

Chapitre 1 : Aperçu sur les techniques d'échantillonnage

Année Universitaire : 2025/2026

1-1 Introduction

L'échantillonnage permet aux statisticiens ou aux chargés d'étude de tirer des conclusions au sujet d'un tout, en n'en examinant qu'une partie. Les chercheurs ne s'intéressent pas 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.

1-2 Quelques concepts de bases

- On appelle **enquête** l'ensemble des opérations qui ont pour but de collecter de façon organisée des informations relatives à un groupe d'individus, d'éléments ou d'unités statistiques.
- Les **individus** ou les **éléments** en question, appelées aussi **unité statistique** peuvent être aussi bien des personnes humaines que des animaux, des familles, machines, entreprises, etc.
- L'ensemble des individus auxquelles on s'intéresse est appelé **population** ou **univers**.
- Lorsque toutes les unités de la population sont observées individuellement, l'enquête est appelée **recensement**.
- La partie de la population qui est réellement observée constitue l'**échantillon**, et l'opération de choix de cette population est précisément l'opération d'**échantillonnage**.
- Une **base d'échantillonnage** ou une **base de sondage** est une liste des unités d'échantillonnage possibles.

1-3 Les caractéristiques d'un échantillon

L'échantillon ou la population (cible) peut être différente selon le problème étudié. Elle est caractérisée par les critères suivants :

a- L'objectif de l'enquête

La population peut être définie dans l'objet même de l'enquête.
Exemple : Enquête auprès des consommateurs sur le prix du Yaourt.

b- Définir la population cible

C'est la population totale pour laquelle on a besoin de l'information. Il faut définir les unités qui composent la population sous forme de caractéristiques l'identifiant (nature des données, emplacement géographique, dates ou encore critères sociodémographiques...). Cependant, la *population observée* est différente de la *population cible*. En effet, la population cible est la population que nous **voulons observer**, tandis que la population observée est la population que nous **pouvons observer**. Et les conclusions ne s'appliqueront qu'à la population réellement observée (cible). L'utilisateur des résultats doit en être informé.

c- Le type d'échantillonnage adopté

La nature des documents disponibles pour construire l'échantillon peut amener à restreindre la population à interroger.

Exemple : si on analyse les listes électorales ; l'échantillon est limité aux personnes dont l'âge dépasse 18 ans (âge de voter).

d- Les contraintes matérielles imposées

Pour des raisons budgétaires ou de délais d'exécution, la population peut être restreinte à une portion plus accessible.

Exemple : Dans une enquête sur l'opinion des jeunes à l'égard du code de la route, on limitera la population aux jeunes qui possèdent un permis de conduire.

1-4 Les méthodes d'échantillonnage

À l'exception des recensements qui se réalisent auprès de l'ensemble des individus composant la population, toutes les autres enquêtes se limitent à interroger un échantillon (une partie) de la population.

L'échantillon doit être une partie représentative de la population et refléter, dans sa composition, la diversité réelle de la population.

Toute la question est de savoir si cet échantillon est représentatif ou pas de cette population. Si ce n'est pas le cas, l'enquête réalisée ne sera pas représentative et ne présentera aucune validité externe.

Pour cela ; on fait appel aux méthodes d'échantillonnage pour sélectionner les individus représentatifs d'une population. Il existe principalement les *méthodes aléatoires ou probabilistes* et les *méthodes empiriques*. La différence entre les deux tient au fait que dans

le cas de l'échantillonnage probabiliste chaque unité a une « chance » d'être sélectionnée et que cette chance peut être quantifiée, ce qui n'est pas vrai pour l'échantillonnage non probabiliste ; dans ce cas, chaque unité incluse à l'intérieur d'une population n'a pas une chance égale d'être sélectionnée.

1-4-1 Méthodes aléatoires ou probabilistes

Tous les membres de l'univers (population) ont une chance égale de faire partie de l'échantillon. Plusieurs méthodes existent :

a. Échantillon au hasard simple ou "Sondage aléatoire Simple"

Dans un échantillon aléatoire simple, chaque élément de la population a la même chance d'être choisi. En outre, chaque échantillon possible d'une taille « n » a la même probabilité d'être sélectionné. Cela implique que chaque élément est choisi indépendamment de tout autre élément. Les unités qui constituent l'échantillon sont désignées par tirage au hasard.

Le chargé d'étude doit dresser une liste de toutes les unités incluses dans la population observée (base de sondage ou d'échantillonnage) pour sélectionner un échantillon aléatoire simple. Un échantillonnage aléatoire simple peut s'effectuer avec remise (tirage non exhaustif) ou sans remise (tirage exhaustif).

Avantages de cette méthode :

- On peut espérer à un échantillon « représentatif » puisque la méthode donne à chaque individu de la population une chance égale d'être sélectionné ;
- Facile à mettre en œuvre.

Inconvénients :

- La difficulté dans la construction de la base de sondage apte à constituer un échantillon aléatoire simple ;
- Augmentation du temps et le coût de collecte des données, si la population est très dispersée géographiquement, de ce fait les frais de déplacement des enquêteurs seront élevés.
- Elle ne permet pas toujours d'obtenir un échantillon représentatif. Bien qu'en moyenne, les échantillons constitués représentent bien la population, il arrive qu'un échantillon aléatoire simple donné présente une image grossièrement déformée de la population ciblée.

b. Échantillonnage systématique

L'échantillonnage systématique est une méthode qui exige aussi l'existence d'une liste de la population où chaque individu est numéroté de 1 jusqu'à N .

Notons n , le nombre d'individus que doit comporter l'échantillon (la taille de l'échantillon). L'entier voisin de N/n sera noté « r » et appelé « raison de sondage » ou « pas de sondage ».

Choisir au hasard un entier naturel d entre 1 et r (cet entier sera le point de départ). L'individu dont le numéro correspond à d est le premier individu, pour sélectionner les autres, il suffit d'ajouter à d la raison de sondage : les individus choisis seront alors ceux dont les numéros correspondent à : $d + r$; $d + 2r$; $d + 3r$, etc.

Avantages : facile à sélectionner parce qu'un seul individu est choisi au hasard. On peut obtenir une bonne précision parce que la méthode permet de répartir l'échantillon dans l'ensemble de la liste.

Inconvénients : dans le cas où le pas de sondage égale à la périodicité des données. Les données peuvent être biaisées à cause de la périodicité.

Exemple : Étude des déplacements en autobus sur 365 jours en prenant un échantillon de taille 60.

($N = 365$ jours et $n = 60$).

Dans ce cas : $r = 365/60 \approx 7$.

Donc, le jour sélectionné au départ sera le même pour tout l'échantillon.

Exemple 2 : On a une population de 400 individus, on veut un échantillon de 100 individus

$R = 4$

On a donc 4 échantillons possibles :

- 1, 5, 9, 397
- 2, 6, 10, ... 398
- 3, 7, 11, ... 399
- 4, 8, 12, ... 400

c. Échantillonnage stratifié

Une strate est un groupement homogène d'unités statistiques ; triées de la population totale et reliées entre elle par un caractère lié à l'objet de l'enquête.

L'univers statistique est découpé en sous-ensembles homogènes. La stratification se fait selon les problèmes étudiés, par exemple :

- Classement des entreprises suivant le personnel.
- Classement des ménages suivant leurs revenus ou le nombre de personne qui les composent.
- Classement de la population suivant les communes.

Soit :

- N : la taille de la population
- N_i : la taille de la strate « i », (i allant de 1 à k).
- K : le nombre de strates
- n_i : l'échantillon lié à la strate : $n_i\% = \frac{N_i}{N} \times 100$

Exemple : Une entreprise possède 3000 clients. Ils sont repartis en 2 strates : 2100 femmes et 900 hommes. On vérifie que l'échantillon total est reparti proportionnellement à l'importance de chaque strate dans la population.

	Population N_i	Taille de l'échantillon ($n_i\%$)
Strate 1 (femmes)	2100	70
Strate 2 (hommes)	900	30
Total	3000	100

TABLE 1 – Répartition des clients d'une entreprise selon le genre

Pour un échantillon de taille 300, il doit être composé de 210 femmes et de 90 hommes.

Avantages :

Il est peu probable de choisir un échantillon absurde puisqu'on s'assure de la présence proportionnelle de tous les divers sous-groupes composant la population.

Inconvénients :

La méthode suppose l'existence d'une liste de la population (base de sondage). Il faut aussi connaître comment cette population se répartit selon certaines strates, c'est-à-dire les critères à adopter pour constituer les strates.

d. Échantillon en grappes

L'échantillonnage en grappe limite la portée de la construction de la base de l'échantillon et des travaux de terrain connexes à un ensemble ou un échantillon, surtout quand la population est dispersée géographiquement. Cette méthode, simple et économique, permet de choisir un échantillon en plusieurs étapes.

Tout d'abord, on divise la population en sous-groupes (grappes) mutuellement exclusifs (chaque individu n'appartient qu'à une seule grappe) et collectivement exhaustif (la somme des éléments de toutes les grappes donnera la population).

Ensuite, choisir au hasard un échantillon en grappes en s'appuyant sur une des techniques d'échantillonnage probabiliste comme l'échantillonnage aléatoire simple. Pour chaque grappe sélectionnée, soit on inclut tous les éléments dans l'échantillon, il s'agit dans ce cas d'un échantillonnage en grappes *simple*, soit on extrait un échantillon d'éléments de façon probabiliste, la procédure est dite d'échantillonnage à deux degrés.

Avantages :

- Réduction des coûts de collecte de données.

Inconvénients :

- Effet de grappe (variance intra-grappes qui est faible) dû à l'existence de similarité entre individus d'une même grappe.

1-4-2 Méthodes empiriques ou non probabilistes

La méthode d'échantillonnage non-probabiliste est utilisée lorsqu'il n'est pas possible de constituer une liste exhaustive de toutes les unités du sondage.

Dans le cas de l'échantillonnage probabiliste, chaque unité a une chance d'être sélectionnée. Ce n'est plus le cas dans l'échantillonnage non probabiliste. On se fixe alors comme règle que l'échantillon retenu doit avoir la même composition que la population mère par rapport à une ou plusieurs caractéristiques.

Elles ne font plus appel au hasard, et elles sont plutôt empiriques. Il existe, en effet plusieurs méthodes :

a. Échantillon par choix raisonné

Cette méthode consiste à procéder par un choix raisonné, c'est à dire, le chargé d'étude choisit les unités les plus représentatives de la population. Plusieurs méthodes sont utilisées :

Méthode des quotas En pratique, c'est la méthode la plus utilisée. Elle consiste à choisir un ensemble de critères appelés « caractères de contrôle » et construire un modèle

réduit de la population totale. Les caractères de contrôle sont, par exemple : revenu, âge, sexe, catégorie socioprofessionnelle, etc.

Les avantages :

- Elle n'exige pas une énumération complète des unités de l'univers.
- Les unités à relever étant choisis par les enquêteurs.

Les inconvénients :

- La méthode ne permet pas le calcul des erreurs, seul les échantillons aléatoires le permettent.
- Les résultats peuvent être entachés d'erreurs systématiques introduite suite à une faute de jugement des enquêteurs (biais de sélection).
- La qualité des enquêtes repose sur la qualité du travail des enquêteurs.

Méthode de convenance Pour faciliter la recherche des personnes à interroger, on peut décider de les interroger là où il y a le plus de chance de trouver les unités directement concernées par le sujet de l'enquête. Par exemple : des personnes interceptées dans la rue, dans des magasins ou dans des centres commerciaux...

Exemple 1 : un journaliste qui réalise un micro-trottoir à l'occasion d'une manifestation.

Exemple 2 : Si un enquêteur doit interroger des acheteurs de matelas, il pourra se placer à la sortie des magasins de literie.

En effet, cette manière de faire n'interroge pas nécessairement la population cible mais aussi des personnes qui ne soit pas concernées par l'étude.

Échantillon de jugement Ce type d'échantillon se veut représentatif dans la mesure où le chercheur va interroger les individus les plus susceptibles d'éclairer et d'apporter une information pertinente sur le problème à résoudre.

b. Échantillon pseudo-aléatoire

Il s'agit de méthodes qui sans être des méthodes probabilistes s'en rapprochent le plus. Elle consiste à imposer à l'enquêteur un itinéraire en lui indiquant exactement les points du circuit. L'identification de ces points d'enquête résulte de la combinaison de tirage au hasard.

1-5 Procédure de choix d'une méthode d'échantillon

Il serait faux de dire que toutes les méthodes d'échantillonnage se valent. Toutefois, nous sommes souvent restreints à choisir une méthode. Ce choix peut être résumé par la figure suivante :

1-6 Détermination de la Taille de l'Échantillon

Combien de personnes doit-on interroger pour que l'échantillon soit représentatif de la population étudiée ? Autrement dit :

1. Est-ce qu'un échantillon de taille 500 suffit pour une population de taille 10 000 ?
2. Quelle est la taille de l'échantillon qui assure tel degré précision ?

Il est impossible de répondre par oui ou par non à ces questions. Un échantillon doit fournir une estimation aussi précise que possible d'une variable.

La taille de l'échantillon dépend généralement des critères suivants :

- Taille de la population d'étude
- Critères statistiques
- Contraintes de terrain observées et des questions auxquelles on désire répondre par le moyen de l'enquête

1-6-1 La taille de la population

Pour déterminer la taille de l'échantillon il est important de connaître la taille de population. En effet, plus la taille de la population est grande, plus la taille de l'échantillon sera grande afin de tenir compte des caractéristiques de la population.

1-6-2 Critères statistiques

- **Niveau de précision (e)** : Marge d'erreur acceptée (ex : 5%)
- **Niveau de confiance** : Probabilité que l'échantillon représente fidèlement la population (ex : 95%)
- **Degré de variabilité (p)** : Mesure de l'hétérogénéité de la population ($p = 0,5$ pour la variabilité maximale)

1-6-3 Calcul de la taille de l'échantillon

Lorsqu'on se limite aux critères cités précédemment, le calcul de la taille de l'échantillon sera déterminé comme suit :

0.0.1 Petites populations (<200)

Il est souvent préférable d'enquêter toute la population (recensement).

0.0.2 Tableaux statistiques

Pour des critères standards (confiance 95%, $p = 0,5$), des tableaux donnent directement la taille de l'échantillon.

0.0.3 Formules mathématiques

0.0.4 Formule simplifiée (sans variabilité)

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

0.0.5 Formule pour les proportions (population infinie)

$$n_0 = \frac{t^2 \times p(1-p)}{e^2} \quad (2)$$

où $t = 1,96$ pour un niveau de confiance de 95%

0.0.6 Formule pour les proportions (population finie)

$$n = \frac{n_0}{1 + \left(\frac{n_0}{N}\right)} \quad (3)$$

Conclusion

Pour déterminer la taille d'un échantillon, on combine des calculs statistiques (basés sur la précision, la confiance et la variabilité souhaitées) avec une évaluation réaliste des contraintes pratiques du terrain.