

Les méthodes d'échantillonnage

Objectif

L'étudiant doit être capable d'appliquer le principe de sélection d'un échantillon à l'aide d'une méthode aléatoire ou d'une méthode non aléatoire.

Introduction

Pour effectuer une étude statistique (enquête, observation ou expérimentation), on se sert généralement d'un échantillon. Celui-ci doit refléter le plus exactement possible l'image de la population. En fait, choisir un échantillon, c'est mettre ensemble un certain nombre d'individus qui composeront une sorte de modèle réduit de la population à laquelle ils appartiennent. Mais comment choisir ces individus? C'est à cette question que nous répondrons dans ce chapitre, en présentant les méthodes d'échantillonnage les plus courantes.

Echantillonner, c'est choisir une partie d'une population pour représenter l'ensemble de la population.

Si vous travaillez sur un échantillon, c'est pour chercher une information sur l'ensemble de sa population mère. Les résultats observés sur cet échantillon n'auront de sens que s'ils sont rapportés à toute la population.

On a donc recours à des techniques bien précises pour construire un échantillon et s'assurer qu'il est représentatif de sa population mère.

Il existe deux grandes catégories de méthodes.

- **les méthodes d'échantillonnage non probabilistes (ou empiriques),**
- **les méthodes d'échantillonnage probabilistes (ou aléatoires).**

Section I- Les méthodes d'échantillonnage probabilistes

Elles regroupent les techniques qui font intervenir le hasard pour désigner les éléments de l'échantillon. Elles nécessitent une base de sondage (liste de tous les individus de la population mère), sont plus coûteuses et longues que les méthodes empiriques. Les échantillons obtenus par ces techniques sont :

- ➔ les échantillons aléatoires simples
- ➔ les échantillons systématiques
- ➔ les échantillons par grappe
- ➔ les échantillons stratifiés

La technique d'échantillonnage aléatoire simple qui est la technique de base, dans cette catégorie.

1- L'échantillonnage aléatoire simple

Comme je vous l'ai indiqué précédemment cette méthode suppose l'établissement d'une base de sondage à partir de laquelle les éléments du sondage seront tirés au hasard.

Le principe de base est le suivant : tous les individus ou éléments constituant la population de base ont la même chance d'être choisis.

Il faudra donc, dans un premier temps établir la base de sondage qui est une liste exhaustive des individus de la population. Pour cela, on aura recours à des fichiers existants ou à des registres. Dans tous les cas de figure, il est très important que ces fichiers soient complets, à jour et adaptés à la population sur laquelle on veut enquêter.

A ce sujet, il est parfois tentant de vouloir utiliser des répertoires comme l'annuaire du téléphone ou les listes d'inscription électorale. Mais ce sont souvent de mauvaises bases de sondage. En effet, on évalue à 6% le nombre de personnes qui ne figurent pas dans l'annuaire. Pour ce qui est des listes électorales, elles sont souvent inexactes en raison du nombre des non-inscrits ou des personnes inscrites mais résidant ailleurs ou l'inverse etc.

Une fois la base de sondage établie il existe trois techniques :

- **le tirage de jetons numérotés ou de papiers avec le nom de chaque individu dans une urne**
- **la désignation de façon aléatoire de l'échantillon à l'aide d'un ordinateur**
- **la sélection de nombre identifiant les individus à l'aide d'une table de nombres aléatoires**

Pour une population restreinte, la première technique suffira largement; pour une population de grande dimension, il est préférable d'attribuer un numéro distinct à chaque individu ou élément de la base de sondage. Ceux qui composeront l'échantillon seront sélectionnés à l'aide d'un programme informatique ou d'une table de nombres aléatoires.

L'intervention du hasard n'est pas synonyme d'anarchie. Il obéit tout de même à des règles strictes qu'il faut respecter dans la constitution d'un échantillon.

Un échantillon construit suivant cette méthode, ne peut être modifié par les enquêteurs en cours d'enquête. En effet, l'enquêteur n'interrogera pas un profil, comme dans l'échantillon par quotas, mais une personne bien précise. Des moyens doivent être prévus pour procéder aux différentes relances nécessaires pour interroger les individus sélectionnés.

En règle générale, ces échantillons portent sur un grand nombre d'individus, environ 1000 ou plus.

Comment utiliser la table des nombres aléatoires : La table des nombres aléatoires est composée de 2 000 nombres compris entre 0 et 9. Par commodité de lecture, ces nombres y figurent par groupe de 5 chiffres, répartis sur 10 colonnes, par bloc de 5 lignes.

Pour utiliser cette table, il faut avant tout **attribuer un numéro à chaque individu** de votre population mère. Puis, vous allez vous fixer une technique de lecture de la table :

1) définir un point d'entrée

2) décider des chiffres que vous allez **retenir** pour choisir les éléments de votre échantillon deux premiers, deux derniers, un sur deux etc.

3) la direction à suivre pour trouver les autres chiffres : gauche à droite, droite à gauche, haut en bas, bas en haut ou diagonale.

Un chiffre tiré au hasard dans la table ne peut être retenu une seconde fois s'il ressort. Par ailleurs, si le chiffre de la table ne fait pas partie des numéros attribués à votre base de sondage il ne doit pas être pris en compte.

Exemple

On veut choisir par échantillonnage aléatoire simple, sans remise, 8 étudiants; parmi un groupe de 60. La sélection de L'échantillon sera faite à l'aide de la table de nombres aléatoires.

-1ère étape : attribuer un numéro à chaque étudiant de 01 à 60

La première étape consiste à numéroter chaque étudiant à l'aide de deux chiffres afin que chacun des étudiants ait la même chance d'être choisi: 01, 02, 03..., 59, 60.

-2ème étape : définition de la technique de lecture de la table

La deuxième étape consiste à choisir un processus de déplacement dans la table et à déterminer au hasard un premier nombre de deux chiffres (car il y a deux chiffres dans 60) qui sera le point de départ. L'échantillon sera alors constitué des huit premiers nombres de deux chiffres n'excédant pas 60, tout nombre apparaissant une deuxième fois étant éliminé. Supposons que l'on ait choisi comme point de départ de la table les deux premiers chiffres à l'intersection de la 21ème ligne et de la 2ème colonne. Supposons aussi que l'on ait décidé d'aller de haut en bas jusqu'à la fin de la colonne en ne prenant toujours que les deux premiers chiffres de chaque nombre, puis de reprendre le même processus en haut de la colonne suivante.

Point de départ : 21ème ligne, 2^{ème} colonne

Chiffre à retenir : les deux premiers chiffres

Sens de lecture : de haut en bas à droite

On obtient alors:

88 à rejeter car supérieur à 60

49 à retenir

49 à rejeter car échantillonnage sans remise

19 à retenir

66 à rejeter car supérieur à 60

01 à retenir

63 à rejeter car supérieur à 60

16 à retenir

73 à rejeter car supérieur à 60

93 à rejeter car supérieur à 60

06 à retenir

92 à rejeter car supérieur à 60

67 à rejeter car supérieur à 60

05 à retenir

10 à retenir

14 à retenir

Le tirage selon la table donnera le résultat suivant : 49 / 19 / 01 / 16 / 06 / 05 / 10 / 14. Les étudiants à qui ces chiffres auront été attribués composeront votre échantillon.

Les avantages et les inconvénients de l'échantillonnage aléatoire simple

Puisque la méthode donne à chaque individu de la population une chance égale d'être choisi, elle permet d'espérer un échantillon "représentatif". La méthode peut cependant poser certaines difficultés. D'abord, elle n'est applicable que s'il existe une liste des individus composant la population. Ensuite, elle peut être fastidieuse si l'échantillon et la population sont tous deux de grande taille. Cela sera particulièrement vrai si la sélection se fait manuellement.

2- Les échantillons systématiques

L'échantillonnage systématique est une méthode demandant moins de manipulations que l'échantillonnage aléatoire simple. Cependant, elle suppose 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. L'entier voisin de N/n sera noté r et appelé la raison de sondage (ou le pas de sondage).

Choisissons ensuite au hasard un entier d entre 1 et N: cet entier sera le point de départ. Pour former l'échantillon, il s'agira de choisir un premier individu comme point de départ; ce sera

l'individu dont le numéro correspond à d .

Pour sélectionner les autres, il suffit d'ajouter à d la raison de sondage:

les individus choisis seront alors les individus dont les numéros correspondent à:

$d ; d + r; d + 2r; d + 3r; \dots ; d + (n-1)r$.

Il faudra reprendre au début lorsque la liste sera épuisée.

Exemple

On veut choisir par échantillonnage systématique 6 étudiants parmi un groupe de 60. La raison de sondage (r) sera 10, car $N/n = 10$. (Si L'on devait trouver un échantillon de taille $n = 7$, la raison serait plutôt 9 puisque $60/7 = 8,57$, arrondi à 9.) Si L'origine choisie au hasard est disons, 3, les étudiants inclus dans L'échantillon correspondront aux numéros suivants:

3, c'est-à-dire le point de départ ; 13, c'est-à-dire $3 + 10$;

23, c'est-à-dire $3 + (2 \times 10)$ 33, c'est-à-dire $3 + (3 \times 10)$ 43, c'est-à-dire $3 + (4 \times 10)$ 53, c'est-à-dire $3 + (5 \times 10)$

Notre échantillon est maintenant complet ; il sera composé des étudiants portant les numéros suivants: 3, 13, 23, 33, 43, 53.

Selon la raison de sondage et le point de départ choisi, il peut arriver que le nombre.. obtenu se situe à l'extérieur de la liste. En pareil cas, il faut revenir au début de la liste. Si L'on reprenait l'exemple précédent et si L'on devait continuer à prélever des étudiants, on obtiendrait le numéro 63, c'est-à-dire $3 + (6 \times 10)$. Comme ce numéro r se situe à l'extérieur de la liste, on reviendrait au début de la liste et on choisirait le numéro 3 (puisque $63 - 60 = 3$).

Les avantages et les inconvénients de l'échantillonnage systématique

L'échantillonnage systématique offre des avantages. D'abord, L'échantillon est facile à sélectionner car un seul nombre est choisi au hasard. Ensuite, cette méthode tend à répartir L'échantillon sur L'ensemble de la liste de la population, ce qui permet d'obtenir une assez bonne précision. Elle offre cependant des inconvénients.

D'abord, on ne peut obtenir qu'un nombre limité d'échantillons différents à partir de la même liste. Dans l'exemple précédent, on constatera par exemple qu'il ne peut y avoir que 10 échantillons différents de taille 6. Un autre inconvénient vient de ce que les données peuvent être biaisées à cause de la périodicité de la méthode. Si L'on voulait par exemple étudier les déplacements par autobus sur 365 jours en prenant un échantillon de taille 52, la raison serait $365/52 = 7$. Si le jour de départ choisi au hasard est un vendredi, tous les autres jours seront aussi des vendredis...

En résumé, L'échantillonnage systématique consiste à choisir n unités statistiques

parmi N unités d'une population dont on possède la liste.

1. On numérote tous les individus de la liste.
2. On calcule la raison de sondage (r), c'est-à-dire l'entier le plus proche de N/n .
3. On choisit au hasard un entier d entre 1 et N .
4. A partir de d , on ajoute autant de fois r que cela est nécessaire pour sélectionner n

unités statistiques.

3- Les échantillons stratifiés

Contrairement aux deux méthodes précédentes, dans L'échantillonnage stratifié, on tient compte des renseignements qu'on pourrait déjà posséder sur la population, renseignements obtenus en particulier lors d'un recensement. La méthode repose en effet sur une division de la population en groupes relativement homogènes, appelés strates, puis sur la sélection d'un échantillon dans chaque strate. C'est une méthode qui permettra d'obtenir un échantillon représentatif, c'est-à-dire un échantillon qui possédera les mêmes caractéristiques que la population dont il a été extrait. Illustrons le processus à L'aide d'un exemple

Exemple

On veut choisir par échantillonnage stratifié 10 élèves dans un groupe de 60, en tenant compte du fait que 50 % d'entre eux sont en première année, 30 % en deuxième année et 20 % en troisième année. Chaque année sera une strate dans laquelle on ira chercher des élèves en tenant compte des pourcentages qu'occupe chaque strate dans la population. Ainsi, on choisira au hasard:

5 élèves en 1ère année, puisque $10 \times 50\% = 5$;

3 élèves en 2ème année, puisque $10 \times 30\% = 3$;

2 élèves en 3ème année, puisque $10 \times 20\% = 2$.

Il ne reste plus qu'à sélectionner un échantillon dans chaque strate, ce qui pourrait être fait par échantillonnage aléatoire simple ou systématique.

Les avantages et les inconvénients de l'échantillonnage stratifié

Cette méthode présente certains éléments positifs. L'échantillon a bien des chances d'être représentatif puisqu'on s'assure la présence proportionnelle des divers sous-groupes composant la population. Si, en plus, il existe une certaine homogénéité au sein de chacun de ces sous-groupes, L'échantillon sera encore meilleur. Il faut savoir que cette méthode, L'échantillonnage stratifié, est particulièrement souhaitable dans les cas où les strates correspondent à des divisions géographiques: il est alors possible de décentraliser les activités reliées au sondage et, par conséquent, de réduire les coûts.

Les défauts maintenant? Tout comme L'échantillonnage simple ou systématique, L'échantillonnage par strates suppose L'existence d'une liste de la population au niveau de chaque strate. Autre difficulté, c'est que, pour utiliser cette méthode, il faut savoir exactement comment la population se répartit dans chacune des strates choisies.

En résumé, L'échantillonnage stratifié consiste à choisir n unités statistiques parmi N unités d'une population dont on possède la liste.

1. On subdivise la population en strates.
2. On calcule combien il faut d'individus au sein de l'échantillon, pour représenter proportionnellement chaque strate de la population.
3. Dans chacune des strates, on choisit au hasard les individus

4- Les échantillons par grappe

Dans chacune des méthodes précédentes, les unités de L'échantillon étaient choisies individuellement. L'échantillonnage par grappes consiste plutôt à choisir plusieurs individus en même temps, c'est-à-dire par groupes. Par exemple, prenons comme population les habitants d'une ville à partir desquels on désire constituer un échantillon de 600 individus. Selon les méthodes précédentes, il faudrait choisir 600 individus disséminés dans toute la ville. Suivant L'échantillonnage par grappes, on pourra choisir les 600 individus dans une vingtaine d'immeubles choisis au hasard, où tous les occupants auront été retenus.

Exemple

Au moyen de L'échantillonnage par grappes, il s'agit de choisir 12 étudiants; dans un groupe de 60. On demande aux étudiants de se regrouper par 6. On choisit ensuite au hasard deux regroupements, par exemple les grappes numéro 4 et 7 en retenant tous les individus de ces deux grappes, on constitue un ; échantillon de 12 étudiants.

Dans l'exemple précédent, la situation était assez simple. En réalité, les choses sont beaucoup plus compliquées. Par exemple, dans une ville, les quartiers et les immeubles ne sont pas composés d'un nombre égal d'individus. Comment peut-on procéder alors pour effectuer un échantillonnage par grappes? Puisque le nombre d'habitants et le nombre d'immeubles et de logements de chacun des quartiers sont généralement connus, il est possible de quadriller la ville en un grand nombre de secteurs (qu'on appelle îlots) ayant des populations à peu près semblables. Pour constituer un échantillon, il suffira alors de choisir certains de ces secteurs et, par conséquent, toutes les personnes qui y habitent.

Les avantages et les inconvénients de l'échantillonnage par grappes

Certains avantages de L'échantillonnage par grappes sont évidents. La méthode ne nécessite pas une liste globale de la population, puisque seuls les individus inclus dans les grappes comptent. La méthode permet aussi de limiter l'échantillon à des groupes proches géographiquement, ce qui permet de réduire les frais de déplacement et de supervision.

En revanche, L'échantillonnage par grappes peut entraîner des résultats imprécis. Les éléments voisins ont en effet tendance à se ressembler. Ainsi, les caractéristiques socio-économiques des habitants d'un quartier présentent en général plus de similitudes que celles de toute la population d'une ville.

En résumé, L'échantillonnage par grappes consiste à choisir n unités statistiques parmi N unités d'une population.

1. On subdivise la population en grappes, autant que possible nombreuses et de taille équivalente. Par exemple, si L'on doit répartir 307 individus dans 10 grappes, on pourra faire 7 grappes de 31 personnes et 3 grappes de 30 personnes.

2. On calcule combien il faut de grappes pour constituer L'échantillon.

3. On choisit au hasard les grappes qui serviront à construire L'échantillon.

4. On sélectionne tous les individus des grappes choisies.

Section II- Les méthodes d'échantillonnage empiriques

Elles regroupent les techniques où les éléments de l'échantillon sont choisis sur le terrain en fonction de jugements sur les caractéristiques de la population ou bien en reconstituant le hasard. Elles ne nécessitent pas de base de sondage, elles sont rapides, et pas très chères. Les échantillons obtenus à l'aide de ces méthodes sont les suivants :

→ les échantillons par quotas ou proportionnel

→ les échantillons « accidentels » ou à l'aveuglette

→ les échantillons de volontaires

Nous reviendrons un peu plus longuement dans quelques instants sur la technique d'échantillonnage par quotas qui est la plus utilisée dans cette catégorie.

1- L'échantillonnage par quotas ou proportionnel

C'est l'une des techniques les plus simples et les plus répandues. Elle repose sur un principe simple : reproduire le plus fidèlement possible les caractéristiques de la population étudiée grâce à l'application de règles très précises.

Par ce procédé, on va chercher à construire un échantillon qui ressemble à la population mère à partir d'informations statistiques (données de cadrage disponibles) sur la

structure de cette population. On réduit ainsi le risque d'avoir un échantillon aberrant. Par ailleurs, cette méthode permet d'obtenir des précisions d'autant plus fine que les critères de quotas sont corrélés à l'objet du sondage (modèle de population)

Ces **données** statistiques, sur les principaux caractères de la population visée par l'étude, **sont appelées variables contrôlées** (sexe, âge, taille de la commune de résidence, la profession du chef de ménage, la profession du répondant, le niveau d'étude etc.).

Le principe sur lequel repose cette méthode est qu'un échantillon identique à la population mère quant à la répartition des variables contrôlées sera peu différent de cette population quant aux autres variables non contrôlées.

Généralement se sont les statistiques produites par l'INS qui sont utilisées pour établir ces quotas. Une fois sur le terrain, l'enquêteur est libre d'interroger qui il veut à condition de respecter ces quotas.

Exemple : s'il y a dans la population 48% d'hommes et 52% de femmes, 25% de jeunes de 25 à 30 ans, etc. L'enquêteur chargé d'enquêter auprès de cent personnes devra interroger 48 hommes, 52 femmes et 25 personnes de 25 à 30 ans, etc.

Les **quotas** retenus **peuvent être « simples »** c'est à dire avec plusieurs séries de distribution qui seront respectées de façon indépendante **ou encore croisée** en combinant plusieurs caractéristiques.

Revenons à notre exemple si l'on sait que 19% des femmes de la population mère a entre 25 à 30 ans et 31% des hommes ont de 25 à 30 ans. L'enquêteur devra respecter ces quotas et interroger 10 femmes et 15 hommes ayant de 25 à 30 ans.

Pour assurer une certaine objectivité aux résultats il faut fixer quelques règles aux enquêteurs comme la gestion rigoureuse des feuilles de quotas, le contrôle strict de la démarche « aléatoire » de l'enquêteur (interdiction d'interroger des personnes qui se connaissent entre elles ou qui connaissent l'enquêteur; l'obligation d'une certaine dispersion géographique avec imposition d'un itinéraire précis à respecter ; ou encore la réalisation des enquêtes à des moments différents.

2- les échantillons « accidentels » ou à l'aveuglette

La personne qui déguste des échantillons de nouveaux produits pour déterminer lequel est le meilleur ou le journaliste qui, au moyen d'entrevues dans la rue, sonde l'opinion du grand public sur un sujet d'actualité, mettent en pratique la méthode de l'échantillonnage à l'aveuglette. Le grand avantage de cette méthode est qu'elle demande peu de planification. D'application restreinte, elle peut malgré tout donner de bons résultats si la population observée est homogène. Par exemple, si l'on désire évaluer la concentration d'un produit

chimique dans un lac ou le taux de glycémie d'une personne, il est raisonnable de supposer que les résultats devraient être assez semblables d'un échantillon à l'autre.

3- Les échantillons de volontaires

Dans le cas d'expériences psychologiques ou médicales, d'enquêtes sur les habitudes de consommation, il ne serait pas pratique de choisir au hasard des individus dans toute la population. Comme l'enquête sera longue, exigeante, quelquefois même désagréable, on préfère réunir des volontaires, d'où le nom d'échantillonnage de volontaires. Néanmoins, il faut toujours craindre un écart entre les caractéristiques des volontaires et celles de la population.

Section III- COMPARAISON ENTRE LES DIFFÉRENTES

MÉTHODES

Le deuxième type de méthode d'échantillonnage est privilégiée par les Instituts de sondage et de marketing tandis que la première est utilisée par l'INS et les statisticiens professionnels.

Chacune de ces méthodes a ses avantages et ses inconvénients mais les études comparatives entre sondages probabilistes et sondages par quota ont montré l'équivalence des performances des deux méthodes, notamment dans le cadre des sondages électoraux.

Il faut toute fois préciser que l'utilisation de la méthode des quotas (empiriques) pour un échantillon de 3 000 individus et plus donnera des estimations biaisées. Tandis que l'utilisation de la méthode aléatoire donnera des résultats d'une grande imprécision pour des échantillons de moins de 80 individus.

Un autre aspect de la complémentarité de ces deux méthodes c'est la combinaison de ces deux méthodes pour la construction d'échantillons complexes.

En conclusion, sur ces techniques d'échantillonnage, il faut retenir que l'utilisation de ces techniques va vous permettre de reproduire fidèlement la population totale que vous étudiez et de ce fait vous garantir la représentativité de votre échantillon. Mais il faut savoir qu'en matière de sondage il n'existe pas de certitude même si l'échantillon a été obtenu par utilisation d'une méthode rigoureuse. Les pourcentages obtenus, résultats du sondage, ont donc toujours une marge d'erreur liée au fait qu'on interroge une fraction de la population.