

Méthodes de sondage – SOL3017 et SOL 6448

Notes de cours

Pourquoi un échantillon ? Les types d'échantillon

Département de sociologie
Université de Montréal

Professeur : Claire Durand

© Claire Durand 2009

1. Qu'est-ce qu'un échantillon, pourquoi un échantillon?

Un échantillon est constitué dès que l'on sélectionne un nombre restreint d'unités à partir d'une population d'unités. **Cette population doit être définie de telle manière que l'on peut toujours savoir si une unité fait partie de la population.**

Pourquoi échantillonner?

- réduire les coûts;
- assurer un meilleur contrôle des opérations et une surveillance plus rigoureuse;
- accélérer la cueillette et le traitement des données;
- à long terme, réduire le fardeau des informateurs.

Note:

"Les chercheurs universitaires effectuent souvent des sondages auprès des étudiants qui suivent des cours magistraux en grands groupes. La facilité et le coût moindre d'une telle méthode expliquent sa popularité mais elle donne rarement des résultats d'une quelconque valeur. Ce type de méthode peut servir pour un pré-test mais ne devrait pas être utilisé pour une étude visant à décrire la population étudiante." Babbie, E. (1977) Survey research methods. p. 108

Définitions:

La **population** est donc constituée de l'ensemble des unités auxquelles les résultats de l'enquête s'appliqueront.

La **base d'échantillonnage ou base de sondage** est constituée par la liste des unités d'échantillonnage (liste matérielle ou conceptuelle), c'est-à-dire la liste des unités à partir de laquelle se fera la sélection. Cette liste doit constituer la meilleure approximation possible de la population : *Chaque membre de la population doit y apparaître une fois et une seule fois.*

Exemple: Si je prends les diplômés de baccalauréat, de maîtrise et de doctorat des cinq dernières années, certaines personnes peuvent apparaître deux fois dans la liste, soit en tant que diplômés de deux cycles différents; elles auront donc plus d'une chance d'être choisies; je devrais donc épurer la liste de toutes les doubles ou les triples entrées avant de tirer l'échantillon.

L'unité d'échantillonnage est constituée de chaque "membre" de la base d'échantillonnage. Elle renvoie au niveau auquel se fait l'échantillonnage; l'unité d'échantillonnage peut être une personne mais également un ménage), une école, un îlot, un village.

A différencier de :

- L'unité d'analyse - le niveau sur lequel porte l'analyse; on sélectionne des ménages mais l'analyse porte sur des individus

- L'unité déclarante - celle relative à l'informateur qui donne des informations sur d'autres personnes; par exemple, une mère interrogée sur ses enfants.
- L'unité de référence - celle sur laquelle on demande des informations; par exemple, les enfants pour lesquels la mère a donné des informations.

2. Les divers types d'échantillons

Après avoir constitué la base de sondage, la deuxième étape consiste à déterminer la meilleure façon d'échantillonner. Plusieurs stratégies s'offrent à nous. Il faut déterminer laquelle sera la plus efficace tout en permettant de généraliser à la population.

2.1 L'échantillon probabiliste:

Chaque unité doit avoir une probabilité connue d'être choisie.

- Cette probabilité ne peut pas être nulle mais elle n'est pas nécessairement égale pour toutes les unités.
- Cette probabilité ne doit pas dépendre de circonstances entraînant un biais (Ex. personnes répondant habituellement au téléphone)

Pourquoi utiliser un échantillon de type probabiliste?

- permet l'utilisation de méthodes d'estimation et de méthodes d'inférence et d'analyse statistique qui sont toutes basées sur la théorie des probabilités.
- permet de connaître et donc de mieux contrôler les biais.

Type d'échantillons probabilistes:

a) Il y a en gros deux "façons" de tirer un échantillon probabiliste.

- **le tirage aléatoire simple** C'est le hasard pur, l'équivalent de tirer des noms d'un chapeau ou du tirage au hasard à la loterie. Pour le réaliser, concrètement, il faut avoir une liste de la population – la base d'échantillonnage-- et "numéroter" chaque unité de la liste; ensuite, il faut soit utiliser une table de nombres aléatoires qui déterminera quelles unités sont choisies ou, si on a une liste sur fichier informatique, utiliser un procédure informatique (présente dans la plupart des programmes d'analyse).

- **le tirage aléatoire systématique** : Il s'agit ici de tirer seulement la première unité de la liste au hasard, et de prendre ensuite les autres unités à un intervalle prédéterminé (une unité à toutes les X unités). Il faut faire ou avoir une liste - au moins conceptuelle - de la population et donc savoir combien d'unités elle comprend. Il faut ensuite déterminer l'intervalle en fonction du nombre

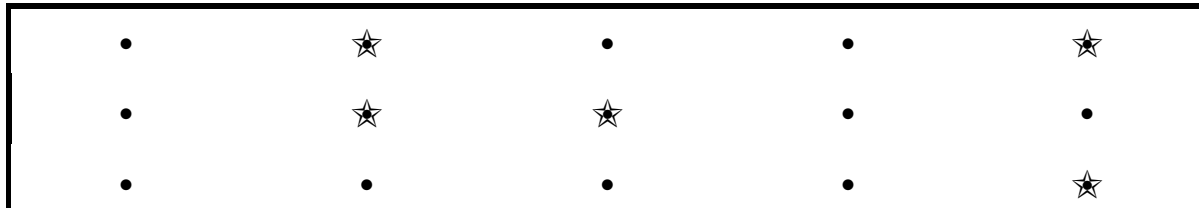
d'unités requises dans l'échantillon et déterminer au hasard la première unité. Pour procéder, on choisit la première unité et on choisit ensuite les unités en en prenant une à tous les intervalles. Ainsi, si l'intervalle est de 10 et que le nombre choisi au hasard est 7, on prendra les unités suivantes : 7, 17, 27, 37,...

Note : La liste ne doit pas avoir un ordre qui pourrait entraîner un biais : Par exemple, si j'échantillonne un logement sur trois dans un îlot où il n'y a que des triplex, je n'aurais dans l'échantillon que des personnes vivant au rez-de chaussée -- et relativement plus de propriétaires -- ou que des personnes vivant au deuxième étage et plus de locataires...

b) La probabilité de sélection peut être égale ou inégale; elle peut dépendre de divers facteurs.

Si l'échantillon est à probabilité égale, chaque unité a exactement la même probabilité d'être choisie. Par contre, dans l'échantillon à probabilité inégale, la probabilité de sélection pourrait dépendre, par exemple, de la fréquence d'occupation d'un lieu, du chiffre d'affaires de l'entreprise, du nombre d'élèves dans l'établissement. Concrètement, je peux faire un échantillon des commissions scolaires au Québec où chacune d'elles, quelque soit sa taille, a la même probabilité d'être choisie. La commission scolaire de Montréal (CSDM) aurait donc la même probabilité d'être dans l'échantillon que celle de la Basse Côte Nord. Par contre, dans un échantillon à probabilité inégale, la probabilité d'être dans l'échantillon pourrait être fonction du nombre d'étudiants; la CSDM aurait alors nettement plus de chance d'être dans l'échantillon que toute autre commission scolaire au Québec.

Échantillon aléatoire simple à probabilités égales, illustration tirée de Dussaix et Grosbras (1994):



Un échantillon peut être stratifié, proportionnel ou non-proportionnel. Dans ce cas, la probabilité de sélection est appliquée sur chaque strate et peut varier ou non selon les strates.

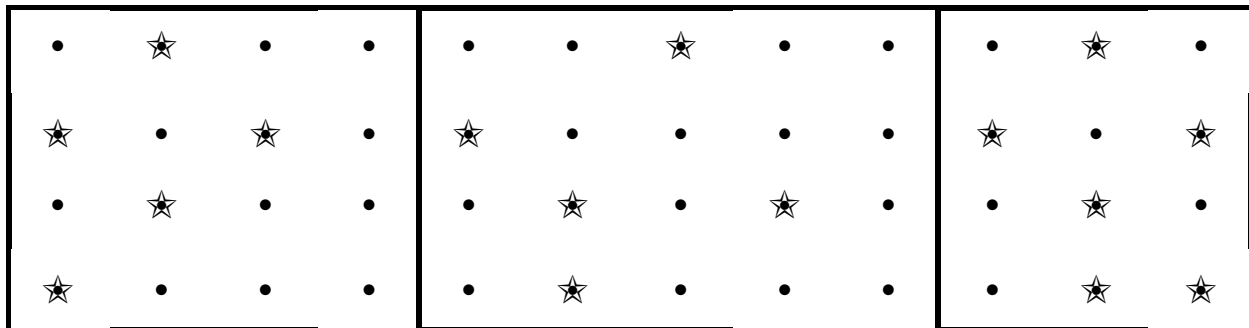
On parle **d'échantillon stratifié** lorsque la fraction de sélection peut différer en fonction de certains groupes de la population (qui peuvent être définis par le sexe, la région, le statut d'emploi dans une entreprise, etc.). On stratifie pour s'assurer que toutes les catégories de la population qui nous intéressent seront représentées en nombre suffisant. Le tirage de l'échantillon proprement dit peut être aléatoire simple ou systématique mais se fait indépendamment pour chacune des sous-populations (strates) définies.

“La stratification est un regroupement des unités de la base de sondage en sections relativement homogènes appelées strates. L’échantillon final est composé d’échantillons prélevés indépendamment dans chacune des strates.... On a recours à la stratification pour réduire l’erreur d’échantillonnage,... pour permettre un meilleur contrôle des coûts, pour avoir des estimateurs par strate,...” (Dussaix et Grosbras, 1994).

Les échantillons stratifiés peuvent être proportionnels, c’est-à-dire que l’on tire proportionnellement le nombre d’unités dans chaque strate; il s’agit alors de s’assurer d’une bonne représentation proportionnelle des strates. Ils peuvent être non-proportionnels, c’est-à-dire que certaines strates sont sur-représentées par rapport à leur proportion réelle dans la population. Il s’agit alors de s’assurer d’avoir un échantillon suffisant dans chaque strate.

Ainsi, les échantillons utilisés pour les sondages électoraux au Québec sont à peu près toujours des échantillons stratifiés non-proportionnels. Un exemple typique est la répartition en trois strates, une pour la région métropolitaine de Montréal, une pour la région métropolitaine de Québec et une pour le reste du Québec. L’échantillon est constitué de telle sorte que la strate de Montréal constitue environ 50% de l’échantillon, celle de Québec 25% et celle du reste du Québec 25%. En procédant ainsi, on “gonfle” l’échantillon de Montréal, ce qui permet d’avoir une meilleure représentation de la diversité des points de vue et des origines (plus d’anglophones et d’allophones dans l’échantillon). En plus, on réduit les coûts puisqu’il y a moins d’appels interurbains à faire (moins d’appels en province).

Échantillon aléatoire stratifié non-proportionnel, illustration tirée de Dussaix et Grosbras (1994) :

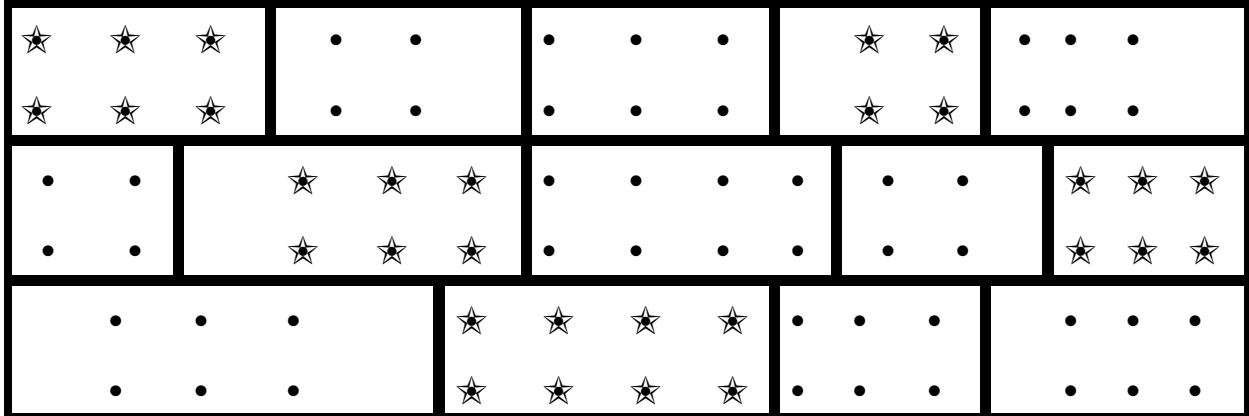


c) L’échantillon peut être tiré par groupes d’unités ou par unité

échantillon aréolaire : Il s’agit ici de sélectionner des ensembles (unités collectives du type unités de recensement, pâtés de maison, écoles, classes). Dans l’échantillonnage aréolaire, on sélectionne directement les ensembles. La base d’échantillonnage aréolaire est constituée de la liste des unités collectives à partir desquelles la sélection se fera. On parle d’échantillons aréolaires par opposition à un échantillon habituel où on sélectionne directement les individus membres de l’échantillon.

On peut sélectionner des unités ou même des groupes d'unités mais prendre également les unités ou groupes d'unités qui sont contigus : Il s'agit d'un *échantillon en grappes*. Ainsi, par exemple, un échantillon en grappes serait constitué si on sélectionnait au hasard un certain nombre d'étudiants de l'université et qu'ensuite, on constitue l'échantillon en prenant tous les étudiants qui suivent un cours en même temps que chacun des étudiants sélectionnés. Chaque groupe d'étudiants suivant les cours choisis est une "grappe".

Échantillon par grappes, illustration tirée de Dussaix et Grosbras (1994) :



d) Un échantillon peut être apparié.

- **échantillon apparié** : Il s'agit de sélectionner des échantillons de telle sorte qu'il y ait une relation deux à deux entre les membres des deux échantillons au plan des caractéristiques ou de certaines relations. Exemple : un échantillon d'étudiants, un échantillon constitué des parents des étudiants sélectionnés et un autre constitué des professeurs des mêmes étudiants.

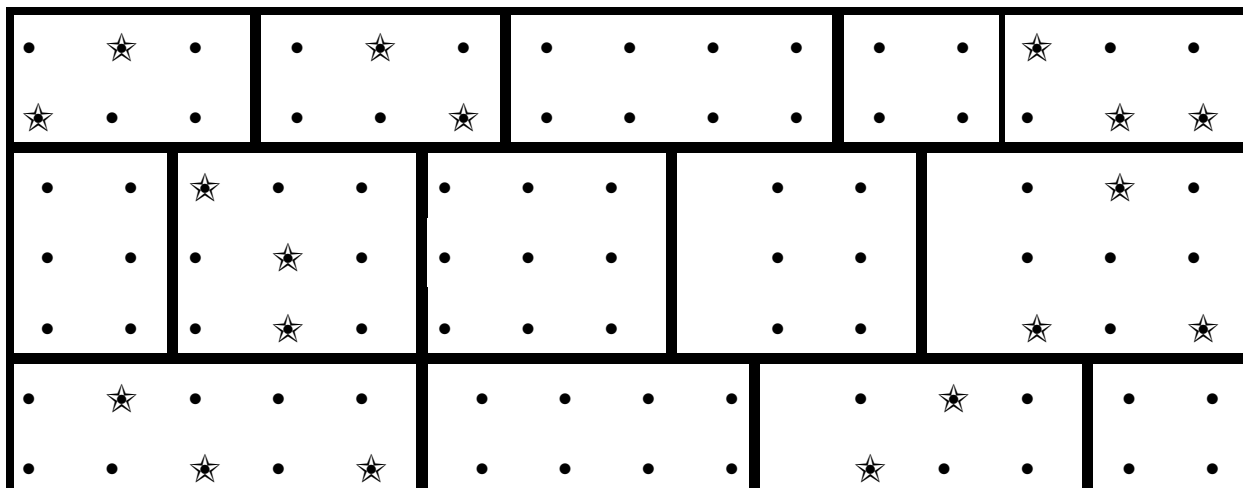
e) Les méthodes de tirage, de regroupement, de stratification peuvent être combinées en plusieurs étapes, degrés, etc.

- **Toute combinaison de ces méthodes** (plusieurs degrés, plusieurs phases)

- *plusieurs phases*: sélection d'unités, prise d'information, sélection sur la base des informations prises, nouvelle prise d'information, ...

- *plusieurs degrés*: sélection de commissions scolaires, d'écoles à l'intérieur des Commissions scolaires, d'élèves et ou d'enseignants à l'intérieur des écoles. **Le cas classique de l'échantillonnage à plusieurs degrés est celui de l'échantillonnage pour les sondages téléphoniques auprès de la population. On sélectionne un ménage (représenté par un numéro de téléphone) et ensuite, une personne à l'intérieur du ménage au moyen d'une grille de sélection au hasard.**

Échantillon à deux ou plusieurs degrés, illustration tirée de Dussaix et Grosbras (1994) :



On devra recourir à une hiérarchie de bases d'échantillonnage: Il peut y avoir des bases à plusieurs niveaux lorsque l'on a une stratégie en plusieurs phases ou à plusieurs degrés. Par exemple, il y aura au premier niveau une "liste" de numéros de téléphone à partir duquel sera tiré un échantillon renvoyant aux ménages. Au deuxième niveau, à l'intérieur des ménages choisis, il y aura une liste des membres du ménage à partir de laquelle on choisira au hasard la personne qui répondra au sondage.

2.2 L'échantillon non-probabiliste:

a) L'échantillon raisonné - au jugé : On sélectionne les "unités considérées caractéristiques" du phénomène à l'étude. Souvent utilisé en méthodologie qualitative, ce type d'échantillon pourrait être justifié pour une étude en sociologie au premier niveau de sélection. Il s'agit ici de choisir des unités (quartiers, îlots, écoles,...) en fonction de certaines caractéristiques. L'échantillonnage à l'intérieur de chacune des grandes unités se fait ensuite au hasard. **On ne peut alors généraliser qu'aux unités choisies.** L'avantage consiste à permettre de recueillir des caractéristiques objectives des grandes unités.

Par exemple, il pourrait être justifié de choisir des pâtés de maison (îlots) "types" en fonction d'un certain nombre de caractéristiques de composition sociale que l'on veut contrôler. Par la suite, on effectue une sélection au hasard des unités d'habitation dans chaque îlot.

b) L'échantillon par quota aussi appelé "échantillon représentatif": Il s'agit ici de déterminer le nombre de personnes possédant chaque caractéristique de base que l'on veut dans l'échantillon et d'arrêter de recueillir les données dès que ce nombre (le quota) est atteint.. L'échantillonnage par quota, encore pratiqué par certaines firmes de sondage, a été longtemps le principal mode d'échantillonnage. Strictement parlant, il ne permet pas l'inférence statistique et il est moins fiable en ce sens que les résultats sont plus variables d'un échantillon à l'autre (voir Vachon,

Durand et Blais, 1999). Ce type d'échantillon est pratiquement le seul utilisé par les firmes de sondage dans plusieurs pays d'Europe, des considérations techniques, pratiques et sociologiques ayant retardé l'utilisation d'échantillons probabilistes. Toutefois, de façon à pallier les problèmes des échantillons par quota, il y a eu introduction d'étapes probabilistes. Par exemple, l'échantillonnage se fera à partir de numéros de téléphone tirés au hasard ou à partir de "routes aléatoires" que les interviewers devront suivre pour interviewer les répondants.

c) L'échantillon de volontaires : "Chercheur conduisant une étude sur ... cherche volontaires pour participer à une expérience; Centre de recherche sur le syndrome pré-menstruel cherche femmes pour participer... si vous avez un ou plusieurs des symptômes suivants, prière de communiquer avec nous....". ***En aucun cas, l'échantillonnage de volontaires ne peut être considéré comme représentatif d'une population.*** On l'utilise uniquement lorsque l'on peut prétendre que les phénomènes étudiés sont intra-individuels et universels, lorsque l'on étudie des processus pour eux-mêmes.