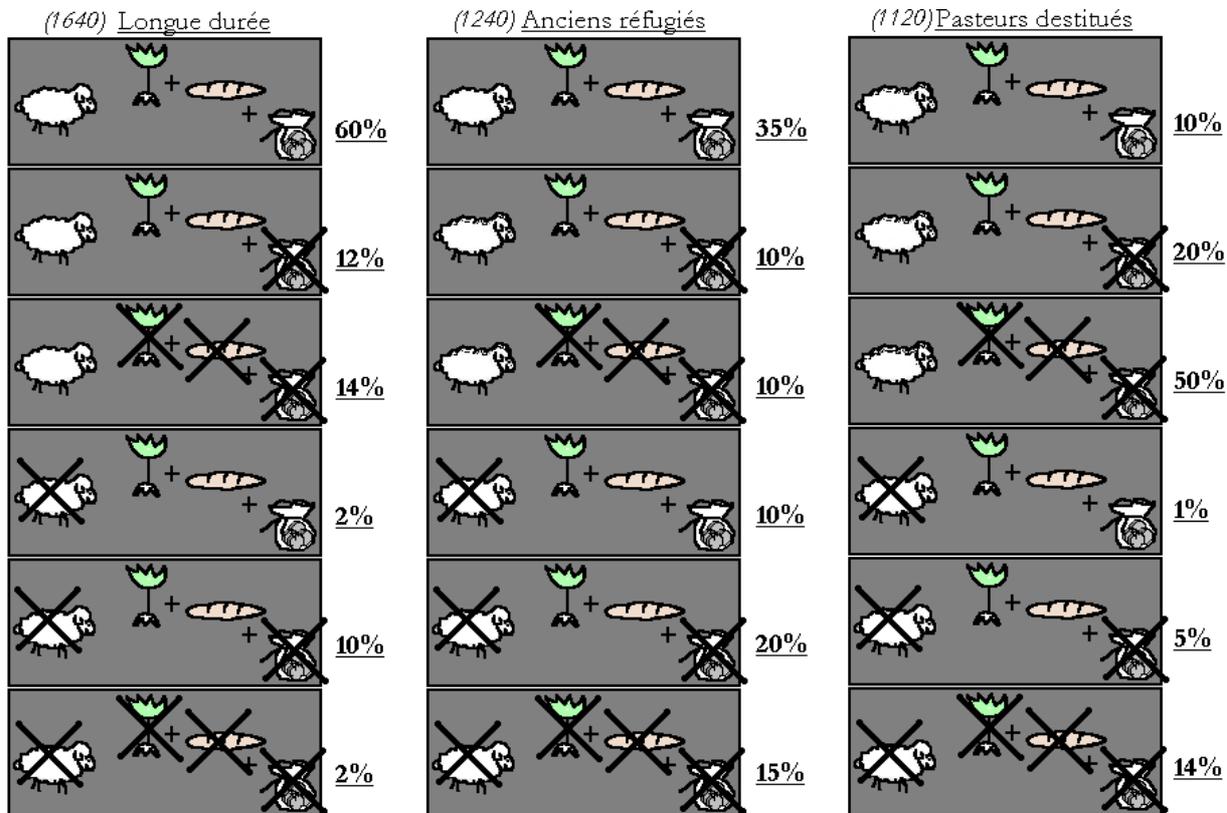
| Maladie | | 1 ^{ère} Visite CNT | |
|---------|-----------------------------------|-----------------------------|----------------------------|
| 0 | 350 | 0 | xx |
| 1 | 150 <small>=500*30/100</small> | 0 | xx |
| | | 1 | <small>=80*150/100</small> |
| Total | 500 | | |

Selon les informations disponibles (effectifs réels, estimations), le remplissage du tableau sera plus ou moins rapide. Pour compléter les informations manquantes, il est judicieux de construire un premier tableau d'échantillonnage AVANT les entretiens de groupes lors du ciblage de la population des bénéficiaires, à l'aide des variables discriminantes a priori. Les entretiens permettront d'affiner la qualité de la stratification. Il convient alors de le remplir avec les données disponibles et de récolter les informations manquantes lors des divers entretiens. (cf utilisation de l'outil lors des entretiens chap. XX partie .. p.xx)





Cas pratique : Les effectifs des types de fermiers sont connus et leurs répartition selon les critères de stratification retenus ont été estimées lors des différents entretiens de groupe (cf techniques d'estimation chap. XX p. xx). La population des bénéficiaires est donc répartie comme suit :



A l'aide de ces indications, nous sommes en mesure de renseigner les effectifs de la dernière colonne dans le tableau d'échantillonnage correspondant (l'effectif total est bien égal à 4000 fermiers) :

| Type de ferme | Bétail | | Autonomie alimentaire > 6 mois | | Génération de bénéfiques | |
|----------------------------|-------------|-------------|--------------------------------|-------------|--------------------------|-----|
| | 1 | 0 | 1 | 0 | 1 | 0 |
| Longue durée 1640 | 1 | 1410 | 1 | 1181 | 1 | 984 |
| | | | 0 | 229 | 0 | 197 |
| | | | 0 | 229 | 1 | 0 |
| | 0 | 230 | 1 | 197 | 1 | 33 |
| | | | 0 | 33 | 0 | 164 |
| | | | 0 | 33 | 1 | 0 |
| Anciens réfugiés 1240 | 1 | 682 | 1 | 558 | 1 | 434 |
| | | | 0 | 124 | 0 | 124 |
| | | | 0 | 124 | 1 | 0 |
| | 0 | 558 | 1 | 372 | 1 | 124 |
| | | | 0 | 186 | 0 | 248 |
| | | | 0 | 186 | 1 | 0 |
| Pasteurs destitués 1120 | 1 | 896 | 1 | 336 | 1 | 112 |
| | | | 0 | 560 | 0 | 224 |
| | | | 0 | 560 | 1 | 0 |
| | 0 | 224 | 1 | 67 | 1 | 11 |
| | | | 0 | 157 | 0 | 56 |
| | | | 0 | 157 | 1 | 0 |
| 4000 | 4000 | 4000 | 4000 | 4000 | 4000 | |





Le tableau peut être complété aux niveaux de stratification supérieurs (« Autonomie alimentaire > 6 mois » et « Bétail ») :

| Type de ferme | Bétail | | Autonomie alimentaire > 6 mois | | Génération de bénéfiques | | Strate | PoidsStrate |
|----------------------------|--------|------|--------------------------------|------|--------------------------|------|--------|-------------|
| Longue durée 1640 | 1 | 1410 | 1 | 1181 | 1 | 984 | 1 | 25% |
| | | | 0 | 229 | 0 | 197 | 2 | 5% |
| | | | 0 | 229 | 0 | 229 | 3 | 6% |
| | 0 | 230 | 1 | 197 | 1 | 33 | 4 | 1% |
| | | | 0 | 33 | 0 | 164 | 5 | 4% |
| | | | 0 | 33 | 0 | 0 | 6 | 0% |
| Anciens réfugiés 1240 | 1 | 682 | 1 | 558 | 1 | 434 | 7 | 11% |
| | | | 0 | 124 | 0 | 124 | 8 | 3% |
| | | | 0 | 124 | 0 | 124 | 9 | 0% |
| | 0 | 558 | 1 | 372 | 1 | 124 | 10 | 3% |
| | | | 0 | 186 | 0 | 248 | 11 | 6% |
| | | | 0 | 186 | 0 | 0 | 12 | 0% |
| Pasteurs destitués 1120 | 1 | 896 | 1 | 336 | 1 | 112 | 13 | 3% |
| | | | 0 | 560 | 0 | 224 | 14 | 6% |
| | | | 0 | 560 | 0 | 0 | 15 | 0% |
| | 0 | 224 | 1 | 67 | 0 | 560 | 16 | 14% |
| | | | 0 | 157 | 1 | 11 | 17 | 0% |
| | | | 0 | 157 | 0 | 56 | 18 | 1% |
| 4000 | | 4000 | | 4000 | | 4000 | | 100% |

Strates retenues

Une fois le tableau d'échantillonnage rempli, on peut comptabiliser le nombre de sous-groupes détectés dans la population étudiée, selon les critères de stratifications retenus. Ce nombre équivaut au nombre de cellules renseignées et différentes de 0 dans la dernière colonne.

Cas pratique : 18 strates distinctes sont renseignées (cf colonne **Strate** dans le tableau précédent).

Pour chacun de ces sous-groupes on obtient son **poids** (représentativité) dans la population totale en calculant le rapport entre l'effectif de la strate et celui de la population totale (**ex : PoidsStrate1 = n1/N**). La somme des poids est égale à 1 (=100%).

Cas pratique : Le poids de la Strate « fermiers de longue durée, avec bétail, et avec récoltes permettant une autonomie alimentaire de plus de 6 mois et des bénéfices liés à leur vente » est égal à $984/4000 = 0.25$. Cela signifie que 25% des bénéficiaires appartiennent à cette Strate (cf colonne **PoidsStrate** du tableau précédent pour les autres poids).

Il peut arriver que certaines strates soient très peu représentées. Cela est vérifié lorsque le **poids est inférieur à 5%**. 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, mais assez grandes pour que l'on soit en mesure d'émettre des conclusions après l'analyse de leur caractéristiques spécifiques. 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** (en Sécurité Alimentaire, les échantillon ont rarement une taille supérieure à 2000 individus).

Cas pratique : Parmi les 18 strates, celles à petits effectifs ont été regroupées, selon la logique de l'arborescence retenue, pour n'en former plus que 10 (cf regroupements par les cercles dans le tableau précédent) :





| Type de ferme | Bétail | | Autonomie alimentaire > 6 mois | | Génération de bénéfiques | | Strate | PoidsStrate | Strate | PoidsEchant |
|----------------------------|--------|------|--------------------------------|------|--------------------------|-----|--------|-------------|--------|-------------|
| Longue durée 1640 | 1 | 1410 | 1 | 1181 | 1 | 984 | 1 | 25% | 1 | 25% |
| | | | 0 | 229 | 0 | 197 | 2 | 5% | 2 | 5% |
| | | | 0 | 33 | 0 | 229 | 3 | 6% | 3 | 6% |
| | 0 | 230 | 1 | 197 | 1 | 33 | 4 | 1% | 4 | 6% |
| | | | 0 | 33 | 0 | 164 | 5 | 4% | | |
| | | | 0 | 33 | 0 | 0 | 6 | 0% | | |
| Anciens réfugiés 1240 | 1 | 682 | 1 | 558 | 1 | 434 | 7 | 11% | 5 | 17% |
| | | | 0 | 124 | 0 | 124 | 8 | 3% | | |
| | | | 0 | 124 | 0 | 124 | 9 | 3% | | |
| | 0 | 558 | 1 | 372 | 1 | 124 | 10 | 3% | 6 | 9% |
| | | | 0 | 186 | 0 | 248 | 11 | 6% | | |
| | | | 0 | 186 | 0 | 0 | 12 | 0% | | |
| Pasteurs destitués 1120 | 1 | 896 | 1 | 336 | 1 | 112 | 13 | 3% | 8 | 8% |
| | | | 0 | 560 | 0 | 224 | 14 | 6% | | |
| | | | 0 | 560 | 0 | 560 | 15 | 0% | | |
| | 0 | 224 | 1 | 67 | 1 | 11 | 16 | 0% | 10 | 6% |
| | | | 0 | 157 | 0 | 56 | 17 | 1% | | |
| | | | 0 | 157 | 0 | 0 | 18 | 0% | | |
| 4000 | 4000 | 4000 | 4000 | 100% | 100% | | | | | |

Tableau d'échantillonnage

Une fois ces considérations de poids et de nombre de strates, le plan d'échantillonnage peut être ébauché. Ce plan contient les informations suivantes :

NumStrate : Numéro de la strate. Il permettra de retrouver un individu (ou un foyer) selon les caractéristiques d'échantillonnage retenues (lorsque la base de données sera disponible, chaque individu sera affecté du numéro de strate auquel il appartient)

LibStrate : libellé de la strate associée

NbBénéf : nombre d'individus faisant partie de la strate parmi la totalité des bénéficiaires (ou des foyers faisant partie de la strate associée) ayant les caractéristiques de la strate associée

PoidsStrate : représentativité en terme de nombre d'individus (ou de foyers) de la strate dans la population totale (rapport de l'effectif de la strate sur l'effectif total)

NbEchant : nombre d'individus à tirer dans chaque strate pour faire partie de l'échantillon

Cas pratique : Le plan d'échantillonnage est le suivant :

| NumStrate | LibStrate | NbBenef | PoidsStrate (%) | NbEchant |
|-----------|--|---------|-----------------|----------|
| 1 | Fermiers de longue date , ayant 1 à 3 bêtes , dont les récoltes permettent une autonomie alimentaire de plus de 6 mois et sont génératrices de revenus supplémentaires | 984 | 25% | |
| 2 | Fermiers de longue date , ayant 1 à 3 bêtes , dont les récoltes permettent une autonomie alimentaire de plus de 6 mois et mais sans revenus supplémentaires | 197 | 5% | |
| 3 | Fermiers de longue date , ayant 1 à 3 bêtes , dont les récoltes ne permettent pas une autonomie alimentaire suffisante | 229 | 6% | |
| 4 | Fermiers de longue date , n'ayant plus de bétail | 230 | 6% | |
| 5 | Fermiers anciennement réfugiés , ayant 1 à 3 bêtes | 682 | 17% | |
| 6 | Fermiers anciennement réfugiés , n'ayant plus de bétail , dont les récoltes permettent une autonomie alimentaire de plus de 6 mois | 372 | 9% | |
| 7 | Fermiers anciennement réfugiés , n'ayant plus de bétail , récoltes ne permettent pas une autonomie alimentaire suffisante | 186 | 5% | |
| 8 | Fermiers pasteurs destitués , ayant 1 à 3 bêtes , dont les récoltes permettent une autonomie alimentaire de plus de 6 mois | 336 | 8% | |
| 9 | Fermiers pasteurs destitués , ayant 1 à 3 bêtes , dont les récoltes ne permettent pas une autonomie alimentaire suffisante | 560 | 14% | |
| 10 | Fermiers de pasteurs destitués , n'ayant plus de bétail | 224 | 6% | |

Pour compléter le tableau nous devons d'abord déterminer la taille de l'échantillon. La partie suivante traite spécifiquement de cet aspect.





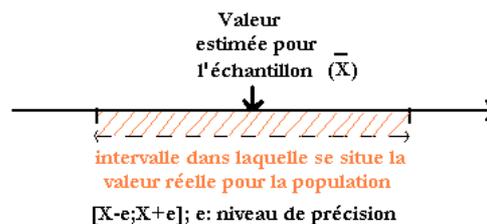
3^{ème} étape : LA TAILLE DE L'ÉCHANTILLON

« Quelle est la taille de l'échantillon que je dois tirer pour qu'il soit valable ? » Cette question est fréquemment posée lorsque l'on parle d'échantillonnage.

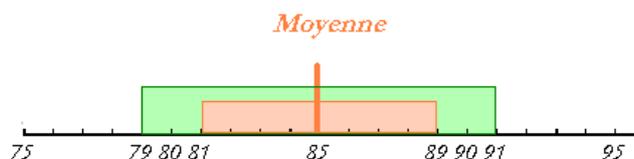
Pour déterminer la taille de l'échantillon à tirer et ainsi permettre des interprétations valables des analyses qui seront effectuées à partir de celui-ci, outre la taille de la population d'étude, il est nécessaire de prendre en compte certains critères statistiques pour calculer la taille de l'échantillon souhaité, par le biais de formules mathématiques éprouvées. Le choix de la taille d'un échantillon dépend aussi des contraintes de terrain observées et des questions auxquelles on désire répondre par le moyen de l'enquête : « S'agit-il simplement de calculer des statistiques descriptives (moyenne, fréquences), ou d'effectuer une analyse approfondie d'une population composée de plusieurs sous-groupes ayant bénéficié de plusieurs types d'aide ? » Plus les objectifs de l'étude sont pointus et la diversité des individus enquêtés est grande, plus la taille minimale requise sera grande. Au delà, il existe des cas particuliers qui nécessitent une attention particulière lors du calcul de la taille de l'échantillon comme *celui des enquêtes rapides, de la comparaison de moyennes ou lorsque la population est enquêtée par grappes* ». Toutes ces considérations sont importantes pour effectuer un choix optimum.

Critères statistiques

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 de 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]) :



En règle générale, le niveau de précision retenu pour les enquêtes en Sécurité Alimentaire est de +/- 5%, mais il peut être plus large selon les grandeurs estimées et surtout l'homogénéité de la population (cf degré de variabilité). Lorsque l'on effectue une comparaison des moyennes par exemple, il faut s'assurer que le niveau de précision est assez fin pour détecter les différences si elles existent.

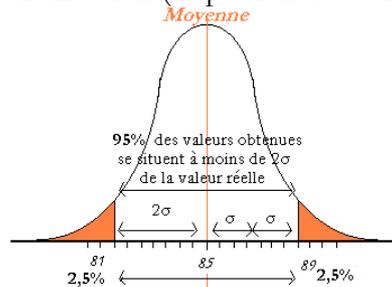




Niveau de confiance

Il y a toujours un risque que l'échantillon sélectionné ne représente pas la population étudiée. Le niveau de confiance (ou marge d'erreur) permet d'indiquer le pourcentage de chances que l'échantillon sélectionné a d'être 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 : 84;85;84;86;... ;87;83;84;84;... ;83;81;87;... ;86;82;79 ;... ;85;85;86;85;89;84 : moyenne=85UM*) sera égale à la moyenne réelle de la population étudiée (*ex : revenu moyen=85UM*). Les valeurs obtenues pour ces échantillons suivent une distribution normale autour de la moyenne réelle. Certaines sont proches de la valeur réelle (*ex : 84;85;86*), d'autres sont plus éloignées (*ex : 79;81;89...*). La déviation standard (ou écart-type³, noté σ) 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 écarts-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 écarts-types. Elles sont considérées comme trop éloignées de la moyenne réelle et les échantillons correspondants ne sont pas représentatifs de la populations d'étude (*ex : on retrouve ces 5% de cas extrêmes dans les « queues » (zone orange) de la cloche du graphique ci-dessus*).

Ainsi, plus le niveau de confiance retenu est fort, moins le risque de tirer un échantillon éloigné de la population étudiée est élevé. 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. La table statistique en Annexe II (p45) donne les valeurs de t équivalentes selon les tailles et niveaux de confiances désirés.

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 (*c'est souvent le cas dans les enquêtes NUT auprès de populations d'enfants malnutris*) 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 (*ex : administration de deux types de lait aux moins de 6 mois*). En revanche, lorsque l'objectif de l'étude est simplement de décrire les comportements de populations que l'on sait a priori différenciées, le degré de précision sera moins fin (*ex : connaître la moyenne du rendement des semences distribuées dans une région ; $e= +/-14%$*).

³ N.B. : Plus l'écart type sera faible, plus les échantillons se ressembleront entre eux et meilleure sera la précision de l'estimation.





Calcul de la taille de l'échantillon

Petite population/ Etudes similaires

Lorsque la population est petite (**200 individus ou moins**), il est préférable de l'enquêter dans sa totalité. Les coûts associés au déploiement de l'enquête seront les mêmes si l'on enquête 50 ou 200 individus, et enquêter toute la population évite les erreurs d'échantillonnage, plus conséquentes lorsque la population totale est petite.

Lorsque des **enquêtes similaires** ont été menées sur la même population, il est préférable d'utiliser la **même taille d'échantillon** qui permettra des comparaisons intéressantes. Cette approche doit être retenue dans le cas où le **plan d'échantillonnage est valide** et répond aux attentes et conditions de l'étude en cours (même population ciblée, mêmes objectifs que l'enquête précédente, mêmes zones accessibles, pas de déplacement de population de masse...).

Tables statistiques

Lorsque les critères de choix sont standards et prédéfinis, on peut se référer aux tables statistiques existantes qui présentent les différentes tailles d'un échantillon aléatoire simple selon la taille de la population et le niveau de précision désiré (pour des niveaux de confiance (95%) et d'hétérogénéité (P=0.5)). A noter que :

- (1) Ces tailles représentent le nombre d'individus effectivement enquêtés. Il est donc important de **prévoir un échantillon complémentaire** pour palier aux phénomènes de réponses erronées et de non-réponse. En pratique on ajoute **10%** de l'échantillon initial si l'enquête se base sur des **réponses déclaratives** (ex : *fermiers, femmes,...*) et **5%** si **l'enquêteur effectue lui même les mesures** (ex : *mesures anthropométriques*).
- (2) La table concernant les populations de petite taille part du postulat que la population suit une distribution normale (la normalité n'est pas automatique avec des populations de petite taille).

| Tables statistiques pour un niveau de précision de 95% et une population d'étude hétérogène (P=0,5) selon : | | | | | | | | | |
|---|--|-------|-------|--------|-------------------------|--|-------|-------|--------|
| (i) la taille de la population d'étude; (ii) le niveau de précision désiré (+/-3%, +/-5%, +/-7%, +/-10%) | | | | | | | | | |
| Taille de la population | Taille de l'échantillon selon la précision : | | | | Taille de la population | Taille de l'échantillon selon la précision : | | | |
| | +/-3% | +/-5% | +/-7% | +/-10% | | +/-3% | +/-5% | +/-7% | +/-10% |
| 100 | - | 81 | 67 | 51 | 500 | all | 222 | 145 | 83 |
| 125 | - | 96 | 78 | 56 | 600 | all | 240 | 152 | 86 |
| 150 | - | 110 | 86 | 61 | 700 | all | 255 | 158 | 88 |
| 175 | - | 122 | 94 | 64 | 800 | all | 267 | 163 | 89 |
| 200 | - | 134 | 101 | 67 | 900 | all | 277 | 166 | 90 |
| 225 | - | 144 | 107 | 70 | 1 000 | all | 286 | 169 | 91 |
| 250 | - | 154 | 112 | 72 | 2 000 | 714 | 333 | 185 | 95 |
| 275 | - | 163 | 117 | 74 | 3 000 | 811 | 353 | 191 | 97 |
| 300 | - | 172 | 121 | 76 | 4 000 | 870 | 364 | 194 | 98 |
| 325 | - | 180 | 125 | 77 | 5 000 | 909 | 370 | 196 | 98 |
| 350 | - | 187 | 129 | 78 | 6 000 | 938 | 375 | 197 | 98 |
| 375 | - | 194 | 132 | 80 | 7 000 | 959 | 378 | 198 | 99 |
| 400 | - | 201 | 135 | 81 | 8 000 | 976 | 381 | 199 | 99 |
| 425 | - | 207 | 138 | 82 | 9 000 | 989 | 383 | 200 | 99 |
| 450 | - | 212 | 140 | 82 | 10 000 | 1000 | 385 | 200 | 99 |
| | | | | | 15 000 | 1034 | 390 | 201 | 99 |
| | | | | | 20 000 | 1053 | 392 | 204 | 100 |
| | | | | | 25 000 | 1064 | 394 | 204 | 100 |
| | | | | | 50 000 | 1087 | 397 | 204 | 100 |
| | | | | | 100 000 | 1099 | 398 | 204 | 100 |
| | | | | | <100 000 | 1111 | 400 | 204 | 100 |



Cas pratique : Taille de l'échantillon

dans notre cas, la population d'étude compte **4000 fermiers** (unité de mesure). Les critères retenus sont standards, la population suivant à priori une distribution normale et présentant un degré de variation acceptable (niveau de confiance : 95%, degré de variabilité par défaut (P=0.5), niveau de précision : **+/-5%**). Dans le tableau d'échantillonnage, au croisement des valeurs 4000 (colonne) et « +/-5% » (ligne), on obtient la taille requise pour disposer d'un **échantillon représentatif de la population** (364 fermiers). Nous devons ajouter 10% de fermiers en plus pour palier aux phénomènes de non réponse et réponses erronées, ce qui donne un échantillon de **364+364x10%=400 fermiers**





Formules mathématiques

Formule simplifiée

Bien que les tables soient très pratiques, vous aurez parfois besoin de calculer la **taille de l'échantillon** pour d'autres valeurs des critères statistiques retenues. La formule suivante permet ce calcul, avec une confiance de 95% et une variabilité maximale de 50% :

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



N = taille de la population
e = niveau de précision

Cas pratique : Taille de l'échantillon

Le niveau de précision retenu est toujours +/-5%, et la taille de la population totale = 4000 fermiers. Nous obtenons la taille d'échantillon suivante :

$$n = 4000 / (1 + 4000 \times (0.05 \times 0.05)) = 4000 / 11 = \mathbf{364 \text{ fermiers}}$$

Nous ajoutons les 10% absorbant les pertes liées aux enquêtes de terrain : Taille de l'échantillon = **400 fermiers**

Formule pour les proportions

Lorsque la prévalence estimative de la caractéristique étudiée (degré de variabilité) est connue (ex : proportion de personnes atteintes de malnutrition chronique=20%, proportion de fermiers ayant amélioré leur rendement=60%, ...), on peut calculer la taille de l'échantillon requise en utilisant la formule suivante :

$$n = \frac{t^2 \times p \times q}{e^2} = \frac{t^2 \times p \times (1-p)}{e^2}$$



e = niveau de précision

p = degré de variabilité (taux de prévalence estimative)

t = valeur type associée au niveau de confiance requis (95% -> 1,96 ; cf table annexes II)

Cette formule renvoie une taille d'échantillon plus grande que précédemment, mais elle est utile lorsque nous ne connaissons pas l'effectif de la population totale.

Cas pratique : Taille de l'échantillon

Imaginons que notre étude porte exclusivement sur la possession ou non de bétail. 75% est la proportion des fermiers de notre population détenant au moins une bête. Nous pouvons indiquer un degré de variabilité $p=0.75$ et $q=1-p = 0.25$. Avec un niveau de précision à +/-5%, un niveau de confiance standard de 95%), nous obtenons la taille d'échantillon suivante :

$$n = ((1.96 \times 1.96) \times (0.75 \times 0.25)) / (0.05 \times 0.05) = 0.9604 / 0.0025 = \mathbf{288 \text{ fermiers}}$$

En ajoutant les 10% absorbant les pertes liées aux enquêtes de terrain : Taille de l'échantillon = **317 fermiers**

Formule pour les proportions, populations finies

Si l'on dispose du taux de variabilité, et de la taille de la population, on peut réduire substantiellement la taille minimum de l'échantillon obtenue par la formule précédente :

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}}$$



N = taille de la population

n₀ = taille de l'échantillon obtenue par la « formule pour les proportions » (ci-dessus)



**Cas pratique : Taille de l'échantillon**

La taille de la population est de 4000 fermiers, la taille de l'échantillon est de 288 fermiers. Avec les mêmes valeurs pour les critères statistiques que précédemment, nous obtenons la taille d'échantillon suivante :

$$n = (288 / (1 + ((288 - 1) / 4000))) = 288 / 0.0025 = \mathbf{269 \text{ fermiers}}$$

Nous ajoutons les 10% absorbant les pertes liées aux enquêtes de terrain : Taille de l'échantillon = **296 fermiers**

Cette formule permet un calcul plus optimal de la taille d'échantillon mais elle nécessite de disposer de plus d'informations avant l'enquête sur la population d'étude et son comportement.

Formule pour l'étude de la moyenne

Cette formule est similaire à celle utilisée pour les proportions, exceptée pour la mesure de la variabilité qui est remplacée par la variance (σ^2) de l'attribut de la population étudié. La bonne estimation de la variance de la population totale pour un attribut étant rarement disponible avant une enquête, on optera le plus souvent pour la formule pour les proportions.

$$n_o = \frac{Z^2 \times \sigma^2}{e^2}$$

e = niveau de précision

N = taille de la population d'étude

σ^2 = variance de l'attribut étudié

Z : valeur type associée au niveau de confiance requis pour la loi Normale (cf annexe II, n grand)

Cas pratique : Taille de l'échantillon

L'estimation de la variance de la variable indiquant si des bénéfices financiers de l'aide sont obtenus par les fermiers est de 0.24 ; Avec un niveau de précision à +/-5%, un niveau de confiance standard de 95%, nous obtenons la taille d'échantillon suivante : $n = ((1.96 \times 1.96) \times (0.24)) / (0.05 \times 0.05) = 0.9604 / 0.0025 = 369$ (+10% : **406 fermiers**)

N.B. : On remarquera que l'échantillon n'augmente pas proportionnelle à la taille de la population totale. Lorsque la population dépasse **100 000 individus**, la **taille minimale de l'échantillon atteint un seuil**, à critères statistiques égaux (ex : 400 si précision = +/-5%, niveau de confiance = 95% et variabilité = 0.5).

Contraintes de terrain

Parfois certaines contraintes empêchent de tirer le nombre d'individus optimum estimé par les formules précédentes pour faire partie de l'échantillon (ex : des inondations empêchent d'accéder à certains foyers de l'échantillon d'enquête; seulement 4 semaines sont prévues pour mener les opérations de terrain; seulement 3 enquêteurs sont mobilisés, ...). Les principales contraintes rencontrées lors d'enquêtes de terrain en sécurité alimentaire sont développer ci-après :

Accès

Les zones d'implantation des programme ACF sont souvent en conflit ou alors dévastées par des catastrophes naturelles. Il arrive donc que des **endroits** soient **enclavés** soit parce qu'il s'agit de lieux inaccessibles, pour cause d'inondations par exemple, soit pour des raisons de sécurité. Les individus de ces zones ne pourront donc pas faire partie de l'échantillon. Il est très important cependant de spécifier leur existence et les raisons de leur retrait de l'échantillon lors de la présentation de la méthodologie. S'il est possible de connaître les caractéristiques des populations concernées, il est important de les étudier à part (étude succincte) pour pouvoir faire bénéficier l'aide apportée à ces populations vulnérables au même titre que les autres. Si les caractéristiques des personnes de ces zones sont très différenciées des autres individus de l'échantillon, il convient alors d'adapter l'aide aux conditions particulières de la zone, et de le spécifier lors de l'inférence des analyses à la population totale.





Temps

Le temps imparti pour une enquête de terrain en sécurité alimentaire varie entre 24h et 6 semaines (mise en place, soumission des questionnaires, saisie, traitement, analyse et rédaction du rapport). Pour les enquêtes de suivi, le temps consacré aux opérations de terrain est plus grand (de 1 à 3 mois).

Nombre d'enquêteurs

Le nombre de personnel disponible à plein temps pour mener des enquêtes de terrain est parfois restreint. Souvent, le programme ACF dispose d'une ou deux personnes salariées et donc disponibles tout le temps; le reste des enquêteurs doivent être recrutés pour les besoins de l'enquête. Il faut donc bénéficier de financement pour ces dépenses supplémentaires dans le budget alloué à l'enquête. De plus, lorsqu'il s'agit d'un projet ayant déjà mené des enquêtes, le personnel est opérationnel rapidement. En revanche, lorsque la personne vient d'être recruté, il convient de le former avant de partir sur le terrain (cf module 3).

Nombre de questionnaires remplis

Le temps imparti, le nombre d'enquêteurs ainsi que leur expérience dans la mise en œuvre d'opérations de terrain sont des facteurs parfois limitant lorsqu'il s'agit de déterminer la taille de l'échantillon à enquêter.

Avant toute chose, il convient de déterminer le **nombre moyen de questionnaires qu'un enquêteur est en mesure de remplir en un jour (n_{Qj})** en tenant compte du temps de travail journalier (ex : 8h : la journée commence à 7h et se termine à 17h), de certaines contraintes contextuelles (il convient d'adapter les journées selon le contexte : ex : les heures de prière dans les communautés musulmanes le vendredi, les heures de travaux dans les champs si l'on enquête les fermiers, les heures de marché si l'on interview les personnes responsables des petits commerces dans le foyer, ...), et des différents critères suivants :

Temps nécessaire pour se rendre dans une zone d'enquête : ___ h ___ min

Temps moyen pour remplir un questionnaire : ___ h ___ min (ce temps sera déterminé lors du testing des instruments d'enquêtes ; Attention ! tenir compte des moments de présentation et de discussions avant et après la soumission du questionnaire)

Temps moyen pour se rendre au foyer suivant : ___ h ___ min

Temps pour la pause du déjeuner : ___ h ___ min (en général 1h de pause, peut être déplacé selon le contexte)

Temps pour rentrer sur le lieu d'hébergement de l'enquêteur : ___ h ___ min

Temps exceptionnel (prière, travaux dans les champs, heures de marché...): ___ h ___ min

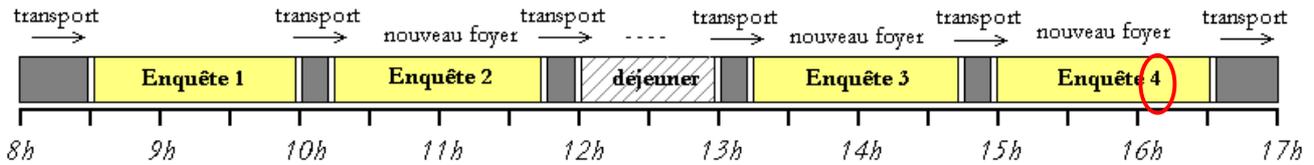
L'astuce est d'imbriquer toutes ces périodes (convertir en minutes) dans le chronogramme d'une journée « type » limitée par les heures de début et de fin de journée devant couvrir le nombre d'heures journalières imposées aux enquêteurs. Cette simulation permet de connaître le nombre de questionnaires qu'il est possible de remplir en une journée de travail. N.B. : sous l'emprise de la fatigue, les enquêteurs négligeront la qualité des données récoltées; il est donc inutile de vouloir trop charger une journée...

Cas pratique : Opérations de terrain

- temps de travail journalier : 8h (sans la pause déjeuner qui dure une heure);
- contrainte supplémentaire : prière le vendredi après-midi
- compter environ 30 minutes en moyenne pour se rendre dans un des villages A ou B à partir de la base ACF (A est plus loin que B, 30 min est une approximation pour les deux)
- une simulation sur le questionnaire à soumettre permet de dire qu'il faut environ 1 heure pour le remplir (incluant les salutations d'usage et la présentation de l'enquête auprès du foyer enquêté)
- 15 minutes sont nécessaires pour un enquêteur (en moyenne) pour aller d'un foyer à un autre

Toutes ces informations permettent de dresser le chronogramme d'une journée d'enquête « type » (cf schéma suivant) qui nous indique le nombre de questionnaires qu'un enquêteur peut remplir par jour qui est : $n_{Qj} = 4$ questionnaires





Une fois déterminé le nombre de questionnaires/jrs/enquêteur on calcule le temps disponible:

Nombre de jours disponibles pour la totalité de l'enquête : ____ jours

Nombre de jours consacrés à la saisie, au traitement, à l'analyse... : ____ jours

nombre de jours impartis pour les opérations de terrain (n_j) : _____ jours

(N.B. : souvent on connaît le nombre de semaines disponibles et $n_j = \text{nbre de jours de travail/semaine} \times \text{nbre de semaines}$)

Cas pratique : Opérations de terrain

- nombre de jours de travail/semaine : 5jrs : (les enquêteurs travailleront le samedi matin à la place du vendredi après midi)

- 7 semaines sont imparties pour la totalité de l'enquête dont :

* 4 semaines pour les opérations de terrain

* 3 pour la saisie, le traitement et l'analyse et la rédaction du rapport

Le nombre de jours d'enquêtes disponibles est $n_j = 5 \times 4 = 20$ jours

Enfin, selon le personnel salarié et les possibilités de recrutement pour les besoins des opérations de terrain, calculer le nombre d'enquêteurs disponibles pour les opérations de terrain (n_e) : _____

Cas pratique : Opérations de terrain

- 1 enquêteur expérimenté travaille à plein temps chez ACF, l'administrateur du projet va coordonner les opérations

- d'autres enquêteurs seront recrutés. Le nombre est à déterminer selon les besoins. Le budget prévoit au maximum 120 jours homme pour l'enquête. Les opérations de terrain sont prévues sur 4 semaines, et chaque semaine comporte 5 jours ouvrés, i.e. le nombre maximum d'enquêteurs que l'on peut recruter sur cette période compte tenu du budget alloué est de $120 / (5 \times 4) = 6$

Le nombre d'enquêteurs disponibles est donc $n_e = 1 + 6 = 7$ enquêteurs

Une fois que l'on a passé en revue tous ces aspects, il est possible de connaître la taille maximum de l'échantillon que nous seront en mesure de couvrir, sachant qu'un questionnaire rempli équivaut à un individu enquêté.

Calcul : taille maximale de l'échantillon que l'on peut prendre en charge

| | | |
|---|---|-----------------------------|
| Nombre d'enquêtes par jour/ enquêteur | X | n_{Qj} |
| Nombre de jours disponibles | X | n_j |
| Nombre d'enquêteurs disponibles | X | n_e |
| Taille maximale de l'échantillon | | n_{max} |

Cas pratique : Opérations de terrain

| | |
|---|------------|
| Nombre d'enquêtes par jour/ enquêteur | 4 |
| Nombre de jours disponibles | 20 |
| Nombre d'enquêteurs disponibles | 7 |
| Taille maximale de l'échantillon | 560 |

La taille maximale de notre échantillon est égale au nombre maximum de questionnaire que l'on peut remplir en 20 jours par 7 enquêteurs : $n_{max} = n_{Qj} \times n_j \times n_e = 4 \times 20 \times 7 = 560$





Autres considérations

La taille d'échantillon la plus fiable statistiquement parlant est celle obtenue par le biais des formules statistiques décrites précédemment. Voici quelques indications supplémentaires dans le cas où certaines contraintes empêchent d'allier la théorie et la pratique :

Enquêtes rapides

Il existe parfois des conditions d'enquête dites « extrêmes » qui contraignent les experts de terrain à agir au plus vite, par le biais d'enquêtes rapides auprès d'informateurs clés, ou de groupes de discussion (*ex : temps d'enquête extrêmement restreint (24h) ; zones d'enquête rendue inaccessible par un blocus militaire ; migration intensive de réfugiés malnutris dans un camps ;...*). En outre ce type d'enquête est souvent nécessaire pour récolter les quelques informations supplémentaires permettant la mise en œuvre d'enquêtes de grand envergure (*ex : détection des sous-groupes de la population d'étude auprès de différents groupes de discussion (personnes influentes, femmes,...) pour bien cibler les populations vulnérables*). Les **effectifs minimaux indicatifs** délivrés ci-après permettent de parer au plus urgent, tout en assurant une **fiabilité minimale des résultats** découlant de ce type d'enquêtes « rapides » :

- Le calcul de la taille de l'échantillon se base sur l'hypothèse de la normalité de la distribution des données recueillies. Ainsi, lorsque la **fréquence d'une variable se situe entre 20% et 80%** (caractéristique de la distribution normale), il est prouvé qu'un échantillon de taille comprise entre **30 et 200 éléments** peut être suffisant (Kish (1965)).
- Lorsque la **distribution est asymétrique**, et notamment lorsque la population d'étude est assez petite, il convient alors de récolter un nombre suffisamment grand de données, voire envisager d'enquêter la totalité de la population pour s'assurer d'une interprétation correcte des résultats (Kish (1965)).

Objectif de l'enquête

Selon que l'on veuille produire des **statistiques descriptives** ou que l'on se lance dans une analyse multivariée d'un phénomène, le nombre d'observations nécessaire est différent. Dans le premier cas, si l'on désire simplement calculer les indicateurs classiques (*ex : moyenne des revenus des foyers, fréquence des femmes qui allaitent leurs enfants de moins de 6 mois, ...*), il n'existe a priori **pas de taille minimale**.

En revanche, lorsque l'**analyse** envisagée est plus **poussée** (*ex : régression multivariée, analyse de la variance, i.e. explication d'une variable par plusieurs autres variables*), un échantillon minimal de **200-500 individus** garantit une interprétation valable des résultats.

Analyse comparative

Lorsque l'étude porte exclusivement sur **deux groupes distincts** (*ex : évaluation d'un projet selon que les individus participent ou non à un programme*), et quelque soit la proportion des uns et des autres dans la population totale (*ex : une population de 10 000 personnes dont 2 000 (20%) ont bénéficié d'une distribution de kits Seeds&Tools*), il est d'usage de mener l'enquête sur des effectifs égaux pour les deux groupes (*ex : 50% des participants et 50% des non-participants*).

De plus, un minimum de **100 observations** est requis par **groupe « majeur »** (Sudman(1976)).

Si, de plus, ces groupes « majeurs » sont composés de diverses strates (sous-groupes) faisant l'objet d'une étude spécifique, **20 à 50 observations par sous-groupes** sont nécessaires pour permettre d'exploiter les résultats au niveau de la population totale.

Tirage en grappes

Enfin, il existe une méthode d'échantillonnage particulière dite « en grappes » qui suscite de l'intérêt dans les enquêtes de Sécurité Alimentaire (cf p.41 pour plus de détails). Néanmoins cette technique implique une perte de précision de l'estimation. La taille optimale de l'échantillon est alors déterminée par le biais de tables spécifiques (Annexes III, p.46-47) qui fournissent la précision de l'estimation associée compte tenu de trois paramètres essentiels.





Plan d'échantillonnage complet

Effectif de l'échantillon par strates

Le plan d'échantillonnage (p.19), renseigné par strates au niveau de la population totale (effectifs et poids), peut désormais être complété au niveau de l'échantillon :

- En premier lieu, nous indiquons **l'effectif total de l'échantillon retenu** après les calculs statistiques et la prise en compte des contraintes de terrain et autres considérations.

Cas pratique : Taille de l'échantillon

- La taille optimale obtenue après les calculs est de 400 (via la formule simplifiée car nous ne disposons pas du degré de variabilité de la population d'étude pour les aspects de revenus et bénéfices qui font l'objet principal de notre étude)

- L'étude des contraintes et autres considérations ne nous obligent pas à réduire cet effectif.

Notre échantillon contiendra donc **400 fermiers**. On inscrit donc cette valeur dans la case verte, qui représente l'effectif total de l'échantillon de l'enquête.

- Ensuite, **l'effectif affecté à chacune des strates** définies dans le plan d'échantillonnage s'obtient en multipliant le poids associée à cette strate par l'effectif total de l'échantillon :

$$n_A = n \times P_A$$

n = effectif totale de l'échantillon

n_A = effectif de l'échantillon pour la strate A

P_A = poids de la strate A dans la population totale

Cas pratique : Plan d'échantillonnage complet

Cas pratique :ex : Strate 4

Le poids de la strate est de 6% dans la population totale ; pour lui affecter le même poids dans l'échantillon on multiplie le poids de la strate par l'effectif total de l'échantillon : $6\% \times 400 = 23$.

Ainsi l'échantillon d'enquête comptera **23 fermiers de longue date, n'ayant plus de bétail (Strate 4)**

| NomStrate | LibStrate | NbBenef | PoidsStrate (%) | NbEchant |
|-----------|--|---------|-----------------|----------|
| 1 | Fermiers de longue date , ayant 1 à 3 bêtes , dont les récoltes permettent une autonomie alimentaire de plus de 6 mois et sont génératrices de revenus supplémentaires | 984 | 25% | 98 |
| 2 | Fermiers de longue date , ayant 1 à 3 bêtes , dont les récoltes permettent une autonomie alimentaire de plus de 6 mois et mais sans revenus supplémentaires | 197 | 5% | 20 |
| 3 | Fermiers de longue date , ayant 1 à 3 bêtes , dont les récoltes ne permettent pas une autonomie alimentaire suffisante | 229 | 6% | 23 |
| 4 | Fermiers de longue date , n'ayant plus de bétail | 230 | 6% | = 23 |
| 5 | Fermiers anciennement réfugiés , ayant 1 à 3 bêtes | 682 | 17% | 68 |
| 6 | Fermiers anciennement réfugiés , n'ayant plus de bétail , dont les récoltes permettent une autonomie alimentaire de plus de 6 mois | 372 | 9% | × 37 |
| 7 | Fermiers anciennement réfugiés , n'ayant plus de bétail , récoltes ne permettent pas une autonomie alimentaire suffisante | 186 | 5% | 19 |
| 8 | Fermiers pasteurs destitués , ayant 1 à 3 bêtes , dont les récoltes permettent une autonomie alimentaire de plus de 6 mois | 336 | 8% | 34 |
| 9 | Fermiers pasteurs destitués , ayant 1 à 3 bêtes , dont les récoltes ne permettent pas une autonomie alimentaire suffisante | 560 | 14% | 56 |
| 10 | Fermiers de pasteurs destitués , n'ayant plus de bétail | 224 | 6% | 22 |
| | | 4000 | 100% | 400 |





Calcul du niveau de précision

Lorsque certaines contraintes empêchent de tirer le nombre d'individus optimum estimé par les formules mathématiques pour faire partie de l'échantillon (*ex : une contrainte de temps (4 semaines) nous empêche d'enquêter le nombre requis (700 foyers) ;*). Inversement, il peut arriver que la taille de l'échantillon soit plus grande que la taille suffisante requise (*ex : la taille de l'échantillon a été fixée de façon arbitraire, ou le budget permettait d'enquêter plus de foyers que prévu*). Dans les deux cas, la baisse ou l'augmentation de l'effectif prévu va influencer sur la précision de l'estimation et donc l'intervalle de confiance.

Pour obtenir le niveau de précision à partir de la taille de l'échantillon, on inverse les formules mathématiques utilisées pour calculer cette taille :

Formule simplifiée

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



$$e = \sqrt{\frac{N - n}{N \times n}}$$

Formule pour les proportions

$$n = \frac{t^2 \times p \times q}{e^2} = \frac{t^2 \times p \times (1 - p)}{e^2}$$



$$e = \sqrt{\frac{t^2 \times p \times (1 - p)}{n}}$$

Formule pour l'étude de la moyenne

$$n_o = \frac{Z^2 \times \sigma^2}{e^2}$$



$$e = \sqrt{\frac{Z^2 \times \sigma^2}{n}}$$

Cas pratique : *ex : Niveau de précision*

Formule simplifiée avec la taille d'échantillon retenue (400 fermiers)

$$e = \sqrt{(4000 - 400) / (4000 \times 400)} = 0.05 = \pm 5\%$$

N.B. : il s'agit bien du niveau de précision indiqué lors du calcul de la taille de l'échantillon

Formule simplifiée avec une taille d'échantillon réduite (ex : un manque de temps nous oblige à n'enquêter que 50 fermiers)

$$e = \sqrt{(4000 - 50) / (4000 \times 50)} = 0.14 = \pm 14\%$$

Formule simplifiée avec le nombre d'enquêtes maximale qu'il est possible de mener (560 fermiers) – cf p. 26

$$e = \sqrt{(4000 - 560) / (4000 \times 560)} = 0.039 = \pm 3.9\%$$

N.B. : si l'on est contraint de réduire la taille de l'échantillon prévu, on perd en précision ; si l'enquête est finalement menée auprès des 560 fermiers, on améliorera sensiblement la précision des résultats obtenus (±3.9% contre ±5%)

Cf p.20 pour les répercussions sur les intervalles de confiance

N.B. : Le fichier Excel « OutilEch » (onglet « TailleEchant ») vous fournit un outil pratique permettant de calculer les différentes tailles d'échantillon selon les critères statistiques désirés, ainsi que de calculer la précision correspondant à une taille d'échantillon spécifique.



Une fois déterminée la taille de l'échantillon, et le nombre d'individus à enquêter par strates, il reste à choisir la méthode de tirage de l'échantillon à adoptée.





4^{ème} étape :

LA METHODE D'ECHANTILLONNAGE

Pour s'assurer de la représentation de la diversité des individus faisant partie de l'échantillon il convient d'utiliser des techniques adéquates (on s'intéresse à des personnes, et non des articles sortis de chaînes de productions automatisées...). La méthode adoptée peut être aléatoire ou non, sachant que :

Dans les méthodes aléatoires, chaque unité (foyer, fermier, enfant...) a une chance que l'on peut quantifier d'être sélectionnée et d'appartenir à l'échantillon. **Dans l'échantillonnage non probabiliste**, on suppose que la distribution des caractéristiques à l'intérieur de la population est égale.

Tirage non aléatoire

Dans l'échantillonnage non aléatoire, le concept de « chance égale » est absent, raison pour laquelle sa « fiabilité » est parfois remise en question. Ne disposant d'aucune information chiffrée sur la population d'étude, il n'existe aucun moyen de mesurer la précision des échantillons retenus.

Ce type de tirage est souvent utilisé dans les sondages effectués en Sécurité Alimentaire lors d'**études exploratoires**, notamment lors des zonages et typologies, pour récolter des informations essentiellement qualitatives. Ces **enquêtes « rapides »** vont permettre aux responsables du programme de cibler les populations vulnérables selon l'objectif du projet mené en dégagant des groupes de population distinctes, selon les variables d'intérêt retenues. Il est parfois aussi utile lors de la phase **d'évaluation du projet**, pour en tirer les enseignements auprès de populations spécifiques (que l'aide ait été efficace ou non).

Les échantillons d'enquêtes se composent d'**informateurs clés ou de groupes de personnes**. La récolte des données vise plus l'information en profondeur plutôt qu'en largeur. Elle permet de juger de l'impact de l'intervention auprès des principaux acteurs et de rectifier l'aide si les premiers résultats sont mauvais (*ex : les semences des kits Seeds&Tools ont-elles bien été plantées ? (ou ont-elles été revendues, échangées ou stockées ?) ; les apports en vitamines sont-ils bien administrés ? (doses respectées, fréquences nécessaires, ... ?) ; le système d'irrigation construit est-il bien entretenu ?, les enfants auxquels on a administré du lait sont-ils mieux portants dans l'ensemble ?, ...*). Il se peut aussi que l'on désire avoir une perspective particulière en interviewant certaines personnes ou groupes (*ex : on veut s'inspirer de l'expérience des foyers ayant déjà bénéficié d'une distribution de semences et qui ont tiré profit, ou les mécanismes d'adaptation des femmes célibataires faisant du petit commerce qui s'en sortent généralement mieux face à une vague de sécheresse*). Il est donc très important de s'assurer que l'on **balaye bien toutes la diversité des situations** présentes au sein de la population malgré le peu de référents.

Les avantages principaux d'utiliser ces méthodes sont les suivantes:

- ❶ Il n'est pas nécessaire d'effectuer un tirage aléatoire à partir d'une base de sondage sous la forme d'une liste (*ex : ménages résidant dans un village, fermes dévastées par l'inondation, individus résidant dans le camps de réfugiés, enfants malnutris fréquentant le centre de santé, ...*).
- ❷ On choisit arbitrairement des individus qui seront inclus dans l'échantillon
- ❸ Elles réduisent substantiellement la mobilisation de moyens temporels, humains et financiers.

Ces méthodes sont utiles lorsque les contraintes de terrain empêchent de faire autrement. Si une méthode aléatoire est possible, elle est à privilégier.





Echantillonnage systématique

Avantages de cette méthode

- Ne nécessite pas de plan de sondage
- Technique de sélection simple et rapide

Limites

- Permet simplement une vision globale des caractéristiques d'une population
- Ne permet pas d'aboutir à des conclusions sur l'impact d'un phénomène sur la population étudiée
- Ne permet pas d'estimer l'erreur d'échantillonnage

Principe

L'échantillonnage systématique est une variante de la méthode aléatoire simple, dans laquelle vous sélectionnez un échantillon à intervalles prédéterminés (*ex : on sélectionne un individu sur 5 dans un assemblée placée par rangs, on enquête les foyers toutes les trois maisons à partir du centre du village*) mais on ne le considère pas comme un échantillon aléatoire pur puisqu'il comporte un élément de prédétermination et que l'on ne dispose pas d'une liste complète de la population d'étude. Cette méthode est très pratique lorsque nous ne connaissons pas la composition de l'échantillon ; le tirage s'effectuera alors sur place selon une sélection systématique des enquêtés.

Comment procéder ?

La population ciblée est localisée dans un espace géographique restreint (*ex : village, quartier*) et les lieux d'habitation (foyers) ou les individus sont bien ordonnés de façon à ce que l'on puisse bien se repérer :

- 1- **Calculer le nombre de foyers** (ou d'individus) dans le village (ou dans l'assemblée). Si nous ne pouvons compter ou ne disposons pas du nombre de foyers du village, nous pouvons en obtenir le nombre approximatif à partir du nombre de population du village et du nombre moyen de personnes par familles qui est habituellement de 5 dans les pays où opère ACF (*ex : population totale=15000 -> 15000/5=environ 3000 foyers dans le village*).
- 2- **Déterminer la taille de l'échantillon** (cf étape 3 du manuel)
- 3- **Déterminer le pas de sondage** (cf méthode échantillonnage aléatoire simple pour plus de détails) :

$$\text{"pas"} = N/n$$

N= taille de la population d'étude
n= taille de l'échantillon

N.B. : Lorsque le chiffre obtenu n'est pas juste on l'arrondi pour obtenir un chiffre rond.
(*ex : pour 3000 foyers et une taille d'échantillon égale à 500, on aura un pas=3000/500=6*)

4- Déterminer la première maison à enquêter

Point de départ de l'enquête : en règle générale, on se positionne au centre du village pour rayonner ensuite autour de ce point. Il faut néanmoins s'assurer que ce positionnement prendra en compte les foyers les plus éloignés, car un foyer basé au centre du village présente des spécificités par rapport à un foyer très éloigné (accès aux boutiques, accès à l'eau à l'électricité, etc...)

Méthode

- On se positionne au centre du village et on fait tourner une bouteille ou un crayon. La direction indiquée au repos désigne la direction de départ de l'enquête.
- Ensuite, on choisit au hasard un nombre entre 1 et 6. Si le chiffre 4 sort, on commencera l'enquête à partir de la quatrième maison (personne) par rapport au point de départ de l'enquête.
- La seconde maison (personne) enquêtée sera la 4+6=10è, puis 10+6=16è... lors de l'enquête, si un foyer ne présente pas d'individus ciblés (*ex : enfant de moins de 10 ans*), on passe au prochain, etc...

Cas pratique : *Le pas ici est de 4000/400=10. Si une représentation géographique de nos fermes bénéficiaires est faisable, on peut donc effectuer l'enquête toutes les 10 fermes à partir de la première choisie par la méthode décrite ci-dessus.*





Echantillonnage aléatoire raisonné (informateurs clés, groupes, volontaires, ...)

Avantages de cette méthode

- Ne nécessite pas de plan de sondage
- Permet de s'informer indirectement sur des populations inaccessibles par le biais d'informateurs

Limites

- Permet simplement une vision globale des caractéristiques d'une population
- Ne permet pas d'aboutir à des conclusions sur l'impact d'un phénomène sur la population étudiée
- Ne permet pas d'estimer l'erreur d'échantillonnage

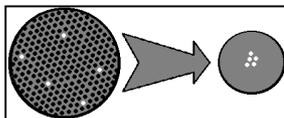
Principe

L'échantillon est établi selon une ou plusieurs caractéristiques fixées à l'avance lors de la phase exploratoire du projet. L'objectif est d'acquérir une connaissance approfondie du contexte sur un sujet donné et détecter des indicateurs pertinents pour définir les populations vulnérables, par le biais d'entretiens semi-directifs qui permet aux personnes de l'échantillon d'exprimer leur point de vue en guidant la logique et l'objet de la discussion (par opposition à des questionnaires « quantitatifs » où les questions sont fermées).

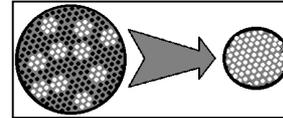
- Il peut s'agir d'**informateurs clés**, très peu nombreux, mais assez proche de la population étudiée pour être au fait des habitudes de vie, de l'organisation alimentaire et économique des foyers et de leurs mécanismes d'adaptation face aux crises (*ex : un instituteur, le maire du village, un « facilitateur », ...*).

- Il peut aussi s'agir de **groupes de 5 à 8 personnes** homogènes (*ex : les notables d'un village, les femmes, les fermiers, les vendeurs du marché,...*) auprès desquelles l'on mène des enquêtes séparées. La complémentarité des différents acteurs permet de s'assurer que l'on détient toute l'information nécessaire (*ex : les hommes nous informeront sur les questions agricoles, les femmes sur les habitudes alimentaires ; les interviewer séparément permet à chaque « communauté » de s'exprimer librement sur les sujets qu'elle maîtrise*).

Informateurs clés



Groupes de discussion



Comment procéder ?

❶ Établissez une liste des personnes clés qui pourraient éventuellement répondre à la question précise à laquelle vous voulez répondre. Il pourra s'agir des personnes suivantes:

- Des experts qualifiés actifs dans la zone de projet (médecins, infirmiers, instituteurs, économistes, spécialistes du crédit, agronomes, membre d'une ONG locale, ...)
- Des dirigeants locaux (chef du village, directeur d'école, autres notables, ...)
- Des personnes bien informées, présentant les mêmes caractéristiques que celles des populations ciblées (commerçants, vendeurs au marché, fermiers, chef de foyers, homme, femmes, personnes âgées, personnes étant considérées comme un cas « extrême », ...)

❷ Sélectionnez ensuite celles de ces personnes qui sont le plus à même de répondre à la question posée. Préférez les personnes volontaires, plus disponibles, lorsque leur intérêt rejoint les objectifs du projet. Ajoutez d'autres personnes si de nouvelles idées apparaissent au cours des interviews.

Cas pratique :

- on s'entretient avec chacun des deux chefs des villages A et B pour avoir une connaissance globale de la taille des cheptels, du niveau d'activité agricole et des capacités d'adaptation des fermiers

- une enquête semi-dirigée avec un groupe de fermiers représentant chaque catégorie détectée (ceux de longue date, les anciens réfugiés et les pasteurs destitués) pour évaluer les effectifs des cheptels, les pertes causées par la sécheresse, et le niveau de dettes accusées pour faire face à la crise

- une enquête auprès des femmes représentant chaque catégorie, permettra de connaître les habitudes alimentaires, avant et après la sécheresse, d'évaluer la quantité de récoltes de sorgho, et les bénéfices liés à la vente des récoltes





Echantillonnage cumulatif (méthode des quotas, snowball,...)

Avantages de cette méthode

- Ne nécessite pas de plan de sondage
- Méthode peu coûteuse et rapide
- Permet de contrôler les proportions de la population
- Permet une description détaillée d'une population particulière établie selon le « voisinage »
- Utile pour établir des comparaisons (choix des individus ne présentant pas la caractéristique étudiée)

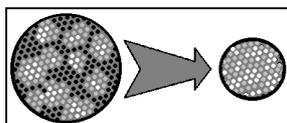
Limites

- Impossibilité d'estimer l'erreur d'échantillonnage (on sélectionne arbitrairement = pas de hasard)
- Ne permet pas d'aboutir à des conclusions sur l'impact d'un phénomène sur la population étudiée

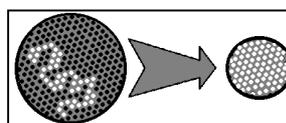
Principe

L'échantillonnage cumulatif s'effectue jusqu'à ce qu'un nombre précis d'unités pour diverses sous-populations ait été sélectionné. En règle général, on interroge les premières personnes rencontrées présentant les caractéristiques étudiées. La notion de hasard est inexistante, et selon les conditions d'enquête (*ex : localisation géographique, moment de la journée, époque de l'année,...*) la représentativité de l'échantillon est remise en cause (*ex : on a interviewé uniquement des foyers étant proches d'un puits : leurs habitudes concernant l'utilisation de l'eau sera spécifique et différente de celles des habitants plus éloignés ; on interview des foyers pendant les travaux des champs : toute une partie de la population concernée sera éludée dans l'enquête ; on interroge des populations sur leurs habitudes alimentaires durant une période de fête : l'information sera biaisée car plus de nourriture est disponible pour ce genre événement particulier, ...*). Ainsi de telles enquêtes nécessitent une bonne connaissance globale de la population pour éviter d'omettre des caractéristiques majeures dans l'échantillon choisi.

Méthode « boule de neige »



Méthodes des quotas



Comment procéder ?

La **méthode « boule de neige »** permet, à partir d'un ou quelques individus présentant les caractéristiques étudiées, de compléter l'échantillon soit par bouche à oreille (*ex : une femme responsable d'un micro-crédit va nous diriger vers une autre femme qu'elle connaît qui est impliquée dans une activité similaire, et ainsi de suite jusqu'à obtenir le nombre d'individus requis*), soit en interviewant toutes les personnes présentant ces caractéristiques au hasard des rencontres (*ex : on se rend dans un village et on passe de maison en maison en enquêtant uniquement les familles de plus de 6 enfants ; on enquête les 50 premiers patients rencontrés dans le centre de santé ; on vérifie les rations de nourriture reçues par les 200 premiers bénéficiaires d'une distribution alimentaire, ...*)

La **méthode des quotas**, quant à elle, construit l'échantillon en respectant la distribution dans la population d'étude en fonction des caractéristiques étudiées (*ex : secteur géographique, catégorie d'âge, sexe, ...*). Une fois la taille de l'échantillon arrêtée, les enquêteurs sont chargés de mener les enquêtes en respectant les proportions (quotas) dans les divers groupes établis de manière à « être à l'échelle » de la population (*ex : un camps de réfugié compte 60% d'hommes et 40% de femmes. L'étude vise connaître les capacités d'adaptation selon le sexe. L'échantillon tiendra compte de cette répartition*).

N.B. : Cette méthode est utile pour effectuer des comparaison entre deux groupes (*ex : bénéficiaires/non bénéficiaires. On sélectionne un échantillon représentatif des bénéficiaires et on sélectionne, par la méthode des quotas, un échantillon témoin de non bénéficiaires de même effectif que le premier*).

Cas pratique : *Si des contraintes nous empêchaient de déployer une enquête détaillée, nous pourrions enquêter les fermiers selon leur catégorie (de longue date, anciens réfugiés, pasteurs destitués) et leur répartition connue (41%, 31%, 28%). Ainsi pour une taille d'échantillon de 60 fermiers, on enquêterait 25 fermiers de longue date, 18 anciens réfugiés et 17 pasteurs destitués. Le risque avec ces enquêtes est de passer à côté d'informations essentielles pour comprendre la situation globale.*





Echantillon aléatoire

L'échantillonnage aléatoire repose sur la **sélection au hasard d'individus d'une population d'étude** pour faire partie d'un échantillon. Il s'agit d'obtenir, sans aucune connaissance préalable ou prise en compte de caractéristiques particulières, un échantillon pouvant être considéré représentatif des principaux acteurs touchés par l'intervention d'un projet. Cette spécificité permet d'utiliser des critères statistiques qui estimeront le niveau de précision de l'échantillon retenu.

Ce type d'échantillonnage est plus complexe, prend plus de temps et est souvent plus coûteux que l'échantillonnage non aléatoire. Toutefois, il est à privilégier car, la propriété aléatoire du tirage confère à **chaque individus la même chance (ou du moins une chance connue) de faire partie de l'échantillon**, que l'on peut quantifier. Cette probabilité permet alors de produire des estimations de l'erreur d'échantillonnage qui nous indique la **fiabilité des résultats** issus de l'enquête menée avant même qu'elle soit menée (*ex : un échantillon de 50 personnes fournit un degré de précision de +/-14%. Cela permet de calculer l'intervalle de confiance associé. Si la précision ne semble pas être assez fine pour détecter des différences entre deux groupes d'individus (dont on connaît globalement le comportement, les approximations des variables mesurées), on va augmenter la taille de l'échantillon pour s'assurer d'une interprétation fiable des résultats issus de l'enquête).*

Ce type de tirage est souvent utilisé dans les enquêtes effectués en Sécurité Alimentaire lors de **l'identification des besoins** (*ex : définition de l'aide mise en place auprès des populations vulnérables*), lors du **suivi de l'aide** apportée (*ex : on quantifie les récoltes issues d'une distribution de kits seeds&Tools auprès d'un échantillon de bénéficiaires*), et parfois lors de **l'évaluation du projet** finalisé (*ex : on recueille à nouveau les quantités de récoltes pour juger de l'évolution de l'impact de l'aide kit Seeds&Tools et pour connaître les nouveaux besoins des bénéficiaire, s'ils existent*). Ils visent à recueillir des données qualitatives et quantitatives. Ces **enquêtes de grande envergure**, permettent de recueillir des informations précises et nombreuses sur la population ciblée. Leur analyse, par le biais de techniques statistiques, aboutira à des conclusions chiffrées et fiables sur la population d'étude et l'impact de phénomène sur cette dernière.

En définitive ce type d'échantillonnage garantit la qualité de la généralisation des résultats obtenus pour un échantillon à l'ensemble de la population d'étude (inférence statistique). Il existe plusieurs méthodes différentes permettant de sélectionner un échantillon probabiliste. Le choix dépend de plusieurs facteurs (existence ou non et précision d'une liste des individus, détection ou non de sous-groupes homogènes d'individus, coût de l'enquête, type et capacité d'analyse statistique).

Tirage pour les méthodes aléatoires

- dresser une liste comprenant la totalité des unités de la population étudiée (tous les villages d'une région, tous les foyers d'un village, tous les enfants fréquentant le centre de santé, ...)
- numéroter chacune des unités de 1 à n (n=taille de la population d'étude)
- choisissez la méthode d'échantillonnage adéquate (cf description des différentes méthodes)
- choisissez la technique de tirage : manuelle ou par ordinateur selon les dispositions

Les avantages principaux d'utiliser ces méthodes sont les suivantes:

- Le caractère aléatoire minimise le risque de non représentativité de l'échantillon
- On peut anticiper sur le degré de précision de l'échantillon obtenu et ainsi éviter une enquête inutile
- Elles permettent la comparaison d'études similaires dans le temps.

N.B. : Le fichier Excel « OutilEch » (onglet « TirageAleat ») simule deux tirages aléatoires sur la base de données de nos 4000 fermiers bénéficiaires (base de données dans l'onglet « BdDFarmers »). Le tableau récapitulatif permet de juger de la représentativité de l'échantillon selon les deux méthodes par rapport à l'échantillon prévu dans le plan d'échantillonnage.





Echantillonnage aléatoire simple

Avantages de cette méthode

- Chaque individu a une chance égale de faire partie de l'échantillon (représentativité)
- permet d'aboutir à des conclusions sur l'impact d'un phénomène sur la population étudiée

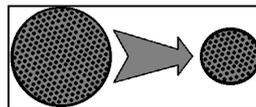
Limites

- Nécessite de disposer d'une liste de tous les individus de la population
- Peut parfois omettre certains groupes de personnes surtout lorsque l'échantillon est petit

Principe

Lorsqu'il n'existe pas de différences majeures entre les individus de la population bénéficiaire (*ex : même situation géographique, même niveau de vie, même surface cultivable, même taille du foyer, même accès à l'eau, même taille du bétail...*) et si l'aide déployée est la même pour tous, il est alors suffisant de choisir au hasard les individus pour faire partie de l'échantillon. Le tirage aléatoire simple consiste à choisir des individus de telle sorte que **chaque membre de la population a une chance égale** de figurer dans l'échantillon (*ex : parmi une population 500 enfants on en choisit 50 au hasard à partir d'une liste comportant les noms des 500 enfants*).

Tirage aléatoire simple



Comment procéder ?

A partir de la liste de la totalité des individus (ou unités) de la population étudiée, nous tirons au sort le nombre requis d'individus pour faire partie de l'échantillon. Le tirage au sort (aléatoire) peut se faire selon plusieurs techniques décrites à la suite :

- l'utilisation d'un tableau de nombres aléatoires (❶ p.35)
- la technique des petits papiers (chapeau) (❷ p.36)
- la technique des pas (❸ p.36)
- la fonction ALEA sous Excel (❹ p.37)

Cas pratique : *ici nous avons une liste de 4000 individus, nous utilisons donc la méthode de tirage par le biais de la formule ALEA d'Excel (cf simulation (SimulEAS) sur la base de données de notre population qui se trouve dans le fichier Excel nommé « Outils Echant » dans l'onglet « BdDFarmers »)*

❶ l'utilisation d'un tableau de nombres aléatoires

Exemple : On veut tirer aléatoirement 6 individus parmi 600 à l'aide de la table suivante :

TABLE DE KENDALL ET BABINGTON SMITH

| | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 5 | 13407 | 62899 | 78937 | 90525 | 25033 | 56358 | 78902 | 47008 | 72488 | 57949 |
| | 50230 | 63237 | 94083 | 93634 | 71652 | 02656 | 57532 | 60307 | 91619 | 48916 |
| | 84980 | 62458 | 09703 | 78397 | 66179 | 46982 | 67619 | 39254 | 90763 | 74056 |
| | 22116 | 33646 | 17545 | 31321 | 65772 | 86506 | 09811 | 82848 | 92211 | 51178 |
| | 68645 | 15068 | 56898 | 84021 | 40115 | 27524 | 42221 | 88293 | 67592 | 06430 |
| 10 | 26518 | 39122 | 96561 | 56004 | 50260 | 68648 | 85596 | 83979 | 09041 | 62350 |
| | 36493 | 41666 | 27871 | 71329 | 69212 | 57932 | 65281 | 57233 | 07732 | 58439 |
| | 77402 | 12994 | 59892 | 85581 | 70823 | 53338 | 34405 | 67080 | 16568 | 00854 |
| | 83679 | 97154 | 40341 | 84741 | 08967 | 73287 | 94952 | 59008 | 95774 | 44927 |
| | 71802 | 39356 | 02981 | 89107 | 79788 | 51330 | 37129 | 31898 | 34011 | 43304 |





Méthode 1

- 1- on numérote nos 100 individus de 1 à 600
- 2- on décide de commencer le tirage à la ligne 10 de la colonne 6 (chiffre 3 marqué d'un ❶)
- 3- on va de la gauche vers la droite et on prend les chiffres pas paquets de 3 (393, 560, 298, 189, 107, 797, 885, 133)
- 4- on ne retiens que les 6 premiers chiffres inférieurs à 600 ((393, 560, 298, 189, 107, 133))
- 5- s'il y a des doublons on tire un numéro supplémentaire
- 6- les individus dont les numéros correspondent constituent l'échantillon

Méthode 2

- 1- on numérote nos 100 individus de 1 à 600
- 2- on décide de commencer le tirage à la ligne 1 de la colonne 31 (chiffre 3 marqué d'un ❷)
- 3- on va de la gauche vers la droite et on prend les chiffres pas paquets de 5 (78902, 47008, 72488, 57949, 50230, 63237, 94083, 93634, 71652, 02656, 57532, 60307, 91619, 48916)
- 4- on ne retiens que les 3 premiers chiffres de ces nombres (789, 470, 724, 579, 502, 632, 940, 936, 716, 026, 575, 603, 916, 489)
- 5- on retient seulement les premiers chiffres qui commencent par un chiffre inférieur à 600 (470, 579, 502, 026, 575, 489)
- 7- s'il y a des doublons on tire un numéro supplémentaire
- 6- les individus dont les numéros correspondent constituent l'échantillon

Méthode 3

- 1- on numérote nos 100 individus de 1 à 600
- 2- on envisage des déplacements particuliers (cf suite débutant au chiffre 1 marqué du numéro ❸)

TABLE DE KENDALL ET BABINGTON SMITH

| | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 13407 | 62899 | 78937 | 90525 | 25033 | 56358 | 78902 | 47008 | 72488 | 57949 |
| | 50230 | 63237 | 94083 | 93634 | 71652 | 02656 | 57532 | 60307 | 91619 | 48916 |
| | 84980 | 62898 | 09703 | 78397 | 66179 | 46982 | 67619 | 39254 | 90763 | 74056 |
| | 22116 | 33646 | 17545 | 31321 | 65772 | 86506 | 09811 | 82848 | 92211 | 51178 |
| 5 | 68645 | 15068 | 56898 | 84021 | 40115 | 27524 | 42221 | 88293 | 67592 | 06430 |
| | 26518 | 39122 | 96561 | 56004 | 50260 | 68648 | 85596 | 83979 | 09041 | 62350 |
| | 36493 | 41666 | 27871 | 71329 | 69212 | 57932 | 65281 | 57233 | 07732 | 58439 |
| | 77402 | 12994 | 59892 | 85581 | 70823 | 53338 | 34405 | 67080 | 16568 | 00854 |
| | 83679 | 97154 | 40341 | 84741 | 08967 | 73287 | 94952 | 59008 | 95774 | 44927 |
| 10 | 71802 | 39356 | 02981 | 89107 | 79788 | 51330 | 37129 | 31898 | 34011 | 43304 |

❷ la technique des p'tits papiers (ou du chapeau)

Cette technique s'inspire directement du jeu de loto. Elle est très utile lorsque la liste des personnes n'est pas encore dressée.



Exemple : On veut tirer aléatoirement 15 individus parmi une assemblée de 100 personnes.



**Méthode :**

- 1- on numérote 100 petits bout de papier de 1 à 100
- 2- on les distribue aux 100 personnes de l'assemblée
- 3- on numérote à nouveau 100 petits bout de papier de 1 à 100
- 4- on les dépose dans un chapeau et on mélange bien
- 5- on retire un premier papier
- 6- on annonce le numéro à l'assemblée
- 7- la personne ayant le numéro correspondant se fait connaître
- 8- cette personne faisant partie de l'échantillon
- 9- un réitère le processus (tirage d'un papier, annonce, ...) jusqu'à ce que l'on obtienne 15 personnes

N.B. : penser à numéroté les absents s'ils font partie de la population d'étude.

③ la technique des pas

Cette technique s'effectue à l'aide de la liste de la population d'étude.

| Numéro | Prénom |
|--------|----------|
| 1 | Edouard |
| 2 | Marie |
| 3 | Eliane |
| 4 | Marc |
| 5 | Anne |
| 6 | Carole |
| 7 | Julien |
| 8 | Régis |
| 9 | Mathilde |



Exemple : On veut tirer aléatoirement 15 individus parmi une assemblée de 100 personnes (ou 15 foyers parmi 100).

Méthode :

- 1- on prend la liste des individus présents numérotés dans l'ordre de 1 à 100 (s'il s'agit de foyers on s'assure qu'on dispose de leurs adresses pour être en mesure de les retrouver lors de l'enquête)
- 2- on compte le nombre total d'individus (ici 100)
- 3- pour obtenir le « pas », on le divise par la taille de l'échantillon retenue (ici 15)
- 4- le « pas » ici est donc 7 (on arrondit), i.e. on tirera au sort un individu sur 7
- 5- pour choisir le premier individu tiré, on choisi au hasard un nombre entre 1 et 7, soit par la technique des p'tits papiers, soit en demandant à quelqu'un de l'assemblée de vous donner un chiffre entre 1 et 7. (ex : 2).
- 6- on entoure l'individu de la liste qui a le numéro 2, puis on saute 7 individus dans la liste pour choisir le prochain qui fera partie de l'échantillon (qui sera le numéro 9) , puis on en saute 7 autres pour trouver le troisième (numéro 16), etc...
- 7- on réitère le processus jusqu'à ce qu'on atteigne la fin de la liste (on peut aussi déterminer les numéros des individus de l'échantillon en additionnant la valeur du pas au numéro du premier individu choisi (ici $2+7=9$). Cela nous donne le numéro du second individu de l'échantillon. Puis on ajoute encore 7 à ce dernier numéro (ici $9+7=16$), et ainsi de suite jusqu'à 100)
- 8- on a ainsi la liste des individus faisant partie de l'échantillon et leur numéro associé (2, 9, 16, 23, 30, 37, 44, 51, 58, 65, 72, 79, 86, 93, 100)





4 la fonction ALEA sous Excel

Cette technique permet de se servir d'Excel pour tirer des nombres aléatoires.



| NumEch | NumAléa | NumEch | NumAléa | NumEch | NumAléa |
|--------|---------------------------|--------|---------|--------|---------|
| 1 | =ALEA.ENTRE.BORNES(1;100) | 1 | 34 | 1 | 4 |
| 2 | | 2 | | 2 | 91 |
| 3 | | 3 | | 3 | 57 |
| 4 | | 4 | | 4 | 3 |
| 5 | | 5 | | 5 | 100 |
| 6 | | 6 | | 6 | 28 |
| 7 | | 7 | | 7 | 55 |
| 8 | | 8 | | 8 | 9 |
| 9 | | 9 | | 9 | 29 |
| 10 | | 10 | | 10 | 95 |
| 11 | | 11 | | 11 | 20 |
| 12 | | 12 | | 12 | 24 |
| 13 | | 13 | | 13 | 91 |
| 14 | | 14 | | 14 | 10 |
| 15 | | 15 | | 15 | 11 |
| 16 | | 16 | | 16 | 53 |
| 17 | | 17 | | 17 | 23 |
| 18 | | 18 | | 18 | 16 |
| 19 | | 19 | | 19 | 86 |
| 20 | | 20 | | 20 | 98 |

| NumEch | NumAléa |
|--------|---------|
| 1 | 4 |
| 2 | 91 |
| 3 | 57 |
| 4 | 3 |
| 5 | 100 |
| 6 | 28 |
| 7 | 55 |
| 8 | 9 |
| 9 | 29 |
| 10 | 95 |
| 11 | 20 |
| 12 | 24 |
| 13 | 10 |
| 14 | 11 |
| 15 | 53 |

Exemple : On veut tirer aléatoirement 15 individus parmi 100 personnes.

Méthode :

- 1- on ouvre un fichier Excel
- 2- on numérote la première colonne de 1 au chiffre correspondant à la taille de l'échantillon(ici 15).
On ajoute quelques numéros supplémentaires dans le cas où il y ait des numéros redondants (ici on numérote jusqu'à 20).

N.B. : Plus la taille de l'échantillon est grande plus il faut prévoir de numéro complémentaires.

- 3- on se positionne sur la cellule de la colonne suivante qui est face au numéro 1
- 4- on y entre la fonction suivante :

= ALEA.ENTRE.BORNES (n1;n2)

n1 : borne inférieure, c'est le premier numéro de la liste des individus (souvent 1)

n2 : borne supérieure. C'est le dernier numéro de la liste (souvent c'est la taille de l'échantillon)

- 5- on appuie sur la touche Entrée
- 6- avec la souris on se positionne sur le coin en bas à droite de cette cellule, on clique sur la souris et on la fait glisser jusqu'à la fin de la liste des numéros
- 7- on effectue un « Edition->Copier » puis « Edition->Collage Spécial->Valeurs » de cette colonne de nombre aléatoires (il s'agit d'une fonction dynamique qui va redonner des nouveaux chiffres dès que l'on effectuera une action dans la feuille de calcul)
- 8- on parcourt la liste des numéros et gardons ceux qui n'ont pas déjà été sélectionnés auparavant pour faire partie de l'échantillon (dans notre exemple, le 13^e nombre tiré est le 91, il avait déjà été tiré lors du 2^e tirage. On le remplace donc par le 16^e tirage (numéro 53), etc...)
- 9- on s'arrête lorsque la taille de l'échantillon est atteinte
- 10- on retiens les individus de la liste complète de la population selon les numéros tirés aléatoirement (ici, l'échantillon se compose des 15 individus ayant les numéros : 4, 91, 57, 3, 100, 28, 55, 9, 29, 95, 20, 24, 10, 11, 53)

Cas pratique : Nous allons tirer aléatoirement 400 numéros de fermiers parmi nos 4000 bénéficiaires (numérotés de 1 à 4000, cf Base de données) en appliquant la formule : ALEA.ENTRE.BORNES(1 ;4000). Les fermiers retenus dans l'échantillon seront les 400 dont le numéro sera tirés.

La simulation « TirageAlea » dans le fichier Excel « OutilsEchant » montre que plus l'échantillon est petit, moins il sera ressemblant de la population (comparaison avec les répartitions par strates)->la répartition est presque respectée pour le tirage simple de 400 fermiers; cela est dû au fait que la taille de l'échantillon est suffisamment grande pour que celui-ci soit représentatif de la population ;en revanche le second tirage (échantillon de 100 fermiers) s'éloigne de la représentativité !





Echantillonnage pondéré

Avantages de cette méthode

- Chaque unité a une chance connue de faire partie de l'échantillon
- Elle est plus fine que l'échantillon aléatoire simple (respecte la répartition de poids des unités d'étude)

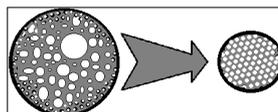
Limites

- La probabilité d'inclusion n'est pas la même pour toutes les unités
- Une représente la population que selon des poids connus et non des catégories différenciées
- Peut amener à omettre certains groupes de personnes surtout lorsque l'échantillon est petit

Principe

L'échantillonnage pondéré est une variante de l'échantillonnage aléatoire simple, mais qui prend en compte le poids des unités de sondage dans le processus de sélection. On entend par poids, le niveau d'importance ou d'influence de l'unité dans la population compte tenu des objectifs de l'étude menée. Ce poids doit pouvoir être quantifié. Il peut dépendre de la taille d'un village en nombre de foyers, il peut dépendre de la taille d'un foyer en nombre d'individus, il peut dépendre de la taille des terres arables pour une ferme... Pondérer l'échantillonnage consiste à donner aux unités plus importantes une probabilité plus forte qu'aux autres d'appartenir à l'échantillon. La différence avec l'échantillonnage aléatoire simple est que **la probabilité de faire partie de l'échantillon diffère d'une unité à l'autre**. On parlera de la probabilité d'inclusion de chaque unité. La somme des probabilité d'inclusion des unités de l'échantillon est égale à 1 (100%). Cela revient à effectuer un tirage aléatoire par la méthode des p'tits papiers en ayant affecté plus de tickets à une unité ayant un poids plus important et moins aux unités ayant un poids faible (ex : on veut effectuer une enquête auprès des foyers de deux villages. Le premier compte 1000 foyers, le second 500. Le poids du premier village est deux fois plus important que celui du second. La probabilité d'être tirée dans l'échantillon pour le premier village (probabilité d'inclusion) est deux fois plus grande que pour le second, et est égale à $1000/500=2/3$ ($1/3$ pour le second). Cela revient à choisir un ticket dans un chapeau parmi 1500, dont 1000 tickets sont affectés au premier village et seulement 500 au second (qui équivaut à choisir un ticket parmi 3 dont 2 affectés au premier et 1 au second)).

Tirage aléatoire pondéré



Comment procéder ?

A partir de la liste de la liste des unités (ex : villages, fermes,...) composant la population étudiée, et de leurs poids respectifs (ex : nombre de foyers, surface de terres arables,...) nous allons :

- 1- déterminer la probabilité d'inclusion pour chaque unité d'étude
- 2- affecter le nombre de tickets correspondant à chaque unité (qui revient à tenir compte de ces différentes probabilités d'appartenir à l'échantillon lors du tirage aléatoire).
- 3- tirer aléatoirement le nombre d'unités requises dans l'échantillon (manuellement si les effectifs sont petits, par le biais de la fonction ALEA sous Excel sinon).

Exemple : On veut constituer un échantillon de 150 foyers parmi les 408 bénéficiaires d'un projet ACF d'un région comportant 10 villages dont les effectifs sont répertoriés dans le tableau page suivante:

- 1- **Le poids de chaque unité dans la population d'étude** est équivalent à son nombre de bénéficiaires (PoidsA=23, PoidsB=38, etc...) divisé par l'effectif total de la population (ici =408 bénéficiaires). La probabilité d'inclusion du village A est $p_A = 23/408 = 6\%$. Pour le village B, $p_B = 38/408 = 9\%$, etc... Ainsi à chaque tirage aléatoire pour la constitution de l'échantillon, le village F ($p_F = 26\%$) qui compte le plus de bénéficiaires, aura 1 chance sur 4 d'être tiré au sort, tandis que le village A en aura seulement 3 sur 50. La somme des probabilités d'inclusion de notre population est égale à 1.





- 2- **Affectation des tickets pour le tirage au sort:** le village A se verra affecter les tickets numérotés de 1 à 23, le village B les 38 suivants (de 24 à 61), etc... le dernier ticket doit avoir le numéro de l'effectif total de la population totale (*ici nombre total de bénéficiaires = 408*)
- 3- **Tirage aléatoire** d'un échantillon de 150 foyers (on l'effectue sous le logiciel Excel en suivant le processus décrit p.33). Si les effectifs sont petits, il est possible de le faire pas l'une des méthodes manuelles utilisées pour le tirage aléatoire simple).
 - On effectue 150 tirages aléatoires en choisissant à chaque fois un ticket dont le numéro est compris entre 1 et le nombre total d'unités dans la population. (*ici entre 1 et 408 : utilisation de la formule ALEA d'Excel-> =ALEA.ENTRE.BORNES(1 ;408)*).
 - Une fois les 150 tickets « uniques » tirés (cf tirage d'un échantillon complémentaire, partie échantillon aléatoire simple), on va les affecter aux unités associées (ex : le ticket 286 est tiré, il correspond au village F auquel sont associés les tickets allant de 202 à 307). Les formules EQUIV et INDEX d'Excel, permettent de retrouver l'unité associée au ticket :

INDEX(« plage où sont listées les unités »;EQUIV(« cellule du ticket concerné »; « plage des tickets maximums correspondants aux unités » ;0);1)

EQUIV : va retrouver l'emplacement d'un ticket issu du tirage aléatoire dans les différentes fourchettes de tickets associées aux différentes unités de la population. Pour ce faire EQUIV va comparer ce ticket à tous les tickets inférieurs des unités de la population et donnera en résultat le numéro de ligne de cette plage où se situe notre chiffre (*ex : EQUIV(G3 ;\$D\$3 :\$D\$12 ;1)=6*)

INDEX : permet de retrouver le nom du village associé à la ligne de la fourchette de tickets où se trouve le ticket tiré aléatoirement (*ex : INDEX (\$A\$3 :\$A\$12 ; 6 ;1)= INDEX (\$A\$3 :\$A\$12 ; EQUIV(G3 ;\$D\$3 :\$D\$12) ;1) =F -> le nom du village associé à la 6^e fourchette de tickets est le village F*)
(\$ permet de figer les cellules concernées lorsqu'on applique une formule à toutes les cellules ; sinon elles sont incrémentées)

| 1 | A | B | C | D | E | F | G | H | I |
|----|---------|---------------------|-------|----------|----------|--------|--------|-------|-------|
| 2 | Village | Nb de bénéficiaires | Poids | ticket - | ticket + | tirage | ticket | EQUIV | INDEX |
| 3 | A | 23 | 6% | 1 | 23 | 1 | 282 | 6 | F |
| 4 | B | 38 | 9% | 24 | 61 | 2 | 56 | 1 | B |
| 5 | C | 45 | 11% | 62 | 106 | 3 | 266 | | |
| 6 | D | 35 | 9% | 107 | 141 | 4 | 253 | | |
| 7 | E | 60 | 15% | 142 | 201 | 5 | 204 | | |
| 8 | F | 106 | 26% | 202 | 307 | 6 | 125 | | |
| 9 | G | 13 | 3% | 308 | 320 | 7 | 154 | | |
| 10 | H | 40 | 10% | 321 | 360 | 8 | 287 | | |
| 11 | I | 17 | 4% | 361 | 377 | 9 | 382 | | |
| 12 | J | 31 | 8% | 378 | 408 | 10 | 172 | | |
| 13 | | 408 | 100% | | | 11 | 387 | | |
| 14 | | | | | | 12 | 106 | | |
| 15 | | | | | | 13 | 333 | | |
| 16 | | | | | | 14 | 307 | | |
| 17 | | | | | | 15 | 157 | | |
| 18 | | | | | | 16 | 297 | | |
| 19 | | | | | | 17 | 55 | | |
| 20 | | | | | | 18 | 354 | | |
| 21 | | | | | | 19 | 75 | | |
| 22 | | | | | | 20 | 150 | | |
| 23 | | | | | | 21 | 56 | | |

N.B. : Le tirage aléatoire pondéré obtenu respecte mieux les répartitions en terme de poids que le tirage aléatoire simple (cf tableau suivant) :

| Village | Nb de bénéficiaires | Population totale | Echantillon Aléatoire Pondéré | | Echantillon Aléatoire Simple | |
|---------|---------------------|-------------------|-------------------------------|------|------------------------------|------|
| A | 23 | 6% | 11 | 7% | 15 | 10% |
| B | 38 | 9% | 15 | 10% | 15 | 10% |
| C | 45 | 11% | 19 | 13% | 15 | 10% |
| D | 35 | 9% | 10 | 7% | 19 | 13% |
| E | 60 | 15% | 20 | 13% | 18 | 12% |
| F | 106 | 26% | 32 | 21% | 16 | 11% |
| G | 13 | 3% | 6 | 4% | 10 | 7% |
| H | 40 | 10% | 18 | 12% | 9 | 6% |
| I | 17 | 4% | 4 | 3% | 17 | 11% |
| J | 31 | 8% | 15 | 10% | 16 | 11% |
| | 408 | 100% | 150 | 100% | 150 | 100% |

Cas pratique : L'échantillonnage basé sur la pondération au niveau des fermiers n'est pas utilisé dans notre exemple. Cependant, on procède à une pondération au niveau des strates d'étude (cf méthode suivante).





Echantillonnage stratifié

Avantages de cette méthode

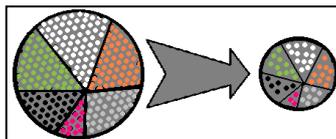
- Chaque unité a une chance connue de faire partie de l'échantillon
- Elle est plus fine que l'échantillon aléatoire pondéré (respecte la représentation proportionnelle)
- Permet d'aboutir à des conclusions sur l'impact d'un phénomène sur la population étudiée

Limites

- Il est nécessaire d'avoir des données détaillées et chiffrées pour déterminer les strates d'étude
- Plus coûteux en terme de temps et de ressources humaines
- Pas adapté dans les échantillons à taille très réduite, empêchant la représentativité

Principe

Le tirage aléatoire proportionnel permet d'améliorer un échantillonnage aléatoire simple en représentant les unités selon leur importance en terme de poids. Bien souvent, la population ciblée présente des différences en son sein, et les personnes sont susceptibles de réagir différemment face à l'aide apportée (ex : les femmes en couple et actives/les femmes seule, avec enfant, sans emploi). Selon le type d'enquête menée, il convient d'approfondir la recherche de « strates » de la population pour s'assurer de la représentativité de cette diversité dans l'échantillon. L'échantillonnage stratifié permet de diviser la population d'étude en sous-groupes d'intérêt homogènes en leur sein et hétérogènes entre eux en fonction de certaines caractéristiques arrêtées à l'avance (ex : l'âge, le sexe, l'appartenance ethnique pour une enquête sur les ménages ou une caractéristique géographique donnée pour une enquête sur l'agriculture). Les variables d'intérêt (de stratification) doivent être simples à utiliser, faciles à observer, étroitement liées au thème de l'enquête.



Tirage stratifié pondéré

Comment procéder ?

Le travail de stratification, préalable au tirage aléatoire, nécessite l'élaboration d'un plan d'échantillonnage. (La procédure détaillée est disponible dans ce manuel dans la deuxième étape, partie « Critères d'échantillonnage »). Les trois étapes majeures de la stratification sont :

- 1- La subdivision de la population en strates (groupes homogènes en leur sein, hétérogènes entre eux)-
Elaboration du plan d'échantillonnage
- 2- Calcul du nombre d'individus par strates dans l'échantillon proportionnellement à leur représentation dans la population d'étude
- 3- Tirage aléatoire simple de ce nombre d'individus indépendamment dans chacune des strates (Excel)

N.B. : Toujours privilégier cette méthode à une autre ! Elle est la plus efficace : la rigueur de recherche d'un échantillon représentatif mène à l'interprétation des résultats la plus fiable (comparativement aux autres méthodes possibles). Malheureusement, la base de données de la population n'est pas toujours disponible (ou pas à jour). Il convient donc de tout mettre en oeuvre pour obtenir des données chiffrées et fiables sur la population d'étude, quitte à réduire le temps d'enquête (et donc la taille de l'échantillon) pour permettre de construire une base de données valable.

Cas pratique : cf étape 2 : p.13 à 18 et p.27 pour le plan d'échantillonnage complet. Les différentes étapes sont :

- 1- détection des variables d'intérêt et choix des strates d'étude selon les effectifs par sous-groupes (elles seront au nombre de 10)
- 2- pour chaque strate : calcul du poids dans la population et du nombre d'unités à tirer (selon la taille de l'échantillon arrêtée)
- 3- tirage aléatoire simple dans chacune des strates avec la formule ALEA d'Excel (cf simulation (SimulStrat) sur la base de données de notre population qui se trouve dans le fichier Excel nommé « Outils Echant » dans l'onglet « Tirage Aleat »)





Méthodes mixtes

Enfin, il existe enfin des cas particulier (**échantillonnage en deux étapes par grappe, sur-représentation d'un strate, représentation paritaire des strates**) faisant l'objet de techniques de choix spécifiques.

Echantillonnage à plusieurs degrés (par grappes)

Avantages de cette méthode

- Ne nécessite pas une liste globale de la population
- Réduit les coûts de déplacement (on interview plusieurs unités au même endroit)
- Réduit la taille de l'échantillon
- Permet d'aboutir à des conclusions sur l'impact d'un phénomène sur la population étudiée

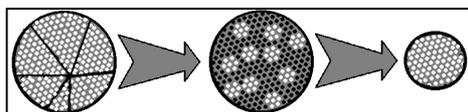
Limites

- Moins fine que l'échantillonnage aléatoire stratifié
- Perte de précision de l'estimation (homogénéité accrue: les unités d'une grappe se ressemblent entre eux)
- Pas adapté dans les échantillons à taille très réduite, empêchant la représentativité

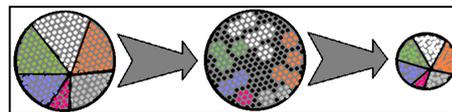
Principe

Lorsque nous ne disposons pas d'informations chiffrées sur les unités d'étude (*ex : nombre, âge, taille, poids des enfants des foyers*) ou lorsque la zone d'enquête est très étendue (*ex : enquête nationale, accès difficile*) ou/et que le temps d'enquête imparti est trop juste (*ex : 1 mois*), nous avons recours à l'échantillonnage par grappes. Celui-ci se caractérise par le fait que l'enquêteur interview plusieurs unités dans un même groupe (*ex : on enquête plusieurs enfants dans un même foyer*) et qu'il construit l'échantillon sur le terrain, ne bénéficiant pas de données suffisantes pour le faire au préalable (*ex : enquête auprès de foyers, on connaîtra la composition des foyers une fois sur place*). De plus, si nous disposons de données chiffrées sur le groupe d'appartenance des unités, appelé « grappe » (*ex : village, communauté, quartier, ...*), nous pouvons en dégager au préalable des sous-groupes distinctifs. Ainsi, avant le début de l'enquête, on pourra effectuer une stratification (construction d'un plan d'échantillonnage (*ex : foyers de pêcheurs, foyers de fermiers, foyers monoparentaux, ...*)) non plus au niveau des unités mais au niveau des grappes. A la différence des autres méthodes, où nous prenons en compte les informations au niveau de l'unité enquêtées, l'échantillonnage par grappe prend en compte le niveaux d'étude supérieur à l'unités (*ex : région, village, foyer, enfant, ...*). Ce type de pratique réduit considérablement la taille de l'échantillon (*ex : on récolte des informations pour plusieurs unités à chaque enquête*) mais, en contre partie, cela implique une perte de précision par rapport à l'échantillonnage stratifié, car les unités enquêtées se ressemblent fortement au sein des grappes (*forte homogénéité : ex : 2 enfants d'un même foyer qui reçoivent la même éducation se ressembleront plus que deux enfants issus de deux foyers différents*).

Echantillonnage aléatoire simple par grappes



Echantillonnage stratifié par grappes



Comment procéder ?

La construction d'un échantillon par grappes se déroule en quatre étapes :

- 1- Stratification au niveau de la grappe
- 2- Calcul de la taille de l'échantillon (utilisation de tables d'échantillonnage spécifiques)
- 3- Tirage des grappes dans chaque strates proportionnellement à leur représentation dans la population
- 4- Tirage des unités au sein de chaque grappe (le nombre maximal d'unités par grappes est défini lors du calcul de la taille de l'échantillon)





Exemple : on veut mener une enquête auprès des enfants de moins de 10 ans d'une région. Des informations quantifiées sont disponibles au niveau des villages (nombre de foyers, liste de ces foyers, types de foyers, situation géographique, niveau de vie ? ...), mais aucune information n'est disponible au niveau des foyers (nombre d'enfants ? nombre d'enfants de moins de 10 ans ? activités ? ...) ni au niveau des enfants (ages, poids, habitudes alimentaires, ...)

1- Stratification (si les données existent) au niveau de la grappe pour s'assurer de la représentativité des unités dans l'échantillon enquêté (*ex : on différencie les types de foyers représentés dans la population de la région pour créer des strates de populations au sein desquelles le comportement des enfants que l'on veut étudier est homogène : urbain/rural ; type d'activité (agricole/administration/ tertiaire) ; niveau de vie ; ...*)

2- Calcul de la taille de l'échantillon (et du nombre d'unités à enquêter par grappes): on a vu plus haut que cette méthode impliquait une perte de précision par rapport à l'échantillonnage stratifié. La précision d'un échantillon en grappes dépend essentiellement de trois paramètres :

- le coefficient de corrélation intraclasse (p) ;
- le nombre de grappes tirées ;
- le nombre d'individus tirés dans chaque grappe.

Des tables d'échantillonnage nous fournissent la précision de l'estimation en fonction de ces différents paramètres (cf table p.46-47 Annexes III). Il est important de savoir utiliser ces tables pour un échantillonnage par grappes car elles permettent de faire des choix selon différentes considérations. Le tableau permet de trouver la taille de l'échantillon par grappes équivalente à la taille minimale requise pour ces critères statistiques lorsque l'on enquête les unités une par une (*ex : un enfant par foyer*), selon le nombre d'unités que l'on veut enquêter par grappe (*ex : 2 enfants par foyer*).

La première ligne indique le degré de certitude exigé pour les intervalles de confiance : 95%. La seconde ligne fournit différentes tailles d'intervalles de confiance qui vont en s'accroissant de la gauche vers la droite. « 0.05s » signifie que l'on exige que les moyennes soient estimées à 0.05 écart type près, et « 2.5% » indique que les proportions doivent être estimées à 2.5% près. Le tableau est ensuite divisé en plusieurs tables selon la valeur de p , le coefficient de corrélation intra-classe. Ce coefficient ne peut être connu à l'avance, on l'estime en général à partir d'autres études ou à défaut on prend la valeur 0.3.

Exemple : on veut estimer le rapport poids taille des enfants à 0.10 écart type près (degré de précision) avec probabilité de 95% (confiance) P est inconnu et estimé à 0.3.

Ici la taille minimale requise pour un échantillon « classique » est de 400. La taille minimale équivalente si l'on enquête 2 enfant par foyer est de 260 et si l'on enquête 5 enfants par foyer on réduit la taille de l'échantillon à 176 foyers.

3- Tirage des grappes dans chaque strate déterminée selon la proportion observée dans la population totale (*ex : à partir de la liste des foyers de la région et pour obtenir un échantillon de 260 foyers (grappes), on tire $30\% \times 260 = 78$ foyers en milieu urbain, et 182 en milieu rural (représentation au niveau régional de ces milieux)*)

4- Tirage des unités : au sein de chaque grappe on tire un certain nombre d'unité (*ex : une fois sur le terrain, dans chaque foyer, on enquête au plus 2 enfants de moins de 10ans*). On tire aléatoirement ces unités parmi la totalité des unités de la grappe (*ex : on choisit au hasard 2 enfants parmi tous les enfants ayant moins de 10 ans dans chaque foyer*).

Cas pratique : si l'on décide d'enquêter plusieurs personnes dans les foyers de fermiers, nous pourrions réduire substantiellement la taille de l'échantillon. 400 est le nombre d'enquête minimum que nous devons effectuer, si nous n'enquêtons que le fermier. Si l'étude menée se penchait plus sur l'organisation au sein du foyer, nous aurions pu enquêter différentes personnes dans le foyer (femme, enfants, employé de la ferme, ...): pour un nombre de 5 enquête par grappes (5 personnes différentes enquêtées par foyer), la taille de l'échantillon minimale aurait été de 176 foyers (à même critère statistiques, et pour un $p=0.3$) (cf table statistique p.46/47)





Echantillonnage aléatoire stratifié non proportionnel (sur-représentation)

Avantages de cette méthode

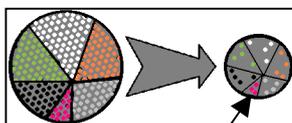
- Permet d'aboutir à des conclusions sur l'impact d'un phénomène spécifique sur la population étudiée

Limites

- Modifie les proportions dans l'échantillon (en tenir compte lors des calculs de moyenne, et analyses)
- Doit être compensé par la réduction d'une autre strate lorsque les moyens d'enquête sont comptés

Principe

On a recours à cette technique lorsqu'une strate représente une petite proportion de la population mais que l'on désire effectuer une étude particulière à son sujet (ex: *foyers bénéficiant d'énergie solaire*). Dans ce cas de figure, il est nécessaire de sur-représenter cette strate d'intérêt spécifique. Lors des calculs de moyennes et autres résultats statistiques, il ne faudra pas oublier de tenir compte de cette sur-représentation (et donc implicitement de la sous-représentation associée) pour ne pas faire d'interprétations erronées.



Sur-représentation d'une strate

Comment procéder ?

On reprend le plan d'échantillonnage élaboré précédemment et on affecte un nouveau poids (probabilité d'inclusion) à la strate que l'on veut sur-représenter. Si le temps ou le budget sont comptés, il faut compenser l'ajout d'unités dans cette strate dans l'échantillon par une suppression d'un nombre équivalent d'unités dans une ou plusieurs strates dans ce même échantillon (on compensera alors le surplus sur les strates comptant le plus grand effectif).

Cas pratique : Notre plan d'échantillonnage prévoit 6% de « pasteurs destitués, n'ayant plus de bétail », ce qui fait un nombre de 19. Une étude spécifique va se pencher sur la raison de la perte de tous leur cheptel et les capacités d'adaptation de ces fermiers face à la crise. Il convient donc d'augmenter la représentation de ces individus dans l'échantillon d'étude. On va augmenter à 15% la représentation de cette strate (i.e. 60 fermiers). Les capacités d'enquêtes sont suffisantes pour absorber cet ajout de fermiers dans l'échantillon (taille maximale possible=560). Si les capacités avaient été réduite, nous aurions compensé cet ajout par une suppression de fermiers de la strate 1 dans l'échantillon (98-38=60).

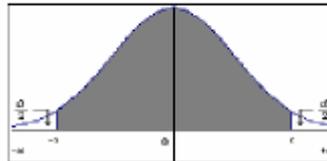
| NbreStrate | LibStrate | NbBenef | PoidsStrate (%) | NbEchant | Nb Echant SurReprésent |
|------------|--|---------|-----------------|----------|------------------------|
| 1 | Ferriers de longue date , ayant 1 à 3 bêtes, dont les récoltes permettent une autonomie alimentaire de plus de 6 mois et sont génératrices de revenus supplémentaires | 984 | 25% | 98 | 98 |
| 2 | Ferriers de longue date , ayant 1 à 3 bêtes, dont les récoltes permettent une autonomie alimentaire de plus de 6 mois et mais sans revenus supplémentaires | 197 | 5% | 20 | 20 |
| 3 | Ferriers de longue date , ayant 1 à 3 bêtes, dont les récoltes ne permettent pas une autonomie alimentaire suffisante | 229 | 6% | 23 | 23 |
| 4 | Ferriers de longue date , n'ayant plus de bétail | 230 | 6% | 23 | 23 |
| 5 | Ferriers anciennement réfugiés , ayant 1 à 3 bêtes | 682 | 17% | 68 | 68 |
| 6 | Ferriers anciennement réfugiés , n'ayant plus de bétail, dont les récoltes permettent une autonomie alimentaire de plus de 6 mois | 372 | 9% | 37 | 37 |
| 7 | Ferriers anciennement réfugiés , n'ayant plus de bétail, récoltes ne permettent pas une autonomie alimentaire suffisante | 186 | 5% | 19 | 19 |
| 8 | Ferriers pasteurs destitués , ayant 1 à 3 bêtes, dont les récoltes permettent une autonomie alimentaire de plus de 6 mois | 336 | 8% | 34 | 34 |
| 9 | Ferriers pasteurs destitués , ayant 1 à 3 bêtes, dont les récoltes ne permettent pas une autonomie alimentaire suffisante | 560 | 14% | 56 | 56 |
| 10 | Ferriers de pasteurs destitués , n'ayant plus de bétail | 224 | 6% | 22 | 60 |
| | | 4000 | 100% | 400 | 438 |





ANNEXES I

Loi de Student



| α | 1 | 0,8 | 0,6 | 0,4 | 0,2 | 0,1 | 0,05 | 0,02 | 0,01 | 0,002 | 0,001 |
|------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| $1 - \alpha$ | 0 | 0,2 | 0,4 | 0,6 | 0,8 | 0,9 | 0,95 | 0,98 | 0,99 | 0,998 | 0,999 |
| $v = \text{ddl}$ | | | | | | | | | | | |
| 1 | 0,0000 | 0,3249 | 0,7265 | 1,3764 | 3,0777 | 6,3137 | 12,706 | 31,821 | 63,656 | 318,29 | 636,58 |
| 2 | 0,0000 | 0,2887 | 0,6172 | 1,0607 | 1,8856 | 2,9200 | 4,3027 | 6,9545 | 9,9250 | 22,328 | 31,600 |
| 3 | 0,0000 | 0,2767 | 0,5844 | 0,9785 | 1,6377 | 2,3534 | 3,1824 | 4,5407 | 5,8408 | 10,214 | 12,924 |
| 4 | 0,0000 | 0,2707 | 0,5686 | 0,9410 | 1,5332 | 2,1318 | 2,7765 | 3,7469 | 4,6041 | 7,1729 | 8,6101 |
| 5 | 0,0000 | 0,2672 | 0,5594 | 0,9195 | 1,4759 | 2,0150 | 2,5706 | 3,3649 | 4,0321 | 5,8935 | 6,8885 |
| 6 | 0,0000 | 0,2648 | 0,5534 | 0,9057 | 1,4398 | 1,9432 | 2,4469 | 3,1427 | 3,7074 | 5,2075 | 5,9587 |
| 7 | 0,0000 | 0,2632 | 0,5491 | 0,8960 | 1,4149 | 1,8945 | 2,3645 | 2,9979 | 3,4995 | 4,7853 | 5,4081 |
| 8 | 0,0000 | 0,2619 | 0,5459 | 0,8909 | 1,3960 | 1,8595 | 2,3000 | 2,9065 | 3,3554 | 4,5000 | 5,0414 |
| 9 | 0,0000 | 0,2610 | 0,5435 | 0,8834 | 1,3830 | 1,8331 | 2,2622 | 2,8214 | 3,2498 | 4,2969 | 4,7809 |
| 10 | 0,0000 | 0,2602 | 0,5415 | 0,8791 | 1,3722 | 1,8125 | 2,2281 | 2,7638 | 3,1693 | 4,1437 | 4,5868 |
| 11 | 0,0000 | 0,2596 | 0,5399 | 0,8755 | 1,3634 | 1,7959 | 2,2010 | 2,7181 | 3,1058 | 4,0248 | 4,4369 |
| 12 | 0,0000 | 0,2590 | 0,5386 | 0,8726 | 1,3562 | 1,7823 | 2,1788 | 2,6810 | 3,0545 | 3,9296 | 4,3178 |
| 13 | 0,0000 | 0,2586 | 0,5375 | 0,8702 | 1,3502 | 1,7709 | 2,1604 | 2,6503 | 3,0123 | 3,8520 | 4,2209 |
| 14 | 0,0000 | 0,2582 | 0,5366 | 0,8681 | 1,3450 | 1,7613 | 2,1448 | 2,6245 | 2,9768 | 3,7874 | 4,1403 |
| 15 | 0,0000 | 0,2579 | 0,5357 | 0,8662 | 1,3406 | 1,7531 | 2,1315 | 2,6025 | 2,9467 | 3,7329 | 4,0728 |
| 16 | 0,0000 | 0,2576 | 0,5350 | 0,8647 | 1,3368 | 1,7459 | 2,1199 | 2,5835 | 2,9208 | 3,6861 | 4,0149 |
| 17 | 0,0000 | 0,2573 | 0,5344 | 0,8633 | 1,3334 | 1,7396 | 2,1098 | 2,5669 | 2,8982 | 3,6458 | 3,9651 |
| 18 | 0,0000 | 0,2571 | 0,5338 | 0,8620 | 1,3304 | 1,7341 | 2,1009 | 2,5524 | 2,8784 | 3,6105 | 3,9217 |
| 19 | 0,0000 | 0,2569 | 0,5333 | 0,8610 | 1,3277 | 1,7291 | 2,0930 | 2,5395 | 2,8609 | 3,5793 | 3,8833 |
| 20 | 0,0000 | 0,2567 | 0,5329 | 0,8600 | 1,3253 | 1,7247 | 2,0860 | 2,5280 | 2,8453 | 3,5518 | 3,8496 |
| 21 | 0,0000 | 0,2566 | 0,5325 | 0,8591 | 1,3232 | 1,7207 | 2,0796 | 2,5176 | 2,8314 | 3,5271 | 3,8193 |
| 22 | 0,0000 | 0,2564 | 0,5321 | 0,8583 | 1,3212 | 1,7171 | 2,0739 | 2,5083 | 2,8188 | 3,5050 | 3,7922 |
| 23 | 0,0000 | 0,2563 | 0,5317 | 0,8575 | 1,3195 | 1,7139 | 2,0687 | 2,4999 | 2,8073 | 3,4850 | 3,7676 |
| 24 | 0,0000 | 0,2562 | 0,5314 | 0,8569 | 1,3178 | 1,7109 | 2,0639 | 2,4922 | 2,7970 | 3,4668 | 3,7454 |
| 25 | 0,0000 | 0,2561 | 0,5312 | 0,8562 | 1,3163 | 1,7081 | 2,0595 | 2,4851 | 2,7874 | 3,4502 | 3,7251 |
| 26 | 0,0000 | 0,2560 | 0,5309 | 0,8557 | 1,3150 | 1,7056 | 2,0555 | 2,4786 | 2,7787 | 3,4350 | 3,7067 |
| 27 | 0,0000 | 0,2559 | 0,5306 | 0,8551 | 1,3137 | 1,7033 | 2,0518 | 2,4727 | 2,7707 | 3,4210 | 3,6895 |
| 28 | 0,0000 | 0,2558 | 0,5304 | 0,8546 | 1,3125 | 1,7011 | 2,0484 | 2,4671 | 2,7633 | 3,4082 | 3,6739 |
| 29 | 0,0000 | 0,2557 | 0,5302 | 0,8542 | 1,3114 | 1,6991 | 2,0452 | 2,4620 | 2,7564 | 3,3963 | 3,6595 |
| 30 | 0,0000 | 0,2556 | 0,5300 | 0,8538 | 1,3104 | 1,6973 | 2,0423 | 2,4573 | 2,7500 | 3,3852 | 3,6460 |
| 40 | 0,0000 | 0,2550 | 0,5286 | 0,8507 | 1,3031 | 1,6839 | 2,0211 | 2,4233 | 2,7045 | 3,3069 | 3,5510 |
| 50 | 0,0000 | 0,2547 | 0,5278 | 0,8489 | 1,2967 | 1,6759 | 2,0086 | 2,4033 | 2,6778 | 3,2614 | 3,4960 |
| 60 | 0,0000 | 0,2545 | 0,5272 | 0,8477 | 1,2958 | 1,6706 | 2,0003 | 2,3901 | 2,6603 | 3,2317 | 3,4602 |
| 70 | 0,0000 | 0,2543 | 0,5268 | 0,8468 | 1,2938 | 1,6689 | 1,9944 | 2,3808 | 2,6479 | 3,2108 | 3,4350 |
| 80 | 0,0000 | 0,2542 | 0,5265 | 0,8461 | 1,2922 | 1,6641 | 1,9901 | 2,3739 | 2,6387 | 3,1952 | 3,4164 |
| 90 | 0,0000 | 0,2541 | 0,5263 | 0,8456 | 1,2910 | 1,6620 | 1,9867 | 2,3685 | 2,6316 | 3,1832 | 3,4019 |
| 100 | 0,0000 | 0,2540 | 0,5261 | 0,8452 | 1,2901 | 1,6602 | 1,9840 | 2,3642 | 2,6259 | 3,1738 | 3,3905 |
| 200 | 0,0000 | 0,2537 | 0,5252 | 0,8434 | 1,2858 | 1,6525 | 1,9719 | 2,3451 | 2,6006 | 3,1315 | 3,3398 |
| ∞ | 0,0000 | 0,2533 | 0,5244 | 0,8416 | 1,2816 | 1,6449 | 1,9600 | 2,3263 | 2,5758 | 3,0903 | 3,2905 |





ANNEXES II

Tables d'échantillonnage, méthode par grappes

Taille des échantillons

| Limites de confiance : 95% | | | | | | | | |
|----------------------------|-------------------|-----------------|-------------------|-----------------|-------------------|-----------------|-------------------|-----------------|
| Taille des grappes | 0.05 s 2.5% | | 0.10 s 5.0% | | 0.15 s 7.5% | | 0.20 s 10.0% | |
| | Nombre de grappes | Nombre d'unités |
| P=0.1 | | | | | | | | |
| 1 | 1600 | 1600 | 400 | 400 | 178 | 178 | 100 | 100 |
| 2 | 880 | 1760 | 220 | 440 | 98 | 196 | 55 | 110 |
| 5 | 448 | 2240 | 112 | 560 | 50 | 250 | 28 | 140 |
| 10 | 304 | 3040 | 76 | 760 | 34 | 340 | 19 | 190 |
| 15 | 256 | 3840 | 64 | 960 | 29 | 435 | 16 | 240 |
| 20 | 232 | 4640 | 58 | 1160 | 26 | 520 | 15 | 300 |
| 30 | 208 | 6240 | 52 | 1560 | 24 | 720 | 13 | 390 |
| 40 | 196 | 7840 | 49 | 1960 | 22 | 880 | 13 | 520 |
| 50 | 189 | 9450 | 48 | 2400 | 21 | 1050 | 12 | 600 |
| P=0.2 | | | | | | | | |
| 1 | 1600 | 1600 | 400 | 400 | 178 | 178 | 100 | 100 |
| 2 | 960 | 1920 | 240 | 480 | 107 | 214 | 60 | 120 |
| 5 | 576 | 2880 | 144 | 720 | 65 | 325 | 35 | 180 |
| 10 | 448 | 4480 | 112 | 1120 | 50 | 500 | 28 | 280 |
| 15 | 406 | 6090 | 102 | 1530 | 46 | 690 | 26 | 390 |
| 20 | 384 | 7680 | 96 | 1920 | 43 | 860 | 24 | 480 |
| 30 | 363 | 10890 | 91 | 2730 | 41 | 1230 | 23 | 690 |
| 40 | 352 | 14080 | 88 | 3520 | 40 | 1600 | 22 | 880 |
| 50 | 346 | 17300 | 87 | 4350 | 39 | 1950 | 22 | 1100 |
| P=0.3 | | | | | | | | |
| 1 | 1600 | 1600 | 400 | 400 | 178 | 178 | 100 | 100 |
| 2 | 1050 | 2100 | 260 | 520 | 116 | 232 | 65 | 130 |
| 5 | 704 | 3520 | 176 | 880 | 79 | 395 | 44 | 220 |
| 10 | 592 | 5920 | 148 | 1480 | 66 | 660 | 37 | 370 |
| 15 | 555 | 8325 | 139 | 2085 | 62 | 930 | 35 | 525 |
| 20 | 536 | 10720 | 134 | 2680 | 60 | 1200 | 34 | 680 |
| 30 | 518 | 15540 | 130 | 3900 | 58 | 1740 | 33 | 990 |
| 40 | 508 | 20320 | 127 | 5080 | 57 | 2280 | 32 | 1280 |
| 50 | 503 | 25150 | 126 | 6300 | 56 | 2800 | 32 | 1600 |





Tables d'échantillonnage (suite)

Taille des échantillons

| Limites de confiance : 95% | | | | | | | | |
|----------------------------|-------------------|-----------------|-------------------|-----------------|-------------------|-----------------|-------------------|-----------------|
| Taille des grappes | 0.05 s 2.5% | | 0.10 s 5.0% | | 0.15 s 7.5% | | 0.20 s 10.0% | |
| | Nombre de grappes | Nombre d'unités |
| P=0.4 | | | | | | | | |
| 1 | 1600 | 1600 | 400 | 400 | 178 | 178 | 100 | 100 |
| 2 | 1120 | 2240 | 280 | 560 | 125 | 250 | 70 | 140 |
| 5 | 832 | 4160 | 208 | 1040 | 93 | 465 | 52 | 260 |
| 10 | 726 | 7260 | 184 | 1840 | 82 | 820 | 46 | 460 |
| 15 | 704 | 10560 | 176 | 2640 | 79 | 1185 | 44 | 660 |
| 20 | 688 | 13760 | 172 | 3440 | 77 | 1540 | 43 | 860 |
| 30 | 672 | 20160 | 168 | 5040 | 75 | 2250 | 42 | 1260 |
| 40 | 664 | 26560 | 166 | 6640 | 74 | 2960 | 42 | 1680 |
| 50 | 660 | 33000 | 165 | 8250 | 74 | 3700 | 42 | 2100 |
| P=0.5 | | | | | | | | |
| 1 | 1600 | 1600 | 400 | 400 | 178 | 178 | 100 | 100 |
| 2 | 1200 | 2400 | 300 | 600 | 134 | 268 | 75 | 150 |
| 5 | 960 | 4800 | 240 | 1200 | 107 | 535 | 60 | 300 |
| 10 | 880 | 8800 | 220 | 2200 | 98 | 980 | 55 | 550 |
| 15 | 854 | 12810 | 214 | 3210 | 95 | 1425 | 54 | 810 |
| 20 | 840 | 16800 | 210 | 4200 | 94 | 1880 | 53 | 1060 |
| 30 | 827 | 24810 | 207 | 6210 | 92 | 2760 | 52 | 1560 |
| 40 | 820 | 32800 | 205 | 8200 | 92 | 3680 | 52 | 2080 |
| 50 | 816 | 40800 | 204 | 10200 | 91 | 4550 | 51 | 2550 |
| P=0.6 | | | | | | | | |
| 1 | 1600 | 1600 | 400 | 400 | 178 | 178 | 100 | 100 |
| 2 | 1280 | 2560 | 320 | 640 | 143 | 286 | 80 | 160 |
| 5 | 1088 | 5440 | 272 | 1360 | 122 | 610 | 68 | 340 |
| 10 | 1024 | 10240 | 256 | 2560 | 114 | 1140 | 64 | 640 |
| 15 | 1003 | 15045 | 251 | 3765 | 112 | 1680 | 63 | 945 |
| 20 | 992 | 19840 | 248 | 4960 | 111 | 2220 | 62 | 1240 |
| 30 | 982 | 29460 | 246 | 7380 | 110 | 3300 | 62 | 1860 |
| 40 | 976 | 39040 | 244 | 9760 | 109 | 4360 | 61 | 2440 |
| 50 | 973 | 48650 | 244 | 12200 | 109 | 5450 | 61 | 3050 |

