

### 3. Notions d'échantillonnage

Il y a plusieurs méthodes d'échantillonnage possibles et la méthode choisie aura un impact direct sur l'exactitude des statistiques produites.

Les enquêtes épidémiologiques sont souvent réalisées sur un échantillon de la population cible.

#### 3.1. Définition

**L'échantillonnage** (ou **sampling** en anglais) est le processus de sélection d'un groupe d'individus qui va être interrogé dans le cadre d'une étude et qui symbolise une population de référence. Il permet de mener des enquêtes à grand échelle en utilisant un **échantillon de la population pour remplacer l'ensemble** et ainsi mener le sondage de manière réaliste.

Quelques définitions des termes les plus fréquemment rencontrés :

- **Population**: cible de l'étude ou nombre total d'individus sur qui porte la recherche
- **Échantillon**: sous-ensemble de la population qui représente la totalité des gens que vous allez étudier
- **Échantillonnage**: méthode et processus de sélection de l'échantillon

#### 3.2. Enjeux et applications

Il est très facile de comprendre l'intérêt du **sampling** pour réaliser des recherches à très grande échelle. Mais il faut savoir que ce processus se révèle très pertinent à utiliser pour tout type d'étude, ciblant des populations de toute taille. **L'échantillonnage permet en effet de rationaliser les efforts et ressources liés à une étude.** En délimitant la recherche vous êtes, de fait, moins contraint par les coûts, le timing ou la complexité des analyses.

C'est d'ailleurs grâce aux échantillonnages que nous sommes capables de mener des enquêtes sur la propagation épidémique ou d'effectuer des recherches à l'échelle nationale .

Il est donc important de bien comprendre les différences entre les différents plans d'échantillonnage possibles. Ceux-ci se séparent en deux types : l'échantillonnage **probabiliste** et l'échantillonnage **non probabiliste**.

#### 3.3. La méthode d'échantillonnage

La différence entre les méthodes d'échantillonnage probabiliste et non probabiliste, tient au fait que dans le cas de l'échantillonnage probabiliste chaque unité a une probabilité d'être sélectionnée qui peut être quantifiée, ce qui n'est pas vrai pour l'échantillonnage non probabiliste. La stratégie diffère en fonction des disciplines étudiées, des domaines de recherche et du type d'étude mener.

## Les deux types d'échantillonnage : probabiliste et non probabiliste

Voici quelques explications simples sur ces deux catégories :

- **L'échantillonnage probabiliste ou aléatoire:** fait référence à la sélection d'un échantillon d'une population lorsque cette sélection repose sur le principe de la randomisation, c'est-à-dire la sélection au hasard ou aléatoire. Il est plus complexe, plus long à mettre en œuvre et habituellement plus dispendieux que l'échantillonnage non probabiliste. Toutefois, comme les unités de la population sont sélectionnées au hasard et qu'il est possible de calculer la probabilité de sélection de chaque unité dans l'échantillon, il permet de produire des estimations fiables et de faire des inférences statistiques au sujet de la population.
- **L'échantillonnage non probabiliste:** L'échantillonnage non probabiliste est une méthode qui consiste à sélectionner des unités dans une population en utilisant une méthode subjective (c'est-à-dire non aléatoire). Comme l'échantillonnage non probabiliste ne nécessite pas de base de sondage complète, c'est un moyen rapide, facile et peu coûteux d'obtenir des données. Cependant, pour pouvoir tirer des conclusions sur la population à partir de l'échantillon, il faut supposer que l'échantillon est représentatif de la population. Il s'agit souvent d'une hypothèse risquée dans le cas d'un échantillonnage non probabiliste, car il est difficile d'évaluer si l'hypothèse est valable ou non. De plus, comme les éléments sont choisis arbitrairement, il n'y a aucun moyen d'estimer la probabilité qu'un élément soit inclus dans l'échantillon. De même, rien ne garantit que chaque élément a une chance d'être inclus, ce qui rend impossible l'estimation de la variabilité de l'échantillonnage ou l'identification d'un éventuel biais.

Pour chacun de ces deux types, il existe **plusieurs méthodes d'échantillonnage**.

### 3.4. Les méthodes d'échantillonnage probabilistes

#### A- Échantillonnage aléatoire simple

Chaque élément de la population de référence a une **chance égale d'être sélectionné pour constituer le panel d'enquête dans ce processus d'échantillonnage**. La sélection est donc parfaitement aléatoire.

Cette **méthode probabiliste** peut être effectuée en anonymisant les membres de la population, puis en attribuant un numéro à chacun et enfin en choisissant les numéros au hasard.

### **Exemple n° 1:**

Pour prélever un échantillon aléatoire simple d'un annuaire téléphonique, il faudrait numéroté en ordre séquentiel chaque entrée ou inscription. S'il y avait 10 000 entrées dans l'annuaire téléphonique et si la taille de l'échantillon était de 2 000 numéros, un ordinateur devrait alors générer au hasard 2 000 numéros entre 1 et 10 000. Tous les numéros auraient la même chance d'être générés par l'ordinateur. Les 2 000 entrées de l'annuaire téléphonique correspondant aux 2 000 numéros aléatoires générés par l'ordinateur composeraient l'échantillon.

C'est la méthode d'échantillonnage la plus couramment utilisée. L'avantage de cette technique tient au fait qu'elle n'exige pas d'autres données dans la base de sondage que la liste complète des membres de la population observée et l'information pour les contacter.

**Cette technique d'échantillonnage est très simple à réaliser et peu coûteuse.** Elle élimine également tout biais potentiel du processus de sampling. Cependant, étant donné son caractère aléatoire, elle peut aboutir à la constitution d'un groupe de répondants non représentatifs de la population de référence.

### **B- Échantillonnage systématique**

Également connue sous le nom de regroupement systématique, cette **méthode d'échantillonnage** consiste à effectuer une sélection qui n'a d'aléatoire que le premier élément choisi. Une règle définie par les chercheurs s'applique alors pour sélectionner les individus suivants.

L'**échantillonnage systématique** signifie qu'il existe un écart, ou un intervalle, entre chaque unité sélectionnée dans l'échantillon.

Bien que **cette technique comprenne de la randomisation**, le chercheur possède ici la capacité de choisir l'intervalle auquel les membres du sous-ensemble sont sélectionnés. Cela lui permet d'effectuer une sélection plus éclairée et précise.

Elle est souvent utilisée dans l'industrie, où l'on sélectionne une unité pour des essais dans une chaîne de production afin de s'assurer que la machinerie et l'équipement sont d'une qualité uniforme. Un testeur dans une usine pourrait, par exemple, soumettre à un contrôle de la qualité chaque 20<sup>e</sup> produit sur une chaîne de montage, en commençant par un point initial choisi au hasard entre 1 et 20.

### **Exemple n° 2**

Les intervieweurs peuvent mettre en œuvre cette technique d'échantillonnage lorsqu'ils interrogent des gens pour une enquête-échantillon. Le responsable d'une étude de marché pourrait sélectionner, par exemple, chaque 10<sup>e</sup> personne qui entrerait dans un commerce, après avoir sélectionné au hasard la première personne. Un enquêteur pourrait interviewer les

occupants de chaque 5e maison d'une rue, après avoir sélectionné au hasard l'une des cinq premières maisons.

Les avantages de l'échantillonnage systématique tiennent au fait que la sélection de l'échantillon ne peut être plus facile : vous n'obtenez qu'un seul nombre aléatoire, l'origine choisie au hasard, et le reste de l'échantillon suit automatiquement. Le plus gros inconvénient de la méthode tient au fait que les échantillons possibles risquent de ne pas être représentatifs de la population s'il existe un trait périodique dans l'ordre d'apparition des unités sur la base de sondage et que ce trait périodique coïncide d'une quelconque façon avec l'intervalle d'échantillonnage.

### **C- Échantillonnage stratifié**

Lorsque l'on utilise l'**échantillonnage stratifié**, on divise la population en groupes homogènes appelés strates qui sont mutuellement exclusifs, puis on sélectionne dans chaque strate des échantillons indépendants. N'importe laquelle des méthodes d'échantillonnage mentionnées dans la présente section peut être utilisée pour sélectionner l'échantillon à l'intérieur de chaque strate. La méthode d'échantillonnage peut être différente d'une strate à une autre. Toute variable pour laquelle on dispose d'une valeur pour la totalité des unités incluses dans la base de sondage (comme l'âge, le sexe, la province de résidence, le revenu, etc.) peut être utilisée pour mettre en œuvre la stratification.

Un des principales avantages est que l'échantillonnage stratifié assure d'obtenir une taille d'échantillon suffisante pour des sous-groupes d'intérêt de la population. Étant donné que chaque strate devient une population indépendante, une taille d'échantillon est déterminée pour chacune d'entre elles.

#### **Exemple n° 3**

Si vous effectuez une recherche sur les habitudes de transport d'un certain ensemble de personnes, il peut être judicieux de séparer les répondants qui possèdent une voiture de ceux qui n'en ont pas. Le fait de stratifier votre liste permet d'obtenir la précision souhaitée pour les deux catégories

La stratification est très utile lorsque les variables de stratification sont :

- simples à utiliser,
- faciles à observer,
- étroitement reliées au thème de l'enquête.

Cette **technique d'échantillonnage probabiliste présente plusieurs avantages** mais implique un risque de biais d'enquête, notamment au moment de la stratification.

## **D- Echantillonnage en grappe**

Il est parfois trop dispendieux d'avoir un échantillon réparti sur l'ensemble du territoire. Les coûts de déplacement risquent de devenir élevés si les intervieweurs doivent sonder des gens d'un bout à l'autre du pays. Les statisticiens peuvent choisir la technique de l'échantillonnage par grappes pour réduire les coûts.

La technique de l'échantillonnage par grappes implique la division de la population en groupes ou en grappes, comme son nom l'indique. Suivant cette technique, un certain nombre de grappes est sélectionné au hasard, puis toutes les unités incluses à l'intérieur des grappes sélectionnées constituent l'échantillon. Aucune unité des grappes non sélectionnées ne fait partie de l'échantillon. Elles sont représentées par les unités des grappes sélectionnées. Rappelons que dans un échantillon stratifié, des unités sont sélectionnées dans toutes les strates. C'est donc l'une des différences entre les deux approches. Entre autres exemples de grappes qui peuvent être utilisées, il y a les usines, les établissements d'enseignement et les régions géographiques telles que les subdivisions électorales.

### **Exemple n° 4**

Supposons que vous représentez une organisation d'athlétisme désirant déterminer quels sports pratiquent les élèves de 2ème années secondaire en Algérie. Il serait trop dispendieux et trop long d'interroger chaque élève 2ème années secondaire ou même deux ou trois élèves de chaque classe. Vous pourriez plutôt sélectionner au hasard 100 écoles dans tout le pays. Ces 100 écoles seraient les grappes échantillonnées. Tous les élèves de secondaire 4 de chacune des 100 écoles pourraient alors être interrogés.

L'échantillonnage par grappes ne permet pas de contrôler totalement la taille finale de l'échantillon, ce qui constitue un inconvénient de son utilisation.

## **E. Echantillonnage à plusieurs degrés**

La méthode d'échantillonnage à plusieurs degrés ressemble à la méthode d'échantillonnage par grappes, sauf qu'un échantillon est prélevé à l'intérieur de chaque grappe sélectionnée. Il y a alors au moins deux degrés. Identification et sélection des grappes au premier degré, suivi d'une sélection des unités au deuxième degré à l'aide de n'importe quelle autre méthode d'échantillonnage. Dans ce contexte, les grappes sont parfois désignées comme les unités primaires d'échantillonnage (UPE) et les unités de la population comme les unités secondaires d'échantillonnage (USE). Lorsque plus de deux degrés sont utilisés, une sélection supplémentaire d'unités tertiaires d'échantillonnage (UTE) est réalisée à l'intérieur des USE, et ainsi de suite jusqu'à l'obtention d'un échantillon final.

### 3.5. Les méthodes d'échantillonnage non probabilistes

#### A- Echantillonnage de commodité

Ici, les personnes ou éléments des échantillons sont sélectionnés en fonction de leur disponibilité. Par exemple, si vous effectuez une enquête auprès d'étudiants sur un campus universitaire, **opter pour l'échantillonnage de commodité** revient à interroger les apprenants qui se trouvent sur le campus, qui ont du temps libre et qui sont disposés à répondre à votre questionnaire.

Ce **type de sampling non probabiliste** peut avoir de la valeur mais comporte un biais important. Il est particulièrement utile en tant qu'étape précoce ou préliminaire à une future enquête plus précise.

**Par exemple:** les études d'opinion réalisées dans la rue

- **Les avantages**
  - la facilité d'application
  - l'absence d'influence de l'investigateur.
- **Les inconvénients:**
  - La non représentativité
  - L'impossibilité d'évaluer le biais associé.

#### B. Échantillonnage à participation volontaire

Dans cette méthode, les répondants sont uniquement des volontaires. En général, les volontaires doivent faire l'objet d'un examen pour obtenir un ensemble de caractéristiques adaptées aux objectifs de l'enquête (par exemple, des personnes atteintes d'une maladie particulière). Cette méthode peut être sujette à d'importants biais de sélection, mais elle est parfois nécessaire. Par exemple, pour des raisons éthiques, il peut être nécessaire de solliciter des volontaires présentant des conditions médicales particulières pour certaines expériences médicales.

- **Les avantages:**
  - Attractif du point de vue de l'éthique
  - utile pour les phases exploratoires.
- **Les inconvénients:**
  - La non représentativité
  - L'impossibilité d'évaluer le biais associé.

#### C- Echantillonnage par quotas

Cette approche vise à obtenir une subdivision de la population (tout comme la méthode probabiliste stratifié) en précisant qui doit **être recruté pour l'enquête sur la base de divers critères**.

Par exemple, votre quota peut inclure un certain nombre d'hommes et un certain nombre de femmes, ou d'individus d'une certaine tranche d'âge ou d'une habitude de consommation spécifique.

Il s'agit de l'une des formes les plus courantes d'échantillonnage non probabiliste. L'échantillonnage est effectué jusqu'à ce qu'un nombre déterminé d'unités (quotas) pour diverses sous-populations soient sélectionnées. L'échantillonnage par quotas est un moyen de satisfaire les objectifs de taille d'échantillon pour les sous-populations.

Les quotas peuvent être basés sur les proportions de la population. Par exemple, si la population compte 100 hommes et 100 femmes, et il faut tirer un échantillon de 20 personnes, 10 hommes et 10 femmes peuvent être interviewés. L'échantillonnage par quotas peut être considéré comme préférable à d'autres formes d'échantillonnage non probabiliste (par exemple, l'échantillonnage au jugé), car il oblige à inclure des membres de sous-populations différentes.

- **L'avantages:**
  - Ne nécessite pas de base de sondage mais uniquement la connaissance de la répartition dans la population selon certaines caractéristiques
  - Le coût est plus faible: personne précise non obligatoire
- **Les inconvénients:**
  - La non représentativité
  - L'impossibilité d'évaluer le biais associé.
  - Non connaissance des taux de réponse

#### **D- Echantillonnage raisonné (au jugé)**

Avec cette méthode, l'échantillonnage est fait en tenant compte des idées préalables sur la composition et le comportement de la population. Un expert ayant une connaissance de la population décide quelles unités de la population doivent être choisies. En d'autres termes, l'expert sélectionne délibérément ce qu'il considère comme un échantillon représentatif.

L'échantillonnage au jugé est soumis aux biais du chercheur et est peut-être encore plus biaisé que l'échantillonnage de commodité.

Dans cette **méthode d'échantillonnage non probabiliste**, les participants sont choisis par les chercheurs en fonction de :

- Leurs connaissances
- Leur compréhension de la question de recherche
- Ou leurs objectifs

Puisque toutes les idées préconçues du chercheur se reflètent dans l'échantillon, des biais importants peuvent être intégrés si ces idées préconçues sont inexactes. Cependant, il peut être utile dans les études exploratoires, par exemple pour sélectionner des membres de groupes de discussion ou pour mener des entrevues approfondies afin de tester des aspects spécifiques d'un questionnaire.

Cette technique est également connue sous le nom de **sampling par jugement** et a peu de chance de mener à la création d'un échantillon représentatif. Cependant, c'est une méthode facile et rapide à mettre en œuvre.

#### **E- Echantillonnage de réseaux**

Cette **méthode d'échantillonnage se révèle très utile** lorsque l'équipe de chercheurs n'a pas de connaissances approfondies de la population cible, et donc peu de moyen de recruter de nombreux participants.

Supposons qu'un chercheur souhaite trouver des individus possédant un trait rare dans la population, qu'il connaisse déjà l'existence de certains d'entre eux et sache comment les contacter. Une approche consiste à contacter ces personnes et à leur demander simplement si elles connaissent quelqu'un comme elles, puis à contacter ces personnes, etc. L'échantillonnage est utile pour les populations rares ou difficiles à atteindre, comme les personnes handicapées, les sans-abri, les toxicomanes ou d'autres personnes qui n'appartiennent pas à un groupe organisé

Par contre, cette technique introduit un biais car cela **favorise une homogénéité de l'échantillon**. Certains groupes seront plus représentés que d'autres et des personnes plutôt isolées ne seront pas du tout interrogées.