

MODULE 3 : Comment construire un échantillon ?

2. COMMENT DÉFINIR ET COMPOSER L'ÉCHANTILLON D'UNE ENQUÊTE ?

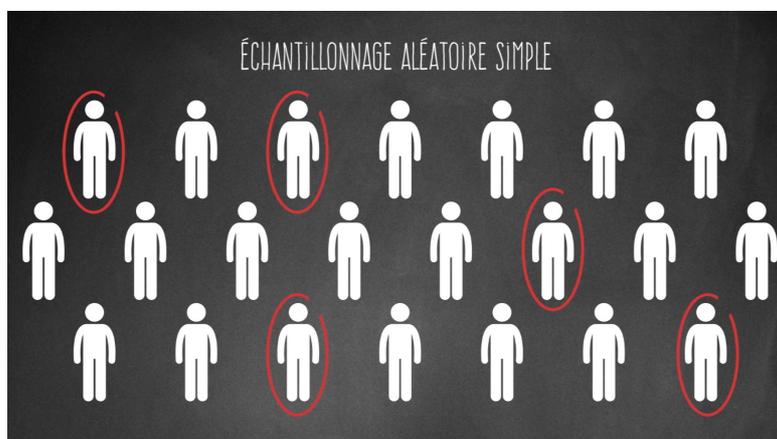
Passons maintenant à la question des échantillons. Un échantillon est une extraction de la population, qui présente des caractéristiques, définies par l'enquête, similaires à la population de référence, et à partir duquel il sera possible d'établir certaines généralisations. Autrement dit, une fois la population définie, il convient de déterminer sur quels critères devra être constitué l'échantillon censé la représenter, ou du moins sur quels critères allons-nous nous assurer de sa représentativité. Cela n'est possible qu'en référence à des hypothèses et à d'autres enquêtes préalables. Ainsi, les critères de représentativité qui président à la constitution d'échantillons pour des enquêtes électorales reposent sur une longue tradition d'enquête... et font suite à de vifs débats méthodologiques. Dans le cadre de ce cours, n'entrons pas dans ces considérations problématiques et méthodologiques, et admettons que nous avons défini nos critères. Comment allons-nous nous y prendre concrètement pour constituer notre échantillon.

Allons-nous faire jouer les probabilités en calculant nos chances d'obtenir une représentativité statistique (on parlera d'échantillon aléatoire), ou allons plutôt employer d'autres techniques, raisonnées, pour constituer ce que l'on appelle un « échantillon empirique » ou « par choix raisonné » ?

2.1 Les méthodes probabilistes d'échantillonnage

La constitution d'un « échantillon aléatoire » repose sur une démarche scientifique validant le fait de laisser intervenir le hasard. Les techniques dont nous allons parler dans un premier temps sont donc celles qui permettent de ne pas trop perturber le hasard et de rester un modèle statistique et probabiliste.

A partir des informations qui permettent de contacter les individus concernés par l'enquête (les numéros de téléphone ou les adresses mail qui composent ce que l'on appelle la « base de sondage » de l'enquête), on va choisir aléatoirement le nombre d'individus requis pour composer notre échantillon. On désigne cette première technique « échantillonnage aléatoire simple ». Ici, laisser faire le hasard consiste à ne rien décider.



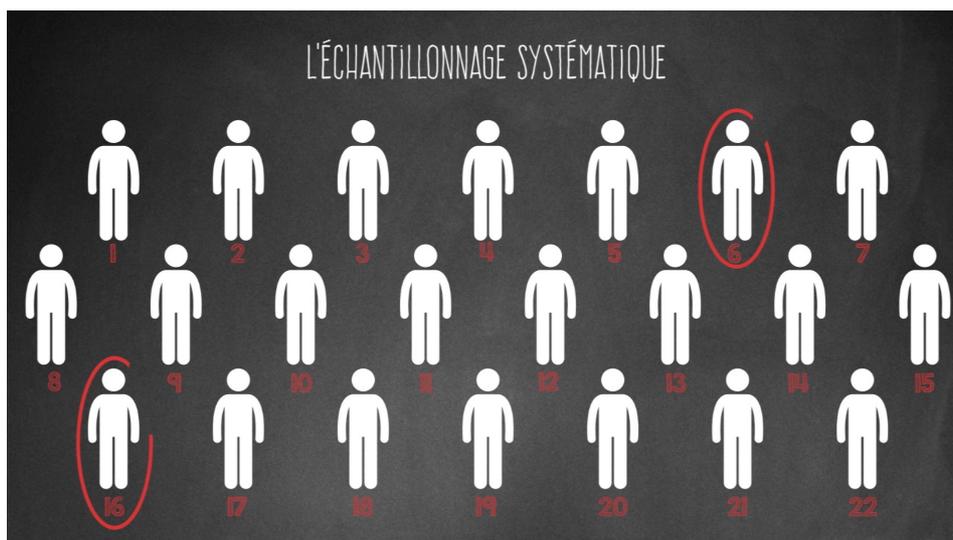
Dans le cas d'un « échantillonnage systématique », on procède au tirage dans la base de sondage de manière systématique, en établissant un « pas de tirage ». Admettons que votre base de sondage compte 5000 individus, et que vous avez déterminé qu'il vous est nécessaire d'obtenir un échantillon de 500 personnes. Pour déterminer le pas de tirage, il vous faut diviser votre base par l'échantillon souhaité > $5000 / 500 = 10$.

ÉTABLISSEMENT DU "PAS DE TIRAGE"

BASE DE SONDEGE = 5000 PERSONNES
=> ÉCHANTILLON = 500 PERSONNES

CALCUL DU PAS DE TIRAGE
= BASE DE SONDEGE / ÉCHANTILLON
= $5000 / 500$
= 10

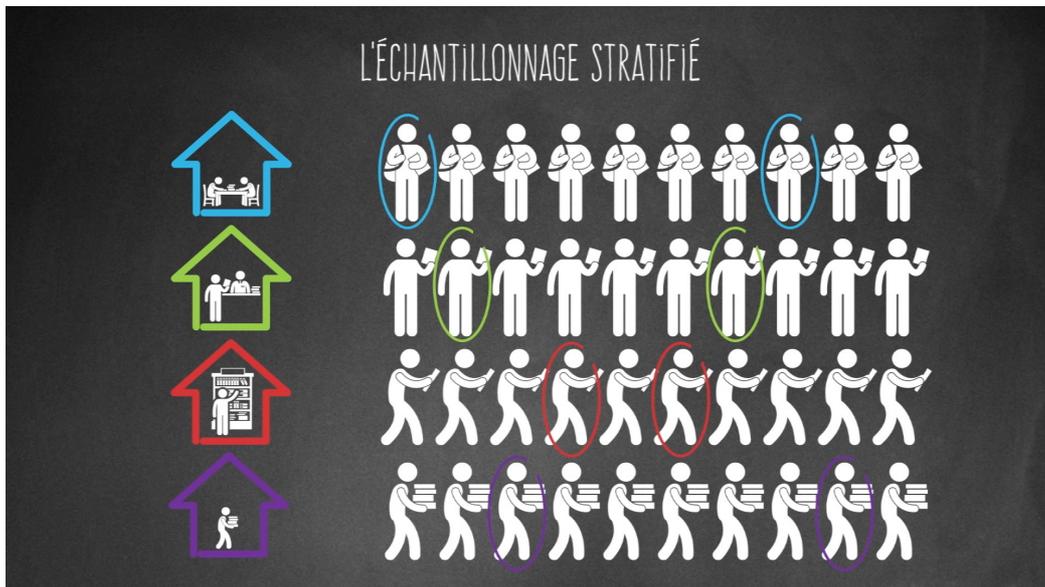
Une fois que vous aurez numéroté les individus dans votre base de sondage, vous allez en tirer 1 tous les 10, à partir d'un numéro (l'origine) choisi au hasard.



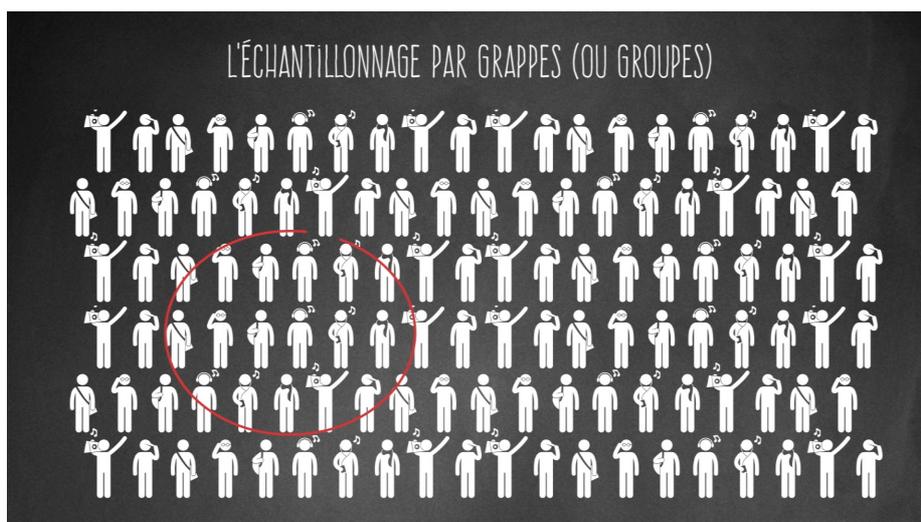
Dans certains cas, vous avez besoin de vous assurer que certaines populations distinctes sont bien présentes dans votre échantillon. Par exemple, vous conduisez une étude dans les bibliothèques d'une ville, et vous voulez vous assurer que tous les sites sont bien représentés dans votre échantillon avant de laisser parler le hasard et les probabilités.



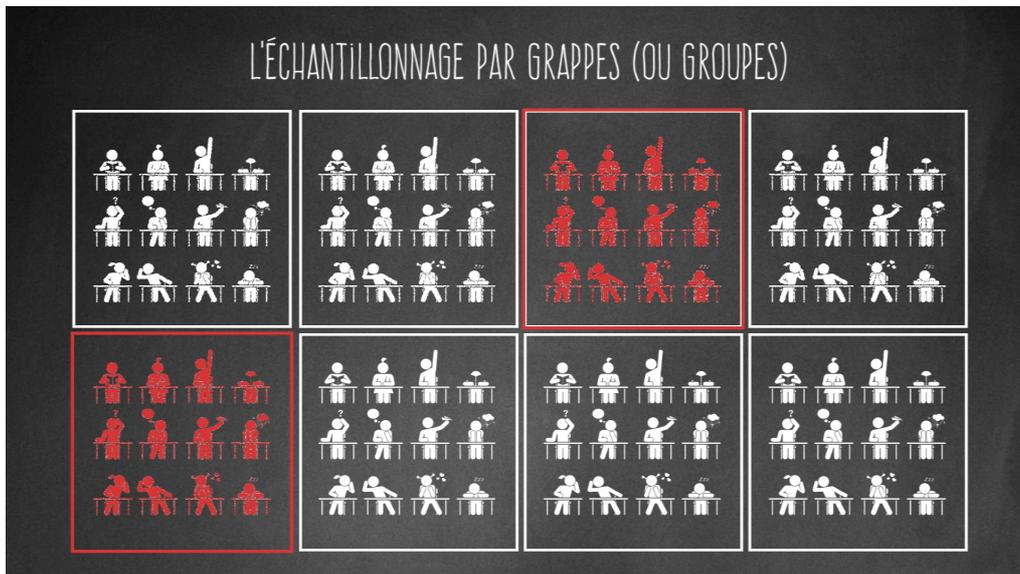
Vous allez donc découper votre population par strates homogènes (les différents sites de votre bibliothèque), et organiser l'échantillonnage aléatoire ou systématique au sein de ces strates. On appelle cette technique probabiliste « échantillonnage stratifié ».



Dernière technique probabiliste, l'échantillonnage en « grappes ». Lorsqu'il est trop complexe ou coûteux d'organiser un échantillonnage systématique ou stratifié, on peut avoir recours à un échantillon en grappes ou en groupes. Dans ce cas, on identifie des poches d'échantillons que l'on inclut intégralement dans l'échantillon. Si l'on souhaite conduire une enquête auprès d'une population de lycéens d'un département, en admettant que l'on puisse en obtenir la liste intégrale pour composer une base de sondage, il peut être fastidieux de se déplacer partout dans le département pour interroger les individus un par un.



On préférera en ce cas choisir des classes (nos grappes) et y conduire une enquête auprès de tous les élèves qui la composent. Petit problème dont il faudra tenir compte : il sera difficile d'obtenir un échantillon ayant une taille prédéterminée.



Voilà pour les principales méthodes probabilistes d'échantillonnage, celles qui permettent de composer un échantillon représentatif en faisant en sorte que les individus d'une population aient les mêmes chances d'être retenus par l'enquête.

2.2 Les méthodes non probabilistes d'échantillonnage

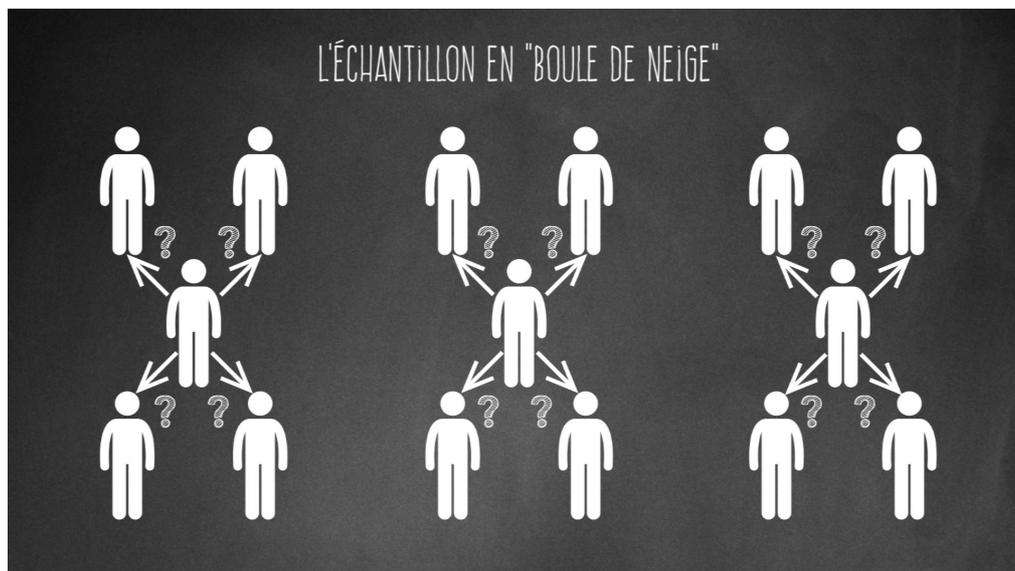
Cependant, les sciences sociales ont recours à un autre mode d'échantillonnage, empirique, qui consiste à déterminer les critères essentiels à l'enquête, et à choisir les individus sur la base de ces derniers. Plus de hasard ici, mais un choix délibéré et contrôlé par l'enquêteur et ses équipes. Il y a, ici aussi, plusieurs méthodes pour composer ces échantillons non probabilistes. La première, la plus simple à comprendre, est celle dite « de convenance » ou « de commodité ». Dans ce cadre, on sélectionne des individus parce qu'ils sont faciles d'accès. Toute la question sera alors celle de la validité de l'échantillon composée, dont on étudiera les caractéristiques au regard de ce que l'on sait de la population de référence. Ainsi, si l'on décide d'interroger les 100 premiers clients d'un cinéma parce que c'est plus pratique qu'en organisant un échantillonnage systématique, il restera à identifier les éventuels biais de ce mode de sélection.

Dans le même ordre d'idées, on peut décider d'interroger les individus qui sont volontaires pour répondre à une enquête, soit parce que cette dernière impose un engagement soutenu, soit parce que des problèmes éthiques et moraux rendent trop complexes un recours au hasard. On comprend aisément les biais de cette technique d'échantillonnage, qui peut conduire à ce que seuls les individus les plus mobilisés par une problématique, ou les plus soumis, ou bien encore les plus vénaux, participent à une enquête.



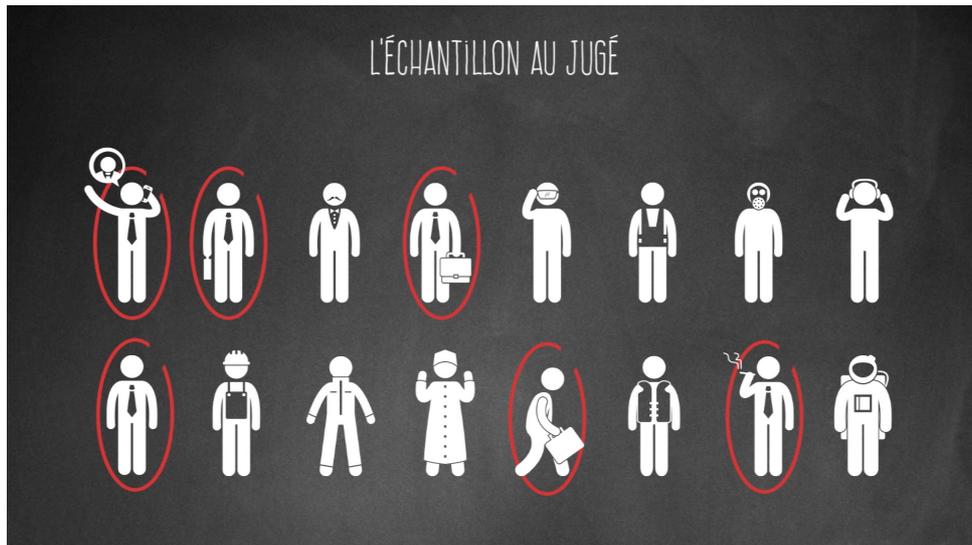
Elle reste cependant intéressante précisément lorsque l'on formule des hypothèses allant dans le sens de ces mobilisations (par exemple des études d'usages portant sur les « lead users », les utilisateurs pionniers qui seront potentiellement les plus intéressés par une enquête d'usages).

Il est également possible de faire appel à la bonne volonté des individus interrogés en composant un échantillon en « boule de neige », ce qui revient à demander à des individus interrogés de désigner dans leur entourage d'autres personnes susceptibles d'être interrogées. Ici encore, les biais sont possibles si l'on ne s'interroge pas sur les motifs ou sur les représentations conduisant un enquêté à en désigner un autre. Cette technique ne va pas non plus dans le sens d'une diversification des profils sociaux de l'échantillon.



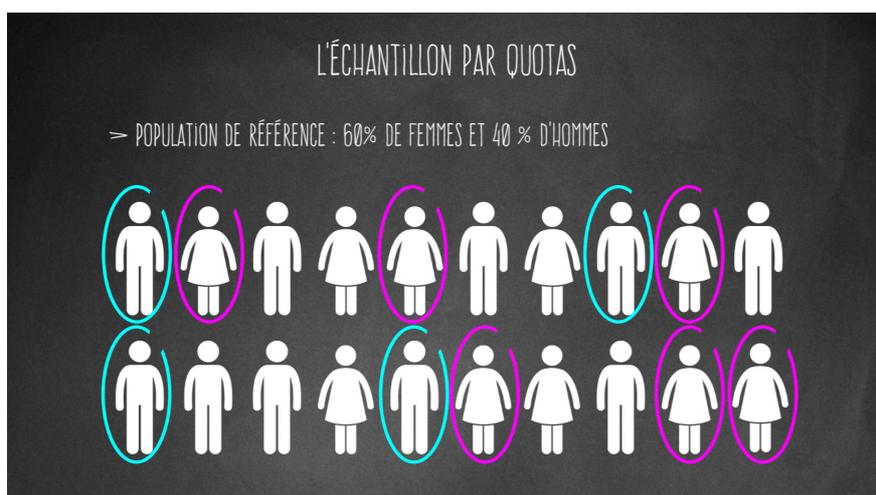
On peut également composer un échantillon « au jugé », selon les caractéristiques ou les compétences que l'on confère aux individus retenus. L'initiative n'est plus dans le camp des enquêtés, mais dans celui de l'enquêteur. Dans ce cadre, les idées préconçues du chercheur ou de l'équipe d'enquêteur interviennent fortement. Une telle technique d'échantillonnage

est intéressante pour mettre ces dernières à l'épreuve, ou dans le cadre d'études préalables, destinées à tester quelques hypothèses avant de lancer la véritable enquête.



La dernière technique, la plus répandue en matière d'échantillons non probabilistes, est celle de l'échantillon par quotas. Elle consiste à composer l'échantillon en fonction de critères retenus au préalable. On peut par exemple décider que, la population de référence étant composée à 60% de femmes et 40% d'hommes, il convient de retrouver cette proportion. Cela ressemble à la technique probabiliste par strates, sauf que dans ce cas on ne laisse plus intervenir le hasard mais que l'on choisit délibérément les individus en fonction de ce critère.

Bien entendu, on peut ajouter plusieurs critères pour composer un échantillon. 60% de femmes et 40% d'hommes, mais aussi 20% de moins de 15 ans, etc.



Enfin, il est possible de croiser ces critères, en s'assurant que l'on a non seulement la même proportion d'hommes et de femmes que dans la population de référence, mais encore la même proportion de femmes de moins de 15 ans, d'hommes de moins de 15 ans, etc. On parlera en ce cas d' « échantillonnage par quotas croisés ».

On voit bien que, composer un échantillon, revient à s'interroger sur la validité statistique des opérations conduites, mais également au sens que cette opération prend au regard de la problématique et des hypothèses de départ. Enfin, les contraintes matérielles, de durée et de budget interviennent nécessairement. Pour reprendre notre illustration du début, disons que s'il est toujours possible de vider entièrement l'étang pour y recenser tous les poissons, on préférera le plus souvent mettre en place des techniques moins précises, mais également plus rapides, plus simples et moins onéreuses...