

## Échantillonnage en grappes

Il est parfois trop dispendieux de disséminer un échantillon dans l'ensemble de la population. Les coûts de déplacement risquent de devenir élevés lorsque les intervieweurs doivent sonder des gens d'un bout à l'autre du pays. Les statisticiens peuvent choisir la technique de l'échantillonnage en grappes pour réduire les coûts.

La technique de l'échantillonnage en grappes entraîne la division de la population en groupes ou en grappes comme son nom l'indique. Suivant cette technique, on sélectionne au hasard un certain nombre de grappes pour représenter la population totale, puis on englobe dans l'échantillon toutes les unités incluses à l'intérieur des grappes sélectionnées. On n'inclut dans l'échantillon aucune unité de grappes non sélectionnées; ces unités sont représentées par celles tirées de grappes sélectionnées. La technique en question diffère de la technique d'[échantillonnage stratifié](#), qui entraîne la sélection d'unités de chaque groupe.

Mentionnons, entre autres exemples de grappes, les usines, les établissements d'enseignement et les régions géographiques telles que les subdivisions électorales. On utilise les grappes sélectionnées pour représenter la population.

**Exemple n° 7** : Supposez que vous représentez une organisation d'athlétisme désirant déterminer quels sports pratiquent les élèves de 11<sup>e</sup> année au Canada. Il serait trop coûteux et trop long d'interroger chaque élève canadien de 11<sup>e</sup> année ou même deux ou trois élèves de chaque classe de 11<sup>e</sup> année au Canada. On sélectionne plutôt au hasard 100 écoles de tout le pays.

Ces écoles fournissent des grappes d'échantillons. On sonde ensuite chaque élève de 11<sup>e</sup> année de chacune des 100 grappes. Les élèves inclus dans ces grappes représentent, en effet, tous les élèves de 11<sup>e</sup> année au Canada.

**Exemple n° 8** : Imaginez que le conseil municipal d'une petite localité veut faire enquête sur l'utilisation par ses résidents des services de soins de santé.

Le conseil demande premièrement à Statistique Canada des cartes des subdivisions électorales qui identifient et étiquettent chaque îlot. Il dresse à partir de ces cartes une liste de tous les îlots. Cette liste servira de base de sondage.

Chaque ménage de la localité en question fait partie d'un îlot et chaque îlot représente une grappe de ménages. Le conseil sélectionne au hasard un certain nombre d'îlots. Il dresse ensuite, à l'aide de la méthode d'échantillonnage aléatoire simple, une liste de tous les ménages inclus dans les îlots sélectionnés; ces ménages composent l'échantillon de l'enquête.

Comme nous l'avons mentionné, réduire les coûts est l'une des raisons d'utiliser l'échantillonnage en grappes. Ce dernier crée des « poches » d'unités échantillonnées, au lieu de disséminer l'échantillon sur tout le territoire. Le fait que l'on ne dispose pas parfois d'une liste de toutes les unités incluses dans la population (une nécessité lorsqu'on effectue un échantillonnage, un [échantillonnage systématique](#) ou un [échantillonnage avec probabilité proportionnelle](#) à la taille), tandis qu'une liste de toutes les grappes est disponible ou facile à dresser, constitue une autre raison d'utiliser l'échantillonnage en grappes.

Dans la plupart des cas, une perte d'efficacité comparativement à ce qu'il en est lorsqu'on utilise un échantillonnage aléatoire simple représente le principal inconvénient de l'emploi de l'échantillonnage en grappes. Il est habituellement préférable de sonder un grand nombre de petites grappes, plutôt qu'un petit nombre de grandes grappes. Pourquoi? Parce que les unités avoisinantes tendent à se ressembler davantage, ce qui donne un échantillon ne représentant pas l'éventail complet d'opinions ou de situations de l'ensemble de la population. Dans les deux exemples précédents, les élèves de la même école ont tendance à pratiquer les mêmes types de sports (suivant les installations dont dispose leur établissement d'enseignement); de la même façon, les gens âgés tendent à vivre dans des quartiers bien précis et à être de grands utilisateurs des services de santé.

L'échantillonnage en grappes ne permet pas de contrôler totalement la taille finale de l'échantillon, ce qui constitue un autre inconvénient de son utilisation. Puisque toutes les écoles ne comptent pas le même nombre d'élèves de 11<sup>e</sup> année et que les îlots ne renferment pas non plus le même nombre de ménages, et que vous devez sonder chaque élève ou ménage inclus dans votre échantillon, ce dernier peut être finalement d'une taille supérieure ou inférieure à ce que vous espériez.

### **Échantillonnage à plusieurs degrés**

La méthode d'échantillonnage à plusieurs degrés ressemble à la [méthode d'échantillonnage en grappes](#), sauf qu'il faut dans son cas prélever un échantillon à l'intérieur de chaque grappe sélectionnée, plutôt que d'inclure toutes les unités dans la grappe. Ce type d'échantillonnage exige au moins deux degrés. On identifie et sélectionne au premier degré de grands groupes ou de grandes grappes. Ces grappes renferment plus d'unités de la population qu'il n'en faut pour l'échantillon final.

Pour obtenir un échantillon final, on prélève au second degré des unités de la population à partir des grappes sélectionnées (à l'aide de l'une des [méthodes d'échantillonnage probabiliste](#) possibles). Si l'on utilise plus de deux degrés, le processus de sélection d'unités de la population à l'intérieur des grappes se poursuit jusqu'à l'obtention d'un échantillon final.

**Exemple n° 9 :** Dans l'exemple n° 8 d'échantillonnage en grappes, on choisissait 100 écoles, puis on interviewait chaque élève de 11<sup>e</sup> année de ces écoles. Dans le cas de l'échantillonnage à plusieurs degrés, vous pourriez plutôt sélectionner davantage d'écoles, vous procurer une liste de tous les élèves de 11<sup>e</sup> année de ces écoles sélectionnées et choisir un échantillon au hasard (un [échantillon aléatoire simple](#), par exemple) d'élèves de chaque école. Ce serait un plan d'échantillonnage à deux degrés.

Vous pourriez aussi obtenir une liste de toutes les classes de 11<sup>e</sup> année des écoles sélectionnées, prélever un échantillon aléatoire de classes de chacune de ces écoles, vous procurer une liste de tous les élèves des classes sélectionnées et finalement choisir un échantillon au hasard d'élèves de chaque classe. Ce serait un plan d'échantillonnage à trois degrés. Le processus se complique chaque fois que nous ajoutons un degré.

Imaginez maintenant que chaque école compte en moyenne 80 élèves de 11<sup>e</sup> année. L'échantillonnage en grappes donnerait alors à votre organisation un échantillon d'environ 8 000 élèves (100 écoles x 80 élèves de 11<sup>e</sup> année). Si vous vouliez un échantillon de plus grande taille,

vous pourriez sélectionner des écoles comptant davantage d'élèves et, pour obtenir un échantillon plus petit, sélectionner des écoles comptant moins d'élèves.

L'un des moyens de contrôler la taille de l'échantillon consisterait à stratifier les écoles en grande taille, en taille moyenne et en petite taille (tailles ici signifiant le nombre d'élèves de 11<sup>e</sup> année) et à sélectionner un échantillon d'écoles de chaque strate. On appelle cette méthode la méthode d'échantillonnage en grappes stratifiées.

Si vous utilisiez un plan d'échantillonnage à trois degrés, vous pourriez sélectionner un échantillon de 400 écoles, puis sélectionner deux classes de 11<sup>e</sup> année par école (en supposant qu'il y ait deux classes de 11<sup>e</sup> année ou plus dans chaque établissement d'enseignement). Finalement, vous pourriez sélectionner 10 élèves par classe. De cette façon, vous finiriez quand même par avoir un échantillon d'environ 8 000 élèves (400 écoles x 2 classes x 10 élèves), mais l'échantillon serait davantage disséminé.

Vous pouvez constater à partir de cet exemple que l'échantillonnage à plusieurs degrés vous offre quand même l'avantage d'un échantillon plus concentré, ce qui réduit les coûts. Cependant, l'échantillon n'est pas aussi concentré que d'autres grappes et sa taille est quand même plus grande que celle d'un échantillon aléatoire simple. Vous n'avez pas non plus besoin de disposer d'une liste de tous les élèves membres de la population. Tout ce qu'il vous faut, c'est une liste des classes des 400 écoles et une liste également des élèves des 800 classes. Il est vrai que ce type d'échantillon exige plus d'information qu'en nécessite l'échantillonnage en grappes. L'échantillonnage à plusieurs degrés épargne cependant quand même beaucoup de temps et d'efforts, parce qu'il ne nécessite pas la création d'une liste de toutes les unités incluses dans une population.

### **Échantillonnage à plusieurs phases**

Un *échantillonnage à plusieurs phases* entraîne la collecte de données de base auprès d'un échantillon d'unités de grande taille et ensuite, pour un sous-échantillon de ces unités, la collecte de données plus détaillées. La forme la plus courante d'échantillonnage à plusieurs phases est l'échantillonnage à deux phases (ou l'échantillonnage double), mais il est également possible d'effectuer un échantillonnage à trois phases ou plus.

L'échantillonnage à plusieurs phases est assez différent de l'échantillonnage à plusieurs degrés, malgré les similarités entre eux sur le plan de leur appellation. Même si l'échantillonnage à plusieurs phases suppose aussi le prélèvement de deux échantillons ou plus, dans son cas, tous les échantillons sont tirés de la même base de sondage et les unités sont structurellement les mêmes à chaque phase. Comme dans le cas de l'échantillonnage à plusieurs degrés, plus l'on utilisera de phases, plus le plan d'échantillonnage et l'estimation deviendront complexes.

L'échantillonnage à plusieurs phases est utile lorsqu'il manque à l'intérieur de la base de sondage des données auxiliaires qui pourraient servir à stratifier la population ou à rejeter à la sélection une partie de la population.

**Exemple n° 10 :** Supposez qu'une organisation a besoin d'information sur des éleveurs de bétail de l'Alberta, mais que dans la base de sondage sont énumérés tous les types d'exploitations agricoles : d'élevage de bétail et de production laitière, de grains, de porcs, de volailles et de fruits

et de légumes. Pour compliquer les choses, la base de sondage ne fournit aucune donnée auxiliaire sur les exploitations agricoles qui y sont énumérées.

On pourrait mener une enquête toute simple dont la seule question serait : « Votre exploitation agricole est-elle en partie ou en totalité consacrée à l'élevage du bétail? » Si elle ne comptait qu'une seule question, cette enquête devrait entraîner un faible coût par interview (surtout si elle était faite par téléphone), ce qui, par conséquent, permettrait à l'organisation de prélever un grand échantillon. Une fois le premier échantillon prélevé, on pourrait en extraire un second, plus petit, à partir des éleveurs de bétail et poser à ces derniers des questions plus détaillées. En utilisant cette méthode, l'organisation éviterait de dépenser de l'argent à sonder des unités ne faisant pas partie du champ d'observation (c'est-à-dire les producteurs agricoles autres que les éleveurs de bétail).

On pourrait utiliser l'échantillonnage à plusieurs phases lorsqu'on disposerait d'un budget insuffisant pour recueillir des données auprès de l'échantillon entier ou lorsque recueillir des données auprès de l'échantillon entier imposerait un fardeau excessif aux répondants ou même quand les coûts de collecte pour différentes questions d'une enquête différeraient énormément.

**Exemple n° 11 :** On pose aux participants à une enquête sur la santé des questions de fond au sujet de leur régime alimentaire et de leurs habitudes au niveau de la consommation de tabac et d'alcool et de la pratique de l'exercice physique. Cette enquête oblige, en outre, les répondants à se soumettre à certains examens médicaux, comme courir sur un tapis roulant ou faire mesurer leur tension artérielle et leur taux de cholestérol.

Remplir des questionnaires ou interviewer des participants sont des procédures relativement peu dispendieuses, mais les examens médicaux exigent la supervision et l'aide d'un professionnel ou d'une professionnelle de la santé qualifié(e), de même que l'utilisation d'un laboratoire équipé, ce qui peut dans un cas comme dans l'autre être assez coûteux. La meilleure façon de mener l'enquête susmentionnée consisterait à utiliser une méthode d'échantillonnage à deux phases. À la première phase, on soumettrait aux interviews un échantillon d'une taille appropriée. On prélèverait à partir de cet échantillon un échantillon plus petit. Ce sont les membres de ce second échantillon qui passeraient des examens médicaux.