

Échantillon Échantillonnage

Échantillonnage probabiliste

L'échantillonnage probabiliste entraîne la sélection d'un échantillon à partir d'une population, sélection qui repose sur le principe de la randomisation (la sélection au hasard ou aléatoire) ou la chance. Il est plus complexe, prend plus de temps et est habituellement plus coûteux 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 d'inclusion de chaque unité dans l'échantillon, on peut, grâce à l'échantillonnage probabiliste, produire des estimations fiables, de même que des estimations de l'erreur d'échantillonnage et faire des inférences au sujet de la population.

Il existe plusieurs méthodes différentes permettant de sélectionner un échantillon probabiliste. La méthode qu'on choisira dépendra d'un certain nombre de facteurs, comme la base de sondage dont on disposera, la façon dont la population sera distribuée, ce que sonder les membres de la population coûtera et la façon également dont les utilisateurs analyseront les données. Lorsque vous choisirez un plan d'échantillonnage probabiliste, votre but devrait consister à réduire le plus possible l'erreur d'échantillonnage des estimations pour les variables d'enquête les plus importantes, tout en réduisant le plus possible également le délai et le coût de réalisation de l'enquête.

Voici les méthodes d'échantillonnage probabiliste les plus courantes :

- [l'échantillonnage aléatoire simple](#);
- [l'échantillonnage systématique](#);
- [l'échantillonnage stratifié](#);
- [l'échantillonnage en grappes](#);
-

1. Échantillonnage aléatoire simple

Dans un *échantillonnage aléatoire simple* (EAS), chaque membre d'une population a une chance égale d'être inclus à l'intérieur de l'échantillon. Chaque combinaison de membres de la population a aussi une chance égale de composer l'échantillon. Ces deux propriétés sont ce qui définit un échantillonnage aléatoire simple. Vous devez dresser une liste de toutes les unités incluses dans la population observée pour sélectionner un échantillon aléatoire simple.

Exemple n° 1 : Il faudrait numéroter dans un ordre séquentiel chaque entrée ou inscription pour prélever un échantillon aléatoire simple d'un annuaire téléphonique. S'il y avait 10 000 entrées dans l'annuaire téléphonique et si la taille de l'échantillon était 2 000 numéros, un ordinateur devrait alors générer au hasard 2 000 numéros entre 1 et 10 000. Chaque numéro aurait la même chance qu'un autre d'être généré par l'ordinateur (ce qui respecterait l'exigence de l'échantillonnage aléatoire simple : une chance égale pour chaque unité). Les 2 000 entrées dans l'annuaire téléphonique correspondant aux 2 000 numéros aléatoires générés par l'ordinateur composeraient l'échantillon.

Un échantillonnage aléatoire simple peut s'effectuer avec ou sans remplacement. Un échantillon avec remplacement signifie qu'il est possible que l'entrée dans l'annuaire téléphonique échantillonnée soit sélectionnée deux fois ou plus. Habituellement, l'échantillonnage aléatoire simple est effectué sans remplacement, parce qu'il est plus pratique et donne des résultats plus précis. Nous ferons référence à l'échantillonnage sans remplacement lorsque nous traiterons de l'échantillonnage aléatoire simple aux fins des présentes descriptions.

L'échantillonnage aléatoire simple est la méthode d'échantillonnage la plus facile à appliquer et la plus couramment utilisée. L'avantage de cette technique tient au fait qu'elle n'exige pas de données additionnelles dans la base de sondage (comme des régions géographiques) autres que la liste complète des membres de la population observée et l'information pour les contacter. Également, puisque l'échantillonnage aléatoire simple est une méthode simple et que la théorie qui la sous-tend est bien établie, il existe des formules-types pour déterminer la taille de l'échantillon, les estimations, etc., et ces formules sont faciles à utiliser.

D'un autre côté, cette technique ne fait aucunement appel aux données auxiliaires contenues dans la base de sondage (p.ex., le nombre d'employés de chaque entreprise) qui pourraient rendre le plan d'échantillonnage plus efficace. En outre, même s'il est facile d'appliquer l'échantillonnage aléatoire simple à de petites populations, le faire peut être coûteux et irréalisable pour de grandes populations, parce qu'il faut en identifier et en étiqueter toutes les unités avant l'échantillonnage. Son application peut également être coûteuse s'il faut effectuer des interviews sur place, puisqu'il est possible que l'échantillon soit géographiquement distribué dans toute la population.

Un tirage à la loterie est un bon exemple d'échantillonnage aléatoire simple. Par exemple, lorsqu'un échantillon de six numéros est généré au hasard à partir d'une population de 49 numéros, chacun de ces derniers a une chance égale d'être sélectionné et chaque combinaison de six numéros a la même chance d'être la combinaison gagnante. Même si les gens tendent à éviter une combinaison comme 1-2-3-4-5-6, cette combinaison a la même chance d'être la série gagnante de numéros que la combinaison 8-15-21-28-32-40.

Exemple n° 2 : Supposez que votre école compte 500 élèves et que vous devez mener une courte enquête sur la qualité des aliments servis à sa cafétéria. Vous déterminez qu'un échantillon de 10 élèves devrait suffire à vos fins. Pour obtenir votre échantillon, vous attribuez à chaque élève de votre école un numéro compris entre 1 et 500. Pour sélectionner cet échantillon, vous utilisez une table de numéros générés au hasard. Tout ce que vous avez à faire consiste à prendre un point de départ à l'intérieur de la table (un numéro de rangée et un numéro de colonne) et à examiner les numéros aléatoires qui y figurent. Dans ce cas, puisque les données ne dépassent pas trois chiffres, les numéros aléatoires devraient renfermer trois chiffres également. Ne tenez pas compte des numéros aléatoires supérieurs à 500, parce qu'ils ne correspondent à aucun des élèves de votre école. Rappelez-vous que votre échantillon est un échantillon sans remplacement et que, si un numéro se répète, vous devez le sauter et utiliser le numéro aléatoire suivant. Les 10 premiers numéros différents entre 001 et 500 composent votre échantillon.

Exemple n° 3 : Imaginez que vous êtes propriétaire d'un cinéma et que vous y organiserez un festival de films d'horreur le mois prochain. Pour déterminer quels films d'horreur vous y présenterez, vous voulez demander à des cinéphiles lesquels parmi les films que vous leur énumérerez ils préfèrent. Pour dresser la liste des films nécessaire à votre sondage, vous décidez

d'échantillonner 100 des 1 000 meilleurs films d'horreur de tous les temps. La population des films d'horreur se divise en proportions égales entre les films classiques (tournés en ou avant 1969) et les films modernes (produits en ou après 1970). L'une des façons d'obtenir un échantillon consisterait à écrire tous les titres des films sur des bouts de papier et à les placer dans une boîte et à tirer ensuite 100 titres et vous auriez alors votre échantillon. En utilisant cette méthode, vous auriez l'assurance que chaque film courrait une chance égale d'être sélectionné.

Vous pourriez aussi calculer la probabilité de sélection de tel ou tel film. Puisque nous connaissons la taille de l'échantillon (**n**) et la population totale (**N**), calculer la probabilité pour un film d'horreur d'être inclus dans votre échantillon deviendrait une simple question de division :

Probabilité de sélection (la même pour chaque film)

$$= (n \div N) \times 100 \%$$

$$= (100 \div 1\,000) \times 100 \%$$

$$= 10 \%$$

Cela signifie que chaque titre de film inscrit sur votre liste aurait 10 % de chances ou 1 chance sur 10 d'être sélectionné.

Vous pouvez constater que l'un des inconvénients de l'échantillonnage aléatoire simple (parce que ce n'est pas le seul, mais c'en est un important) est le fait que vous savez que la population se compose de 500 films classiques et de 500 films modernes et que vous connaissez la date de parution sur les écrans de chaque film à partir de la base de sondage, mais qu'on n'utilise aucunement cette information. L'échantillon en question pourrait renfermer 77 films classiques et 23 films modernes, qui ne seraient pas représentatifs de la population entière des films d'horreur.

Il existe des moyens de surmonter ce problème de non-représentativité (moyens dont nous traiterons brièvement dans la section [Estimation](#)). Cependant, il y a aussi des moyens de tenir compte de cette information (moyens dont nous traiterons aussi plus loin, dans la section [Echantillonnage](#)).