

Guide Rapide pour l'Echantillonnage d'une Enquête

L'échantillonnage est l'une des parties les plus importantes de vos enquêtes. Dans la plupart des cas, nos enquêtes ne disposent pas des ressources nécessaires pour collecter des informations auprès de tous les membres de la population cible. Une stratégie d'échantillonnage correcte nous permet de conduire l'enquête uniquement sur une partie de la population tout en s'assurant que les résultats sont valables pour l'ensemble de la population cible. Cependant, l'échantillonnage est sujet à de nombreuses erreurs qui peuvent compromettre la validité de vos données. IndiKit a donc préparé un bref guide pour vous aider à assurer la qualité maximale de l'échantillonnage de votre enquête.

CHOISIR QUI ENQUÊTER

Tout d'abord, définissez votre **population cible** – les personnes dont nous souhaitons collecter les données. Les populations cibles les plus courantes sont :

- les personnes que votre intervention vise à aider (ses bénéficiaires directs)
- d'autres personnes qui pourraient avoir bénéficié indirectement (peuvent être utilisées pour évaluer les répercussions de l'intervention)
- d'autres parties prenantes de vos interventions dont vous devez évaluer les opinions et autres données
- les membres de votre groupe de comparaison (voir ci-dessous)
- divers groupes de population vivant dans une zone donnée (couramment utilisés lors d'évaluation des besoins).

Gardez à l'esprit qu'une enquête peut avoir plusieurs populations cibles différentes. Par exemple, une enquête de base sur une intervention agricole sensible à la nutrition pourrait cibler les enfants de moins de 5 ans (pour mesurer la prévalence de la malnutrition), leurs accompagnants, les agriculteurs formés ainsi que les vulgarisateurs agricoles. Ces enquêtes nécessitent alors des échantillons distincts pour chaque groupe cible.

Groupes de Contrôle et de Comparaison

La plupart des enquêtes humanitaires et de développement s'intéressent à l'évaluation des changements apportés aux populations concernées. Une hypothèse fréquente est qu'en comparant les données de base et les données finales, nous pouvons observer les changements apportés par notre intervention. Cependant, comment savons-nous que ces changements ont été causés par notre intervention et non par des facteurs externes tels que les changements climatiques (affectant la récolte), les fluctuations des prix du marché (influant sur l'accès à la nourriture) ou certaines tendances que le pays est en train de vivre, telles que la réduction de la pauvreté ? L'approche la plus commune pour évaluer **quels changements peuvent être attribués à notre intervention** est de comparer la situation des bénéficiaires de notre intervention avec la situation d'un groupe similaire de personnes qui n'ont pas été exposées à l'intervention. Le résultat devrait nous dire ce qui se serait passé si notre intervention n'avait pas été mise en œuvre. Il y a deux approches très courantes :

- **Utilisation d'un Groupe de Contrôle** : Cette approche divise de manière aléatoire les bénéficiaires potentiels de l'intervention en deux groupes : le groupe de traitement (qui reçoit l'aide prévue) et le groupe de contrôle (qui ne reçoit aucune aide). Elle compare ensuite les changements vécus par les membres du groupe de traitement et par ceux du groupe de contrôle. En raison de problèmes éthiques et de programmation évidents, cette approche est surtout utilisée lorsque **l'objectif principal de l'intervention est d'obtenir des preuves** (ex : de l'efficacité d'un produit ou d'une approche donnée). Elle est adaptée aux interventions purement axées sur la recherche ou pour des interventions pilotes qui sont censées fournir des preuves nécessaires pour diffuser l'approche testée (produit, etc.) à plus grande échelle. Les membres du groupe de contrôle ne doivent pas nécessairement être laissés sans assistance - une **conception dite progressive** permet au groupe de traitement de participer en premier tandis que le groupe de contrôle reçoit l'assistance une fois le test initial terminé. Ainsi, tous les participants peuvent recevoir la même assistance, à des moments différents.

- Utilisation d'un Groupe de Comparaison : Dans de nombreux contextes, l'utilisation de groupes de contrôle n'est pas une option appropriée. Dans ces cas, la meilleure approche consiste à utiliser un **groupe de comparaison** composé de personnes ayant des caractéristiques très similaires à celles des bénéficiaires de votre intervention. Par exemple, si votre projet cible des agriculteurs vivant dans 50 communautés, votre groupe de comparaison peut être des agriculteurs vivant dans des communautés qui ne sont pas exposées à votre intervention mais qui ont un environnement, des pratiques agronomiques, un accès aux infrastructures et d'autres caractéristiques essentielles similaires.

Sélection de l'Echantillon

Une fois que vous savez quelle est votre population cible, la question suivante est de savoir comment sélectionner les participants à l'enquête. Vous pouvez utiliser les approches suivantes qui sont les plus courantes :

- L'Echantillonnage Aléatoire Simple donne à chaque membre de la population cible une chance égale d'être choisie pour participer à l'enquête. La méthode de sélection la plus courante consiste à préparer une liste de tous les membres de la population cible, puis à les sélectionner en fonction de nombres générés de façon aléatoire (fournis, par exemple, par la fonction ALEA d'Excel). Si la liste n'est pas disponible (généralement pour les populations plus importantes), utilisez d'autres techniques d'échantillonnage, telles que l'Echantillonnage en Grappes ou l'Echantillonnage en Grappes à Plusieurs Degrés.
- L'Echantillonnage en Grappes est une bonne alternative si vous n'avez pas une liste de la population cible entière ou s'il y a une grande zone géographique à couvrir. Il s'agit de sélectionner au hasard des groupes et non des individus. Par exemple, si nous voulons étudier la population d'une ville entière, nous pouvons la diviser d'abord en unités géographiques (grappes), comme les quartiers ou les rues, qui ont à peu près la même population et les mêmes caractéristiques (ceci est très important). Ensuite, nous sélectionnons au hasard un certain nombre de ces grappes et sondons tous les membres de notre population cible vivant dans les grappes sélectionnées.
- L'Echantillonnage en Grappes à Plusieurs Degrés implique la sélection aléatoire sur deux niveaux ou plus. Par exemple, comme dans le cas ci-dessus, il s'agit de diviser la zone en unités géographiques (grappes) ayant à peu près la même population et les mêmes caractéristiques. Par la suite, sélectionnez au hasard plusieurs grappes. Le nombre dépendra de votre capacité et de vos ressources. La prochaine étape consiste à ne pas interroger tous les membres de la population cible vivant dans la grappe, mais à les sélectionner au hasard dans chaque grappe, généralement en sélectionnant les ménages et en interrogeant un membre du ménage faisant partie de votre population cible. La sélection aléatoire des ménages est basée sur une estimation approximative du nombre de membres de la population cible vivant dans la grappe donnée (parfois, vous pouvez également utiliser Google Maps pour estimer). Sélectionnez aléatoirement un point de départ (par exemple, une intersection sélectionnée au hasard). Quand vous arrivez à cet endroit, faites tourner un stylo pour choisir la direction dans laquelle vous marcherez en premier. Pour décider quel ménage vous visiterez en premier, choisissez au hasard un nombre de 1 à 10 et visitez le ménage donné (par exemple, si vous choisissez au hasard 5, visitez le 5ème ménage). Ensuite, continuez l'enquête tous les X ménages (par exemple, tous les 7 ménages) en vous basant sur le nombre total de ménages et la taille de votre échantillon (ce que l'on appelle l'intervalle d'échantillonnage). Lorsque vous arrivez à la fin de la zone selon la direction donnée, vous pouvez tourner à nouveau le stylo et continuer dans une nouvelle direction jusqu'à ce que vous ayez interrogé le nombre de répondants requis.
- L'Echantillonnage Stratifié permet l'identification de sous-groupes (strates) au sein d'une population et la création d'un échantillon qui reflète leur représentation réelle dans la population. Les individus membres de chaque sous-groupe sont choisis par échantillonnage aléatoire simple. La taille des sous-groupes de l'échantillon global devrait être proportionnelle à l'ensemble de la population cible. Par exemple, si les femmes représentent 60% des bénéficiaires d'un projet d'urgence, la proportion de femmes interrogées devrait être de 60%.

- L'Echantillonnage de Convenance (Echantillonnage Raisonné) est utilisé lorsque la sélection aléatoire est trop compliquée ou impossible. L'enquêteur sélectionne comme répondants les membres de la population cible qui sont disponibles à proximité, comme les gens dans la rue, dans un marché, etc. Bien que cette approche fasse gagner du temps, les données fournies ne sont probablement pas représentatives de la population. L'échantillonnage de convenance peut mener à des conclusions trompeuses - par exemple, interroger les clients d'une clinique rurale vous dira très peu de choses sur les besoins de santé de la population en général, car vous n'aurez pas recueilli de données auprès de personnes qui ne fréquentent pas les cliniques de santé (ex : parce qu'ils ne peuvent pas se le permettre ou sont en bonne santé).
- L'Echantillonnage Boule de Neige consiste à trouver plusieurs membres de la population cible, puis à les utiliser pour trouver d'autres répondants (en leur demandant de nous diriger vers d'autres personnes). Chaque « vague » de répondants est utilisée pour contacter les suivants, de sorte que l'échantillon grandit lentement comme une boule de neige roulant sur une colline. Bien que cette technique ne fournisse pas un échantillon représentatif de la population cible, elle est utile pour accéder à des répondants difficiles à atteindre.

Gardez en tête que l'échantillonnage est **sujet à de nombreux biais de sélection**, tels que les enquêteurs qui remplacent de manière incorrecte un participant indisponible par un autre répondant ou la réalisation d'une évaluation des besoins pendant un jour de marché alors que les personnes économiquement plus actives et plus jeunes sont absentes, fournissant des données ayant une représentativité limitée. Par conséquent, **donnez toujours des instructions claires** sur le moment adéquat pour collecter les données, comment sélectionner les participants à l'enquête et que faire lorsque les participants prévus ne sont pas disponibles. De plus, utilisez les checklists CAQV d'IndiKit pour contrôler et améliorer la qualité du travail des enquêteurs.

CALCUL DE LA TAILLE DE L'ÉCHANTILLON

Les enquêtes de base, finales et autres nécessitant des données quantitatives précises doivent utiliser un échantillon représentatif de répondants. Sa taille requise peut être facilement calculée en utilisant le [Calculateur de Taille d'Echantillon](#) en ligne d'IndiKit. Les seules informations que vous devez entrer sont :

- **La Marge d'Erreur** (également appelée intervalle de confiance) est la quantité d'erreur que les résultats de votre enquête peuvent admettre. Par exemple, si votre échantillon utilise une marge d'erreur de 5% et que votre enquête montre que 63% des femmes allaitent exclusivement, vous pouvez être sûr que si vous aviez interrogé toutes les femmes allaitantes dans la population donnée, le résultat serait entre 58% (63-5) et 68% (63+5). Plus la marge d'erreur choisie est faible, plus la taille de l'échantillon sera grande. Alors que la marge d'erreur la plus courante est de 5%, les enquêtes axées sur des données qui ne devraient enregistrer que de faibles changements (tels que la prévalence de la malnutrition) n'utilisent souvent que 2-3%
- **La Taille de la Population** est le nombre total de personnes à partir duquel vous choisissez votre échantillon aléatoire. Cela peut, par exemple, être le nombre total de vos bénéficiaires ou le nombre de femmes vivant dans un district donné.
- **Le Niveau de Confiance**, vous indique à quel point vous pouvez être sûr de vos résultats. Comme il est très rare que les enquêtes d'aide humanitaire et de développement utilisent autre chose qu'un intervalle de confiance de 95%, le [Calculateur de Taille d'Echantillon](#) d'IndiKit entre automatiquement cette valeur.

Une erreur fréquente se produit lorsque les enquêtes incluent des **questions qui peuvent seulement être répondues par une partie des répondants**. Par exemple, alors que la population cible d'une enquête pourrait être les mères ayant des enfants âgés de 0 à 23 mois, certaines des questions peuvent se concentrer uniquement sur les mères ayant des enfants âgés de 0 à 6 mois ou de 12 à 15 mois. Comme seule une petite partie des répondants à l'enquête a des enfants de cet âge, le nombre de réponses sera relativement faible, ce qui se traduira par une précision limitée des résultats de l'enquête (augmentant souvent la marge d'erreur jusqu'à 10%). Dans un tel cas, pour que l'enquête fournisse des données suffisamment précises, vous devrez augmenter la taille de l'échantillon de manière significative.

Taille de l'Echantillon pour les Enquêtes de Suivi

Les enquêtes de suivi sont généralement des enquêtes à moyen terme et / ou finales. Alors que la plupart d'entre elles nous permettent d'utiliser un nouvel échantillon de répondants (suivant la même méthodologie que l'enquête de base), certaines enquêtes demandent à être conduite à nouveau sur les mêmes répondants qui ont participé à l'enquête de base. Cependant, ces répondants peuvent ne pas être disponibles ou refuser de participer à l'enquête de suivi. Cela se traduit par ce qu'on appelle **l'attrition de l'échantillon**, parfois aussi appelée **perte lors du suivi** – le pourcentage de répondants qui ont participé à l'enquête de base mais qui n'ont pas participé à l'enquête à moyen terme ou finale. La perte lors du suivi est calculée comme suit :

$$\text{perte lors du suivi} = \frac{\text{nombre de répondants enquête finale}}{\text{nombre de répondants enquête de base}} \times 100$$

Une grande perte lors du suivi peut biaiser les résultats, car les données de base et finales ne sont pas directement comparables. La « **règle 5-20** » peut être utilisée pour interpréter la validité des résultats. Cette règle stipule que :

- si moins de 5% des répondants de l'enquête de base sont perdus, la perte n'entraîne probablement qu'un impact minime sur la validité des résultats ; et
- si plus de 20% des répondants de l'enquête de base sont perdus, la perte menace la validité des résultats (dans ce cas, il est recommandé d'être prudent pour tirer des conclusions basées sur les résultats obtenus)

> Avez-vous une suggestion pour améliorer le contenu de ce Guide Rapide? [Envoyez-la nous, s'il vous plaît !](#)

> Voulez-vous que ce Guide Rapide soit disponible dans une autre langue ? [Contactez-nous !](#)

Auteur : Petr Schmied, [People in Need](#), 2018

RESSOURCES UTILISEES POUR PREPARER CE GUIDE:

- PIN (2014) Data Collection: A Practical Guide to Collecting Data
- PIN (2014) Impact Evaluation: A Practice Guide to Designing and Administering Impact Evaluations
- FAO, [Analysing the Data and Reporting the Results](#)