

CHAPITRE 7 LES ENQUÊTES PAR SONDAGE

TABLE DES MATIÈRES

1. [Comment bien choisir l'échantillon?](#)
2. [La loi des grands nombres](#)
3. [Le sondage et ses limites](#)
4. [Des sondages à toutes les sauces](#)
 - [Exercices supplémentaires](#)
 - [Dossier](#)

Les sondages sont très présents lors des campagnes électorales et il est alors bien vu de les critiquer : on prétend qu'ils se trompent (comme les prévisions météorologiques), qu'ils influencent les indécis (ces derniers devraient plutôt réfléchir au lieu de se laisser mener par le bout du nez) et qu'ils vident le vote de son sens (comme si la partie était jouée d'avance).

Les sondages électoraux sont généralement absents au niveau local. Les mordus de la politique sont alors obligés d'y aller de leurs propres pronostics et la plupart d'entre eux passent tout à fait à côté de la réalité. Les organisations partisans, avec leurs listes de pointage et leurs armées de bénévoles déploient un luxe d'efforts et finissent par se tromper tout aussi lourdement.

À quoi bon blesser un éléphant avec un boulet de canon si on peut le tuer avec une carabine à plomb? (Ramou)

On est alors obligé de reconnaître que l'efficacité des sondages est impressionnante : il suffit d'interroger un millier de personnes (« bien » choisies) pour avoir une idée assez précise de l'opinion de millions d'électeurs et d'électrices. De façon similaire, pendant que Scotland Yard dépouille, à grands frais, les casiers judiciaires de tous les Londoniens fichés, Hercule Poirot après avoir examiné deux ou trois indices (bien choisis aussi) peut déterminer que l'assassin a mis de la moutarde sur ses œufs au déjeuner, qu'il est membre du *Regent Club* et qu'il a oublié de fermer le robinet à gaz en quittant son cottage dans le Kent le weekend précédent.

Bien sûr, il est toujours possible que le sondage se trompe plus ou moins souvent et à un degré plus ou moins grand, mais ces possibilités d'erreurs sont faciles à évaluer. Il suffira alors de choisir un échantillon suffisamment fiable pour que les possibilités d'erreurs ne dépassent pas les marges tolérées.

Lorsqu'on désire connaître certaines caractéristiques d'une population, on se livre à une *enquête*. Dans beaucoup de situations, le sondage, qui consiste à observer une population à travers un échantillon bien choisi, permet d'effectuer cette enquête à peu de frais.

Au terme de ce chapitre, vous devriez être en mesure de répondre aux questions suivantes :

- Comment peut-on tirer des conclusions fiables au sujet d'une population à partir des données fournies par un échantillon?
- Comment doit-on choisir un bon échantillon?
- Comment l'échantillonnage par grappes et l'échantillonnage stratifié peuvent-ils nous aider à obtenir des résultats fiables à moindres frais?

1. COMMENT BIEN CHOISIR L'ÉCHANTILLON

Tout le monde connaît l'histoire du fou qui prend les allumettes dans une boîte, les allume l'une après l'autre, et les remet dans la boîte en disant chaque fois : « celle-là est bonne ». Pour notre part, nous avons un ami qui, avant de partir en camping, prenait la précaution de vérifier ses boîtes d'allumettes réputées à l'épreuve de l'humidité. Contrairement au fou, notre campeur se contentait heureusement de brûler une allumette de chaque paquet. Et cet ami continua ainsi à prélever son *échantillon*, jusqu'au jour où il eut l'idée de remplacer les allumettes par un briquet (pas fou, mais personne n'y avait pensé).

1.1. L'échantillon : une partie à l'image du tout

Pour connaître une chose, il suffit parfois d'en prélever une partie ou un morceau. Une cuillerée suffit pour déterminer si la soupe est assez salée et une goutte de lait sur le creux du poignet permet de juger si le biberon est assez chaud. Parfois, on ne peut pas se contenter d'un échantillon aussi petit. Ce n'est pas parce que le premier homme (ou la première femme) que vous avez rencontré est le roi des crétins que tous les hommes sont comme lui. Ce n'est pas non plus parce que la première personne croisée dans la rue a l'intention de voter pour le candidat Jos Bleau (ou Tartinovitch) que celui-ci a des chances de remporter les élections. Le caractère et les opinions politiques *varient* beaucoup d'une personne à l'autre, c'est pourquoi il faut recueillir un échantillon « suffisant » avant de se faire une idée générale sur l'ensemble de la population.

L'échantillon représente une partie de la population que l'on observe, dans l'espoir de mieux connaître cette population.

Comme on vient de le voir, le mot population peut être pris au sens propre (l'ensemble des électeurs d'un pays) ou au sens figuré (l'ensemble des allumettes du paquet). Faute de pouvoir observer toute une population, on doit parfois se contenter d'en observer un échantillon. Mais ce dernier n'est pas nécessairement un reflet parfait de la population. Lorsqu'on utilise un échantillon pour se faire une idée d'une population, c'est-à-dire lorsqu'on la *sonde*, on fait alors une enquête par *sondage*, ou, plus simplement, un sondage.

Plus l'échantillon est grand, plus il donne une image fidèle de la population (nous apporterons des preuves un peu plus loin). Mais cela n'est pas suffisant. Si votre quinzième fiancé (rencontré dans le même bar que les 14 autres) ne vaut pas plus cher que les précédents, cela ne veut pas dire que tous les hommes ne valent rien. On peut par contre se questionner sur le genre d'endroit que vous fréquentez. Si les 150 premières personnes interrogées prétendent vouloir voter pour le même parti et que vous avez rencontré tous ces gens à la sortie d'un film de Falardeau, votre échantillon ne vous renseigne pas beaucoup sur les intentions de vote de la population en général. L'échantillon doit être suffisamment grand, mais il doit aussi être choisi selon les règles de l'art.

Lorsque les caractéristiques que l'on veut connaître sont très variables, pourquoi ne pas observer directement l'ensemble de la population?

Bien sûr, lorsqu'on observe toute la population on n'a pas à se soucier de la représentativité de l'échantillon (la population est l'échantillon le plus fidèle!). Mais, cela mis à part, l'enquête par *sondage* possède beaucoup d'avantages par rapport à un *recensement* complet de la population. Le sondage coûte beaucoup moins cher, car un échantillon de quelques milliers d'individus est suffisant dans la plupart des cas, et ce peu importe que la population compte 100 000 ou 100 millions d'individus. Du même coup, le sondage permet d'obtenir de l'information plus rapidement. Cela est

particulièrement utile, par exemple, lorsqu'on veut connaître à l'avance les tendances d'un vote. Il serait en effet très malheureux de ne connaître les prévisions que trois mois après la publication des résultats : ce serait aussi intéressant que d'écouter les prévisions météorologiques pour la veille! Enfin, dans un sondage, on peut se permettre d'aborder des sujets plus délicats et plus compliqués : les enquêteurs, mieux formés parce que moins nombreux que lors d'un recensement, seront plus en mesure de rassurer et de guider la personne interrogée. C'est pourquoi une enquête par sondage se révèle parfois plus fiable qu'une enquête par recensement.

Parfois, le recensement de la population est tout bonnement impossible. Comme nous l'avons vu avec le fou qui teste ses allumettes. Cette impossibilité s'applique à toutes les situations où l'observation modifie, ou détruit, l'objet observé. Les tyrans, qui craignaient par-dessus tout être assassinés par leurs proches, ne touchaient aucune nourriture avant qu'elle eût été goûtée par des serviteurs. Ces goûteurs officiels se contentaient de prélever un simple *échantillon* de chaque plat : il n'était évidemment pas question qu'ils mangent tout. Si les goûteurs avaient tout *recensé*, les tyrans n'auraient rien eu à manger. Malgré toutes ces précautions, Agrippine, la mère de l'empereur Néron, réussit à se débarrasser de son impérial époux en empoisonnant seulement la moitié du plat de champignons, et il s'agissait justement de la moitié que le goûteur n'avait pas sondée. Voilà une erreur d'échantillonnage, sans doute voulue, qui a eu des conséquences tragiques.

Recenser la population présente tout de même un avantage. Le recensement fournit des informations précises et détaillées qui permettent de vérifier la qualité des échantillons choisis et de remettre périodiquement les pendules à l'heure.

1.2. Le hasard : père d'un bon échantillon

Tous les gens qui jouent aux cartes savent qu'à la longue c'est l'habileté, bien plus que la chance qui permet de gagner. Bien sûr, on peut avoir du jeu plusieurs fois de suite et même, à l'occasion, toute la soirée, mais il n'y a pas de miracle. Les bonnes et les mauvaises cartes finissent toujours par « s'équilibrer », même si dans certains cas *exceptionnels*, la chance peut se maintenir très longtemps. C'est ce qu'on appelle vulgairement la loi des grands nombres.

Un échantillon est probabiliste lorsque tous les éléments de la population sondée ont les mêmes chances d'être choisis dans l'échantillon.

Si vous ne tirez jamais l'as, vous allez rapidement vous demander si le jeu contient vraiment quatre as, si les cartes ont été mêlées convenablement, ou si quelqu'un triche. Vous savez que le hasard devrait faire que les cartes que vous aurez en main (votre échantillon) finiront par refléter à peu près la composition du paquet (la population). Le mot clé est ici le *hasard*. Pour que l'échantillon soit représentatif, il faut absolument que tous les éléments de la population sondée (toutes les cartes du paquet) aient les mêmes chances de figurer dans l'échantillon. On parle alors d'échantillon probabiliste, puisque la probabilité qu'un élément fasse partie de l'échantillon est égale pour tous les éléments de la population.

Avec cette méthode, les chances de tomber sur un échantillon non représentatif sont d'autant plus faibles que la taille de l'échantillon est grande. De plus, il est possible d'évaluer le risque que l'échantillon tombe plus ou moins loin de la réalité.

Revenons maintenant sur la notion de *représentativité* de l'échantillon. Comment vérifier, par exemple, qu'un échantillon de Québécois est représentatif de la population du Québec? Cet échantillon devrait théoriquement compter 51 % de femmes, 6 % de protestants, 16 % de personnes âgées de 65 ans et plus, 3 % de Saguenéens, 20 % de myopes, 23 % propriétaires de chiens, etc.

Cela fait beaucoup de caractéristiques pour un simple échantillon. Combien de Saguenéens protestants myopes et propriétaires d'un chien devraient figurer dans cet échantillon? Au moins 30 ou 40 pour que la loi des grands nombres puisse s'appliquer. Pour tenir compte de toutes les caractéristiques énumérées, il faudrait alors tirer un échantillon considérable, ce que l'on cherche justement à éviter. Or il ne faut pas oublier que le sondage repose sur le principe qu'un échantillon modeste permet de connaître une population à peu de frais et de façon rapide.

Pour être *représentatif*, un échantillon doit refléter toutes les facettes du sujet étudié.

Ce problème est facile à résoudre. Pour être représentatif, un échantillon doit refléter toutes les facettes du sujet étudié et non pas toutes les facettes de la population. Si on s'intéresse au partage des travaux ménagers, par exemple, il pourrait être souhaitable que l'échantillon soit représentatif des variables suivantes : le sexe, l'âge, le niveau de scolarité. Il serait par contre inutile que l'échantillon soit représentatif de la proportion de myopes ou de propriétaires de chiens dans la population : il est peu probable que le partage des tâches ménagères soit influencé par de telles variables.

Dans certains cas, il n'est même pas nécessaire que la proportion de chaque caractéristique soit la même dans l'échantillon que dans la population. Si on s'intéresse au tabagisme, par exemple, on peut très bien constituer un échantillon formé à parts égales de fumeurs et de non-fumeurs, même si dans la population les fumeurs sont cinq fois moins nombreux que les non-fumeurs (Source : [Statistique Canada, Tabagisme 2012](#)). Dans ce cas, les fumeurs sont surreprésentés; cette surreprésentation se justifie parce qu'ils nous intéressent particulièrement. Il suffira d'en tenir compte lorsqu'on examinera les résultats du sondage.



Un coup de dé

Le mot hasard vient de l'arabe *az-zahar* qui signifie à la fois fleur et chance (ou coup de dé). En latin, le mot *alea* signifie aussi coup de dé. Comme disait Jules César en franchissant le Rubicon avec son armée : *alea jacta est!* (les dés sont jetés). Rien d'étonnant à ce qu'un échantillon choisi au hasard soit aussi appelé un échantillon aléatoire ou probabiliste. Choisir au hasard ne veut surtout pas dire choisir le premier venu. Si le hasard est aveugle, c'est justement parce qu'il obéit à des lois et qu'il peut être mis à l'épreuve (du latin *probare*, prouver, qui a aussi donné le mot *probabilis*, probable). La chance (du latin *cadere*, tomber, qui a aussi donné le mot cadence en français et *caer* en espagnol) est ce qui tombe du ciel... ou du cornet à dés. Le dé (du latin *dare*, donner) serait lui-même à l'origine un jeton ou un pion que l'on distribue au moment où se fait la « donne » du jeu. Enfin, le mot échantillon viendrait du « latin de cuisine » *scandilia*, qui signifiait mesurer, sonder, vérifier la mesure*.

Sources : Daniel Reig, Dictionnaire Arabe Français, Larousse, 1993; Walther von Wartburg, Dictionnaire étymologique de la langue française, PUF, 1968; Henri Goelzer, Dictionnaire Latin-Français, Garnier Flammarion, 1966.

1.3. Un exemple d'échantillon bien choisi

La population canadienne est recensée tous les cinq ans. Toutefois, compte tenu de l'ampleur de la consultation, les questions posées aux millions de résidents canadiens sont relativement simples et peu nombreuses. Pour mieux connaître les habitudes de vie de la population, Statistique Canada effectue régulièrement des enquêtes par sondage.

En 1994, cet organisme publiait les résultats d'une enquête portant sur la famille et les amis. Cette enquête s'inscrivait dans le cadre plus général de l'*enquête sociale générale* qui porte sur des sujets aussi variés que la santé ou les loisirs. Pour les chercheurs que cela intéresse, les données de ces enquêtes sont facilement accessibles sous forme de rapports écrits et sous forme de données informatisées (voir le site de [Statistique Canada](http://www150.statcan.gc.ca)).

Les sujets abordés par l'enquête sur la famille et les amis étaient les suivants : la vie de couple, la procréation, le partage des tâches ménagères, l'aide à la famille et aux amis, les contacts familiaux et la situation plus particulière des personnes âgées. Le tableau 7.1 donne un exemple des résultats obtenus. Nous reviendrons sur cette enquête tout au long du chapitre. Pour le moment, nous nous intéresserons uniquement à la façon dont l'échantillon a été choisi.

Tableau 7.1a - Combien d'amis proches avez-vous?

(Réponses en %)

	Aucun	1 à 2	3 à 5	6 à 9	10 et plus	Total
Hommes	7	14	30	17	31	100
Femmes	6	18	37	16	22	100

Tableau 7.1b - Où avez-vous rencontré votre meilleur ami?

(Réponses en %)

	Hommes	Femmes	Ensemble
À l'école	31	28	30
À la maison ou dans le voisinage	21	24	23
Au travail	24	19	21
À un club ou une organisation	8	7	8
Par l'entremise d'un membre de la famille	5	8	6
Par l'entremise d'un ami	5	7	6
Ailleurs	6	7	6
Total	100	100	100

Source : Statistique Canada, Enquête sociale générale (la famille et les amis), 1994. Données de 1990.

Note : Comme les chiffres ont été arrondis, il se peut que le calcul des totaux ne donne pas exactement 100.

Pour couvrir des sujets aussi variés, et pour que l'échantillon demeure suffisamment fiable au niveau régional ou selon le groupe d'âge, par exemple, il a fallu interroger un nombre considérable de personnes : en tout, 13 495 ménages ont été sondés. À titre de comparaison, il suffit généralement d'un millier de personnes pour sonder l'état de l'électorat.

L'élaboration de l'échantillon s'est faite selon les deux principes fondamentaux énoncés plus haut : l'échantillon doit être probabiliste (chaque élément a la même chance d'être sélectionné) et représentatif des caractéristiques étudiées.

On définit la *population observée*.

On a commencé par définir la population étudiée : il s'agit des ménages canadiens. Le ménage est un groupe de personnes qui habitent ensemble et qui partagent certaines tâches et certaines décisions : une sorte de famille au sens large. À la limite, le ménage peut être constitué d'une seule personne. Par contre, les pensionnaires d'une prison ou d'un foyer pour personnes âgées ne constituent pas un ménage. D'autre part, on a décidé de ne s'intéresser directement qu'aux personnes de 15 ans et plus et, pour des raisons pratiques, on a exclu la population des Territoires.

On détermine une stratégie pour constituer un *échantillon valable*.

Grâce aux compagnies de téléphone, on a pu constituer une liste de numéros associés à chaque ménage. On a pris soin d'éliminer, par exemple, les numéros des entreprises ainsi que les numéros supplémentaires que possèdent certains ménages (pour éviter que ces derniers aient plus de chances d'être choisis). Évidemment, les ménages qui n'ont pas le téléphone (moins de 2 % des ménages) n'ont pas été consultés : cela aurait été très regrettable si l'enquête avait porté sur l'opinion des sans-abris, mais, dans ce cas-ci, l'inconvénient s'avère négligeable compte tenu des coûts économisés.

La stratification consiste à diviser la population en groupes relativement homogènes (les strates) et à sélectionner ensuite un échantillon indépendant dans chaque strate.

Une fois la liste des numéros de téléphone établie, on l'a découpée en différents paquets correspondant à chaque région (grâce aux indicatifs régionaux). On a tiré au hasard 18 325 numéros de téléphone en faisant en sorte de respecter dans l'échantillon les mêmes proportions régionales que dans la population*. Ce principe, que l'on appelle la stratification, permet par exemple d'éviter de tirer, par hasard, un échantillon qui ne contiendrait aucun Québécois. Dans le même ordre d'idée, on pourrait diviser la population en sous-groupes basés sur l'origine ethnique, l'âge et même le sexe avant de constituer un échantillon visant à sonder les intentions de vote des Québécois. Cela permettrait, moyennant un petit effort supplémentaire, de diminuer grandement le risque de tomber sur un échantillon aberrant.

Ces proportions sont connues grâce au recensement : on voit, à ce propos, que les deux méthodes servant à étudier une population (le *recensement* et le *sondage*) peuvent se compléter.

On joignait alors chaque ménage sélectionné au téléphone, on vérifiait auprès de la personne qui avait répondu à l'appel que le ménage était admissible, c'est-à-dire qu'il faisait partie de la population choisie, et on notait certaines données sur chaque membre du ménage : âge, sexe, état matrimonial et lien avec un membre du ménage choisi comme référence. On choisissait alors *au hasard* un seul répondant pour le ménage et on faisait venir cette personne au téléphone (à moins qu'elle ne soit déjà au bout du fil) pour lui poser la série de questions (environ 300 questions) sur l'enquête proprement dite. Aucune entrevue par personne interposée n'était acceptée.

Parmi les 18 325 ménages appelés, 3206 ne répondirent pas (dont 1 884 qui refusèrent carrément de parler à l'enquêteur). Sur les 15 119 ménages restants (soit 18 325 – 3206), il y avait encore 1624 non-réponses (dans 438 cas, l'individu sélectionné dans le ménage refusait de répondre, dans 786 cas la communication a été coupée avant que l'enquête ne soit terminée, dans 13 cas on s'aperçut, après coup, que le ménage ne faisait pas partie de la population étudiée et dans 387 autres cas, les

réponses étaient trop incomplètes pour être traitées). Il restait, au bout du compte, 13 495 réponses (soit 15 119 – 1624).

Les non-réponses diminuent la fiabilité d'un sondage, car elles ne sont pas nécessairement l'effet du hasard. En effet, les personnes qui refusent de répondre, par exemple, partagent peut-être des caractéristiques communes, qui exercent une influence sur les variables mesurées par l'enquête. Dans notre exemple, le taux de réponse est quand même raisonnable compte tenu de l'envergure de l'enquête : 13 495 ménages sur 18 325, soit 74 %.

$$\text{Taux de réponse} = \frac{\text{Nombre d'individus qui répondent}}{\text{Nombre d'individus dans l'échantillon}}$$

$$\text{Taux de réponse} = 13\,495/18\,325 = 0,74 = 74 \%$$

EXERCICES 1

1. La représentativité de l'échantillon

Pour chacun des cas suivants, nommez trois caractéristiques qui, selon votre connaissance du sujet, devraient être bien représentées dans l'échantillon, et nommez une caractéristique dont il serait inutile de tenir compte.

- a) Une enquête sur la conduite dangereuse sur la route.
- b) Une enquête sur les préjugés raciaux.
- c) Une enquête sur le tabagisme.
- d) Une enquête sur les élections provinciales.

2. L'enquête sur la famille et les amis

Dans l'enquête sur la famille et les amis (dont il est question dans cette section), quelles sont les précautions qui ont été prises pour s'assurer que l'échantillon soit représentatif?

2. LA LOI DES GRANDS NOMBRES

Observez le tableau 7.2 ci-dessous. On y retrouve les résultats, pour le Québec, des élections fédérales du 26 octobre 1993 à côté du dernier sondage de la maison Gallup (un [dossier complet](#) est présenté en annexe du prochain chapitre sur cette élection historique.). Si le sondage tombe si près de la réalité, ce n'est nullement par un mystérieux acte de sorcellerie. L'échantillon (un millier de personnes interrogées du 17 au 20 octobre) est suffisamment grand pour refléter, avec un bon degré de fidélité, l'opinion des quelque 3 700 000 électeurs qui s'apprêtaient à voter une semaine plus tard.

Tableau 7.2 - Un sondage électoral

		Sondage Gallup 17-20 octobre 1993	Élections 26 octobre 1993
(Réponses en %)			
Bloc Québécois	BQ	50	49,2
Libéral	PLC	31	33,7
Conservateur	PC	14	12,9
Autres		5	4,2
Total		100	100,0
Marge d'erreur 19 fois sur 20 (95%)		3,1	
Taille de l'échantillon (n)		1011	

Source : La Presse, 22 octobre 1993, Le Devoir, 27 octobre 1993.

Depuis le début de ce chapitre, nous avons affirmé à plusieurs reprises que les chances de tomber sur un échantillon non représentatif diminuent lorsque la taille de l'échantillon augmente, à condition bien sûr que l'échantillon soit vraiment choisi au hasard. Il est maintenant temps de justifier cette affirmation.

2.1. Des anthropologues sur le terrain

Un congrès sur la famille, qui se tient dans un hôtel des Caraïbes, réunit, pour le weekend, 6 hommes et 6 femmes, originaires de 6 pays différents. Nous l'avouons, cette situation est imaginaire, mais le nombre d'enfants correspond réellement à l'indice de fécondité pour chaque pays en 1990-1995.

- Une Ivoirienne et un Ivoirien, tous deux parents de 6 enfants,
- Une Jordanienne et un Jordanien, tous deux parents de 5 enfants,
- Une Salvadorienne et un Salvadorien, tous deux parents de 4 enfants,
- Une Albanaise et un Albanais, tous deux parents de 3 enfants,
- Une Chinoise et un Chinois, tous deux parents de 2 enfants,
- Une Espagnole et un Espagnol, tous deux parents d'un enfant unique.

But de la recherche : estimer le nombre moyen d'enfants par couple.

Trois anthropologues se présentent à l'hôtel et veulent connaître le nombre moyen d'enfants (la *variable* étudiée) des personnes qui participent au congrès (la *population*). Comme ils sont régis par un décret, ces chercheurs n'ont pas le droit d'interroger plus d'une personne par jour.

Le premier anthropologue, une personne très sérieuse, décide d'interroger une femme le samedi et un homme le dimanche. Il leur demandera combien d'enfants ils ont, respectivement. En faisant ensuite la moyenne des deux résultats (*l'échantillon*), le chercheur compte obtenir un chiffre assez proche de la moyenne de toutes les familles représentées au congrès.

Le deuxième anthropologue interroge une femme le samedi et décide de passer le dimanche au bord de la piscine pour compiler ses résultats. Pour se donner bonne conscience, ce chercheur se répète la devise des goûteurs : « inutile d'avaler plus d'une cuillerée quand on veut savoir si la soupe est chaude ».

Le troisième anthropologue compte se reposer deux jours avant de commencer ses travaux, d'autant plus que sa banque de congés de maladie est en train de déborder. Sa devise : « ne fais jamais demain ce que tu peux faire après-demain ». Le chercheur, une fois bien en forme, compte occuper les 12 jours qui suivent à interroger successivement chacune des personnes qui participent au congrès*

Salut les anthropologues! Tout ça, c'est des farces, mais il nous fallait une tête de Turc.

2.2. Les résultats de l'enquête

Inutile de dire que le troisième anthropologue n'obtiendra aucun résultat : qui trop embrasse, mal étreint. Son projet d'étudier toute la population (de faire un *recensement* en somme) est bien trop lourd pour être mené à bien dans le temps voulu.

Nous savons quant à nous que la moyenne du nombre d'enfants par personne est de $[(2 \times 6) + (2 \times 5) + (2 \times 4) + (2 \times 3) + (2 \times 2) + (2 \times 1)]/12 = 42/12 = 3,5$. La question est la suivante : lequel des deux premiers anthropologues a des *chances* de *tomber* le plus près de la moyenne avec son échantillon? En passant, vous rappelez-vous que le mot *chance* vient justement d'un verbe latin qui veut dire *tomber*?

Le deuxième anthropologue, qui n'a interrogé qu'une personne, a autant de chances de tomber sur un élément très éloigné de la moyenne (1 enfant ou 6 enfants) que de tomber sur un élément très proche de la moyenne.

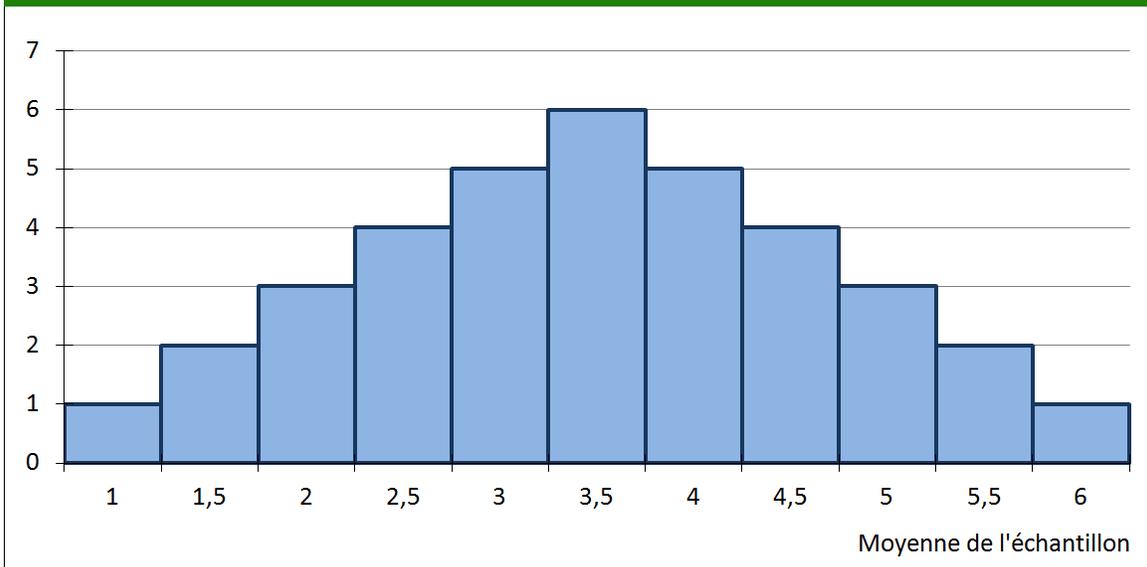
Le premier anthropologue est dans une bien meilleure position. Examinons la liste des échantillons de 2 éléments qu'il pourrait avoir recueillis (tableau 7.3). Dans cette population de 6 femmes et 6 hommes, il existe 36 échantillons possibles de 2 éléments comprenant un homme et une femme. On constate dans le tableau 7.3 qu'il y a beaucoup plus d'échantillons concentrés autour de la moyenne de la population (il y a 6 échantillons qui donnent une moyenne de 3,5) que d'échantillons éloignés de cette moyenne (il n'y a qu'un seul échantillon qui donne une moyenne de 2 ou une moyenne de 6).

Tableau 7.3 - Liste des échantillons de 2 éléments tirés de la population

Él. 1	Él. 2	Moy.	Él. 1	Él. 2	Moy.	Él. 1	Él. 2	Moy.	Résumé du tableau	
									Moyenne de l'échantillon	Nombre d'échantillons
6	6	6,0	4	6	5,0	2	6	4,0	6,0	1
6	5	5,5	4	5	4,5	2	5	3,5	5,5	2
6	4	5,0	4	4	4,0	2	4	3,0	5,0	3
6	3	4,5	4	3	3,5	2	3	2,5	4,5	4
6	2	4,0	4	2	3,0	2	2	2,0	4,0	5
6	1	3,5	4	1	2,5	2	1	1,5	3,5	6
5	6	5,5	3	6	4,5	1	6	3,5	3,0	5
5	5	5,0	3	5	4,0	1	5	3,0	2,5	4
5	4	4,5	3	4	3,5	1	4	2,5	2,0	3
5	3	4,0	3	3	3,0	1	3	2,0	1,5	2
5	2	3,5	3	2	2,5	1	2	1,5	1,0	1
5	1	3,0	3	1	2,0	1	1	1,0	3,5	36

Ce phénomène est encore plus évident lorsqu'on distribue les valeurs moyennes de chaque échantillon possible sur un graphique (figure 7.1). On voit clairement que ces valeurs sont concentrées autour de la moyenne. En fait, la distribution commence un peu à ressembler à une courbe normale (la courbe en forme de cloche vue au [chapitre 3](#)). Nous pouvons déjà affirmer que le premier anthropologue a beaucoup plus de chances, en interrogeant 2 femmes au hasard de tomber sur un échantillon plus représentatif que son collègue qui dispose d'un échantillon plus petit.

Figure 7.1 - Distribution des moyennes d'échantillon de 2 éléments tirés de la population



2.3. Une méthode rigoureuse

Grâce à la moyenne d'un échantillon, on peut obtenir une *estimation* assez précise de la moyenne de la population.

Plus l'échantillon sera grand, moins il sera probable que la moyenne de cet échantillon s'écarte sensiblement de la moyenne de la population (nous vous laissons le soin de le vérifier en exercice avec des échantillons de 3 éléments). Évidemment, la situation que nous avons présentée est volontairement simplifiée. En pratique, les échantillons doivent contenir au moins une trentaine d'éléments pour que leurs moyennes suivent la distribution de la courbe normale. On pourra alors utiliser les propriétés de la loi normale pour évaluer le degré de fiabilité de l'échantillon, en déclarant, par exemple, qu'il y a une probabilité de 95 % que les résultats fournis par l'échantillon du tableau 7.2 vu au début de cette section (les élections du Bloc québécois) ne s'écartent pas de plus de 3 % de ce qu'on aurait obtenu en observant toute la population.

Nous verrons au prochain chapitre comment mesurer précisément ces marges d'erreur lorsqu'on cherche à estimer les caractéristiques d'une population à partir d'un simple échantillon. Notre propos était seulement ici de montrer qu'avec un échantillon raisonnablement élevé, il était relativement facile de sonder, avec un degré raisonnable de certitude, une population inconnue.

Pour conclure, mentionnons une douloureuse expérience de jeunesse. Notre grand-frère, bien que peu studieux à l'école, avait saisi très vite tout l'intérêt des méthodes quantitatives. Tous les samedis, il nous conviait à jouer notre argent de poche aux dés. Avant de lancer les deux dés, qu'il avait « empruntés » au jeu de Monopoly des voisins, il nous encourageait à choisir le 12 (chiffre prestigieux), tandis qu'il se contentait du 7 (chiffre malchanceux). Nous ignorions alors que le 7 (moyenne de 3,5 par dé) sort 6 fois plus souvent que le 12 (moyenne de 6 par dé). Il suffit de revoir le tableau 7.3 (ou la figure 7.1) pour s'en convaincre.

Nous espérons une fois de plus que vous utiliserez les connaissances acquises dans ce manuel à bon escient, c'est-à-dire en luttant contre les tricheurs. Nous déclinons d'avance toute responsabilité quant aux tentatives d'extorsion de fonds inspirées par la lecture de cette section de chapitre.

EXERCICES 2

1. Le quatrième anthropologue

En vous inspirant du tableau 7.3 et de la figure 7.1, faites la liste de tous les échantillons de 3 éléments. Chaque échantillon doit comporter 3 personnes différentes, peu importe le sexe. Calculez le nombre moyen d'enfants de chaque échantillon et tracez-en la distribution sur un graphique.

2. Alea jacta est!

Pour cet exercice, vous devez utiliser deux dés (obtenus par des moyens honnêtes). Si vous n'avez qu'un seul dé, vous pouvez le lancer deux fois (pas fou, n'est-ce pas?). Pour rendre l'exercice plus intéressant, vous pouvez travailler en équipe de deux personnes.

- a) Lancez 6 fois les dés, et notez chaque fois la moyenne (de la valeur des deux dés) obtenue. Quand vous avez fini, tracez une courbe de distribution du même genre que celle de la figure 7.1.
- b) Même chose en lançant les dés 36 fois.
- c) Même chose en lançant les dés 216 fois.
- d) Comparez les trois courbes de distribution et commentez.

3. LE SONDAGE ET SES LIMITES

Le sondage, qui consiste à estimer les caractéristiques d'une population par l'observation d'un échantillon, n'est pas une méthode d'enquête parfaite (demandez au mari d'Agrippine ce qu'il en pense). Le sondage peut donner lieu à des erreurs d'estimation plus ou moins grandes. Ces erreurs sont de deux types. Il se peut tout d'abord que, malgré toutes les précautions prises, le hasard fasse que l'échantillon observé ne soit vraiment pas représentatif de la population. Ce type d'erreur est facile à évaluer, à l'aide de calculs relativement simples (nous expliquerons comment procéder à ces calculs dans le [prochain chapitre](#)). D'autres erreurs, par contre, ne peuvent être mises sur le dos du hasard : il s'agit d'erreurs de méthodes contre lesquelles les plus savantes formules mathématiques ne peuvent rien.

Avant de faire l'inventaire des principales erreurs à éviter, nous parlerons de deux moyens d'améliorer à peu de frais la fiabilité de l'échantillon : la méthode des grappes et la stratification.

3.1. L'échantillonnage par grappes : pour que le sondage porte fruit

Une grappe est un sous-groupe de la population défini selon la proximité géographique. L'échantillonnage par grappes consiste à choisir des grappes au hasard et à inclure dans l'échantillon un certain nombre de membres de chaque grappe choisie.

Lorsque l'envergure de l'enquête exige l'observation d'un grand échantillon, on a souvent recours à une astuce pour diminuer les coûts. Au lieu de tirer chaque élément de l'échantillon de façon individuelle, on fonctionne par grappes. Supposons qu'on ait décidé de choisir au hasard un échantillon de 1000 ménages. Qu'est-ce qui vous paraît le plus fiable : choisir au hasard les 1000 logements où résident ces ménages dans 100 immeubles (ou 100 pâtés de maisons) de 10 logements chacun, ou choisir au hasard ces 1000 logements à travers toute la ville? La première option présente un risque de distorsion plus grand, car les habitants d'un même immeuble ou d'un même quartier possèdent souvent des affinités qui n'ont rien à voir avec le hasard. Par contre, la seconde option, à cause du grand nombre de déplacements, est plus onéreuse. Le véritable choix, compte tenu des coûts, pourrait bien être le suivant : 1000 ménages (pris dans 100 immeubles de 10 logements chacun) ou bien 200 ménages (pris dans 200 logements éparpillés au hasard)? Lorsqu'on a l'intention de regarder à la loupe certaines caractéristiques plus ou moins rares de la population, il est essentiel de travailler sur un échantillon suffisamment grand. Dans ce cas, l'échantillonnage par grappes peut s'avérer plus fiable sans coûter plus cher.

Comment améliorer à peu de frais la fiabilité de l'échantillon.

L'échantillonnage par grappes n'est possible que lorsque la population étudiée se présente sous forme d'une hiérarchie (par exemple : pays, régions, villes, immeubles, ménages, individus). Lorsque cette condition est remplie, l'échantillonnage par grappes (qui permet dans le même temps ou pour le même coût d'interroger plus de monde) peut s'avérer plus représentatif qu'un échantillonnage purement probabiliste. C'est pourquoi Statistique Canada utilise cette méthode pour calculer les taux de chômage.

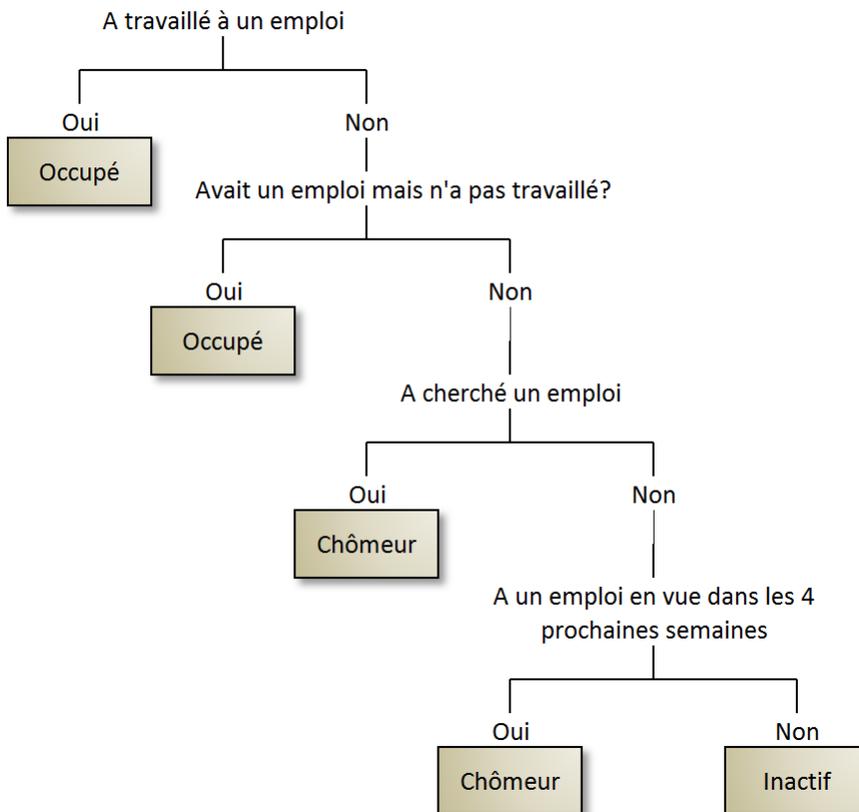
Un exemple d'échantillonnage par grappes : le sondage sur le chômage.

L'enquête sur l'emploi, qui se fait en effet par sondage et non en dépouillant la paperasse de l'Assurance-emploi, doit être à la fois peu coûteuse (l'enquête est répétée chaque mois) et très fiable (les taux de chômage régionaux sont utilisés notamment pour calculer les prestations versées). Étant donné que la personne choisie dans l'échantillon doit être interviewée en personne au moins une

fois, le coût et le temps de déplacement deviennent alors une contrainte importante. La méthode de l'échantillonnage par grappes permet de sonder rapidement un échantillon suffisamment grand (environ 62 000 ménages en tout dans les dix provinces canadiennes) pour pouvoir le découper ensuite par région.

La situation d'un individu sur le marché du travail peut être rangée dans une des trois catégories suivantes : chômeur, occupé à un emploi, inactif. Ces trois catégories sont exclusives (on ne peut faire partie de plus d'une catégorie à la fois) et exhaustives (on fait nécessairement partie d'une de ces catégories) : on reconnaît là les caractéristiques d'une échelle nominale. Le concept de chômeur est simple, mais la réalité l'est moins, c'est pourquoi les catégories sont définies de façon rigoureuse (voir la figure 7.2). Il ne servirait à rien de faire de savants calculs si la notion de chômeur variait chaque mois ou dans chaque région, ou selon l'enquêteur. Compte tenu de ces définitions, certaines personnes sont exclues, en toute logique, de la population observée (les personnes de moins de 15 ans, par exemple, ne devraient pas se trouver sur le marché du travail). D'autres personnes ont été exclues pour des raisons pratiques : les gens vivant dans les Territoires, par exemple, sont trop éparpillés pour qu'on puisse les rejoindre si souvent.

Figure 7.2 - La définition d'un chômeur



Le poids d'échantillonnage est le nombre de sujets de la population représentés par chaque sujet de l'échantillon.

Le Canada compte, au moment de l'enquête en 1994, environ 23 millions de personnes âgées de 15 ans et plus. On peut en déduire, que chacune des 62 000 personnes de l'échantillon interrogé représente approximativement 370 personnes dans la population. Ce rapport s'appelle le poids d'échantillonnage.

$$\text{Poids d'échantillonnage} = \text{Taille de la population} / \text{Taille de l'échantillon}$$

$$\text{Poids d'échantillonnage au Canada} = 23 \text{ millions} / 62\,000 \approx 370.$$

Ce poids d'échantillonnage varie d'ailleurs d'une province à l'autre. Au Québec et en Ontario, chaque personne sondée représente environ 400 individus dans la population. Dans une province comme Terre-Neuve, où la population est plus faible et plus dispersée, le poids d'échantillonnage est d'environ 200 (toutes proportions gardées, on interroge deux fois plus de monde qu'au Québec). La population de 15 ans et plus étant de 5 730 000 personnes au Québec (en 1994), on peut en déduire que l'échantillon interrogé au Québec compte environ 14 000 sujets.

$$\text{Taille de l'échantillon} = \text{Taille de la population} / \text{Poids d'échantillonnage}$$

$$\text{Taille de l'échantillon au Québec} = 5\,730\,000 / \text{environ } 400 \approx 14\,000.$$

Cet échantillon est suffisant pour que la marge d'erreur soit faible, même au niveau régional.

3.2. La stratification : une garantie supplémentaire

Mettre toutes les chances de son côté en raffinant l'échantillon.

La *stratification*, dont nous avons parlé dans la première section à propos de l'[enquête sur la famille et les amis](#), permet d'éliminer des échantillons aberrants. Sans stratification, il se pourrait, par hasard, qu'un échantillon de Canadiens ne contienne aucun Québécois. En outre, la stratification rend plus fiables les estimations par catégorie lorsque ces catégories correspondent aux strates choisies.

Comment vérifier, après coup, la représentativité de l'échantillon?

Supposons qu'après avoir dépouillé les réponses données à un sondage on constate deux choses : d'une part, 44 % des Québécois(e)s âgés de 20 à 64 ans estiment avoir un poids excessif et d'autre part, 60 % des personnes ayant répondu au sondage sont des femmes. Le fait que les femmes soient surreprésentées dans l'échantillon observé peut sûrement fausser les résultats, car il y a relativement plus de femmes que d'hommes qui croient avoir un poids excessif, même si cette perception n'est pas conforme à la réalité.

Des résultats valables avec un échantillon boiteux.

Dans un cas comme celui-ci, il est relativement facile de rectifier les résultats, car nous connaissons, grâce au recensement, la répartition des sexes dans la population. Il suffit alors de pondérer les résultats de chaque catégorie en fonction de cette répartition. Sachant, d'une part, après avoir à nouveau observé les questionnaires dépouillés que 38 % des hommes et 48 % des femmes estiment avoir un poids excessif, et, d'autre part, que les hommes et les femmes représentent chacun 50 % de la [population de 20 à 64 ans*](#), on peut en déduire que la proportion de personnes qui estiment avoir un poids excessif est en réalité de 43 %, et non de 44 %.

Pour être plus précis, il y a (en 1994), chez les Québécois(e)s âgés de 20 à 64 ans, 50,1 % d'hommes et 49,9 % de femmes.

$$\text{Résultat corrigé} = (\text{Score de la catégorie 1} \times \text{Poids de la catégorie 1}) + (\text{Score 2} \times \text{Poids 2}) + \dots$$

$$\text{Résultat corrigé} = (48 \times 0,5) + (38 \times 0,5) = 43.$$

Somme toute, l'écart entre le résultat initial et le résultat corrigé n'est pas très grand : cela est dû au fait que notre échantillon est quand même relativement représentatif et que les différences entre les perceptions des hommes et des femmes ne sont pas énormes.

3.3. Prévoir les erreurs possibles

Examinons à nouveau l'enquête sur la famille et les amis pour faire l'inventaire des biais ou erreurs possibles, autres que ceux qui seraient dues au hasard.

L'erreur de couverture provient de la différence entre la population cible et la population observée.

Dans cette enquête, comme cela est souvent le cas, on exclut, pour des raisons de coût, ceux qui n'ont pas le téléphone. Ça fait peu de monde en apparence (donc le biais devrait être faible), mais dans certains cas, il y a une relation entre les questions et les groupes exclus. Par exemple, ceux qui n'ont pas le téléphone ont de bonnes raisons de ne pas appeler aussi souvent leur grand-mère (voir le tableau 7.4). Dans un cas comme celui-ci, le biais auquel on doit se résigner serait alors plus grand. Par contre, on peut remarquer que dans le tableau 7.4, les gens qui vivent avec leurs grands-parents ont été éliminés de la population observée. Étant donné que ces gens communiquent rarement par lettre, voilà une précaution qui s'avère à la fois peu coûteuse et fort utile.

Tableau 7.4 - Allô grand-maman!

Fréquence des contacts par lettre ou par téléphone avec les grands-parents chez les Canadiens âgés de 15 ans et plus et ne vivant pas avec leurs-grands-parents

	15-24 ans	25-44 ans	45 ans et plus	15 ans et plus
	(en milliers d'individus)			
Tous les jours	151	57	...	211
Au moins une fois par semaine	577	242	...	822
Au moins une fois par mois	857	457	...	1330
Moins d'une fois par mois	890	1119	...	2033
Pas au cours de 12 derniers mois	670	1058	38	1766
Non déclaré	15	8	37	14
Total	3160	2941	75	6176

Source : Statistique Canada, Enquête sociale générale (la famille et les amis), 1994. Données de 1990.

Les gens vivant en institution ont été exclus de l'échantillon. Cela est dû à une raison pratique : rappelez-vous, on choisissait, dans chaque foyer appelé, une personne au hasard. Cette méthode peut causer un certain biais dans les questions touchant les personnes âgées. On peut d'ailleurs avoir une idée du biais en sachant que 9 % des personnes de 65 ans et plus vivent en institution, ce qui n'est pas négligeable, mais qui n'est pas énorme non plus.

Les refus de répondre amoindrissent la représentativité de l'échantillon.

Les refus de répondre ne sont pas le résultat du hasard : ceux qui refusent de répondre correspondent à un groupe particulier (par rapport au sujet étudié). C'est pourquoi on lutte toujours contre les refus de répondre qui amoindrissent la valeur des résultats.

Les personnes qui ne donnent pas de réponse dans l'enquête sur la famille diffèrent peut-être du reste de l'échantillon : 14 % des non-réponses proviennent du ménage et les 10 % restant proviennent de l'individu. Le taux de réponse global est assez bon dans les circonstances, mais il faut aussi tenir compte des non-réponses à certaines questions, ou d'une mauvaise compréhension des questions, ou de l'influence de certaines réponses sur d'autres réponses.

Tout était beau sur papier, mais...

Imaginez que la base de sondage (annuaire téléphonique ou fichier administratif contenant la liste d'une population) soit incomplète, fautive, désuète, difficile à consulter. On a de la difficulté à rejoindre les gens choisis.

Le degré de précision ne correspond pas à ce qu'on attend. Au cours de l'enquête, on a eu des problèmes avec l'échantillon (erreurs, refus de répondre, abandons en cours de route, remaniement de l'équipe de chercheurs) et la marge d'erreur est devenue trop grande par rapport à ce qu'on s'était fixé au départ.

Certaines personnes comprennent mal les questions. D'autres sont intimidés par des questions qu'ils jugent trop personnelles. Certains ne savent pas vraiment quoi répondre. D'autres enjolivent la réalité ou se contredisent.

On le voit, les embûches qui peuvent surgir lors de l'administration d'un sondage sont nombreuses, c'est pourquoi il faut bien prévoir les coups.

EXERCICES 3

1. La grappe

L'enquête sur la population active, basée sur un échantillon de 14 000 personnes au Québec, indique que la population gaspésienne de 15 ans et plus est, en 1994, de 83 700 individus (pour 5 730 000 au Québec). On suppose d'autre part que le poids d'échantillonnage est de 400 dans toutes les régions du Québec.

- Quelle est la proportion de Québécois de 15 ans et plus résidant en Gaspésie?
- Utilisez la proportion calculée en *a* pour déterminer approximativement la taille de l'échantillon interrogé en Gaspésie.
- Utilisez le poids d'échantillonnage pour déterminer approximativement la taille de l'échantillon interrogé en Gaspésie.

2. Comment obtenir de bons résultats avec un mauvais échantillon

Un sondage révèle que les non-francophones s'apprêtent à voter « non » à 90 % et que les francophones s'apprêtent à voter « oui » à 65 %. On constate également que l'échantillon observé comporte 40 % de non-francophones (données fictives).

- D'après vous, est-ce que le « oui » a des chances de passer? Justifiez votre réponse avec des chiffres.
- On s'aperçoit soudain que l'échantillon n'est pas vraiment représentatif car, d'après le dernier recensement, les non-francophones représentent 20 % de la population en âge de voter). Compte tenu de ce nouveau fait, y a-t-il encore moyen de faire une prévision sérieuse avec ce sondage? Si oui, que prévoyez-vous faire?

3. Qu'est-ce qu'un chômeur?

Selon une enquête de Statistique Canada, le Québec comptait, en 1956, 1 535 000 personnes occupées et 80 000 chômeurs sur une population totale de 4 628 000 individus. Quelle était alors la proportion de chômeurs dans la population active (ou *taux de chômage*)?

4. DES SONDAGES À TOUTES LES SAUCES

On associe souvent les sondages aux campagnes électorales. Mais si ce genre d'enquête constitue le moyen privilégié de sonder les cœurs, il peut s'appliquer à d'autres domaines. On peut sonder l'opinion des gens (caractéristiques subjectives), mais aussi leur état de santé (caractéristiques objectives), ou encore leur comportement et les conséquences de leur comportement. Les trois exemples que nous vous proposons maintenant correspondent à ces trois types de sujets d'enquête.

4.1. L'opinion des gens (vis-à-vis d'eux-mêmes et vis-à-vis des autres)

Qu'est-ce qu'un politicien? Un politicien est une personne qui est capable de discourir avec facilité sur n'importe quel sujet. Qu'est-ce qu'un homme d'État? Un homme d'État est une personne qui est capable de se taire.

Juger les autres et les catégoriser constituent des activités préférées des êtres humains en quête de sécurité intellectuelle et affective. C'est pourquoi les sondages d'opinion, comme celui du tableau 7.5, sont toujours populaires. Si on se fie au sondage, on doit admettre que les politiciens n'ont pas la cote. Et vous, que pensez-vous des résultats? Signifient-ils que les gens font plutôt confiance à ceux qu'ils considèrent comme plus savants qu'eux?

Tableau 7.5 - Un sondages sur l'opinion des gens : Qui est le plus honnête?

Canada 1992 : Qui vous paraît le plus honnête et le plus intègre? (Répartition)

États-Unis 2013 : Comment évaluez-vous le degré d'honnêteté et d'intégrité des gens exerçant les professions suivantes? (Proportion de réponses *Très élevé / Élevé*)

Canada 1992		États-Unis 2013	
	(en %)		(en %)
Les médecins	24	Les infirmières	82
Les scientifiques	21	Les instituteurs	70
Les professeurs d'université	21	Les médecins	69
Les gens d'affaire	12	Les membres du clergé	47
Les journalistes	10	Les banquiers	27
Les avocats	4	Les journalistes	21
		Les vendeurs	
Les politiciens	2	d'automobiles	9
Autres	6	Les membres du Congrès	8
<i>Taille de l'échantillon</i>	1500		

Sources : Canada : sondage fait en décembre 1992 par Decima Research à travers le Canada dans l'Almanach moderne 1994. États-Unis : Gallup.

Note : Seules 8 des 22 professions choisies par les sondeurs sont reproduites pour les États-Unis.

4.2. L'état des gens

Le cas suivant diffère du précédent à plusieurs égards. On y observe en effet quelque chose d'objectif (le poids des Canadiens et Canadiennes), et on utilise les grands moyens (l'échantillon est grand, les variables nombreuses et soigneusement définies). Ce type d'enquête est donc relativement coûteux. Le tableau 7.6 montre que l'excès (ou l'insuffisance) de *poids* varie avec le *lieu de résidence*, l'*âge*, le *sexe* et le *niveau de revenu*.

Les deux dernières lignes du tableau 7.6 concernent non plus des caractéristiques objectives, mais des opinions (subjectives). Les résultats illustrent l'écart qui existe parfois entre une réalité et la perception qu'on se fait de cette réalité.

Tableau 7.6 - Un sondage sur l'état des gens : l'excès de poids

	Poids sous la normale	Poids recommandé	Excès de poids possible	Excès de poids	Non disponible	Total
	(IMC inférieur à 20)	(IMC entre 20 et 25)	(IMC entre 25 et 27)	(IMC supérieur à 27)		
	(en %)					
Canada	9	47	18	23	3	100
Par région au Canada						
Terre-Neuve	5	41	22	31		100
Québec	11	47	18	22	2	100
Ontario	9	48	17	21	5	100
Colombie- Britannique	9	49	19	23		100
Par sexe au Québec						
Hommes	4	44	24	28		100
Femmes	18	51	12	17	2	100
Par groupe d'âge au Canada						
20-24 ans	15	57	15	10	3	100
55-64 ans	5	39	23	30	3	100
Par niveau de revenu au Canada						
Quintile inférieur	15	44	15	22	4	100
Quintile supérieur	6	50	18	25		100
Perception de soi au Canada						
Hommes	6	55	38			100
Femmes	4	47	48			100

Source : Statistique Canada, Enquête sociale générale (la famille et les amis), 1994. Données de 1990.

Notes : Sur un échantillon choisi de 14 875 personnes de 15 ans et plus, 11 924 personnes ont accepté de répondre (le taux de réponse est donc de 80,2 %). Le poids a été mesuré en indice de masse corporelle (IMC). L'IMC, qui est le rapport entre le poids (en kg) et le carré de la taille (en m), tient compte du fait que le poids et la taille sont reliés mais de façon non proportionnelle. L'IMC ne s'applique qu'aux personnes de 20 à 64 ans. On a constaté qu'un poids sous la normale (IMC < 20) ou excessif (IMC > 25) augmente les risques de problèmes de santé (mais il ne s'agit là que d'un élément parmi d'autres).

4.3. Le comportement des gens

Deux personnes en état d'ébriété avancée rentrent chez elles en auto. « Fais donc un peu plus attention, dit le passager, tu as failli nous envoyer dans le fossé. — Ah bon? répond le chauffeur, je croyais que c'était toi qui conduisais. »

Le tableau 7.7 montre qu'environ une personne sur 8 conduit parfois avec un verre de trop dans le nez. Par ailleurs, l'âge et le niveau de revenu semblent avoir une influence sur ce comportement. Les résultats doivent cependant être interprétés avec prudence, car pour conduire en état d'ébriété, il est indispensable d'avoir une voiture à sa disposition, ce qui n'est pas donné à tout le monde. Si, par contre, les gens les plus riches sont les plus délinquants, est-ce parce qu'ils ont moins peur des contraventions? Est-ce parce qu'ils ont plus d'automobiles? Est-ce parce qu'ils boivent plus que les autres? Voilà encore une fois un bon sujet de recherche.

Tableau 7.7 - Un sondage sur le comportement des gens : l'alcool au volant

Proportion des consommateurs d'alcool ayant conduit moins d'une heure après avoir pris deux verres d'alcool et plus

Selon l'âge du consommateur		Selon la tranche de revenu du consommateur	
	(en %)		(en %)
15-19 ans	12	Quintile inférieur	12
20-24 ans	14	2e quintile	8
25-34 ans	16	Quintile moyen	12
35-44 ans	14	4e quintile	14
45-54 ans	13	Quintile supérieur	17
55-64 ans	9		
65 ans et plus	4		

Source : Tendances sociales canadiennes, Automne 1995, p. 23.

Note : Les pourcentages ne s'additionnent pas : il ne s'agit pas d'une distribution.

EXERCICES 4

1. Ce qu'on peut sonder

Donnez des exemples d'enquêtes qui porteraient sur chacun des types de sujets suivants : l'opinion des gens, l'état des gens, le comportement des gens, les conséquences d'un comportement.

2. Sommes-nous trop gros? Trop maigres? Trop parfaits?

(Utilisez les données du [tableau 7.6](#) pour répondre aux questions.)

- a) Identifiez les variables étudiées.
- b) L'excès de poids selon le sexe : commentez les écarts entre la réalité observée et la perception que se font les hommes et les femmes de cette réalité.
- c) D'après vous, de quelle catégorie faites-vous partie : poids sous la normale, poids recommandé ou poids excessif?
- d) Calculez votre IMC (en divisant votre poids en kilogrammes par le carré de votre taille en mètres). Comparez le résultat à votre perception telle que spécifiée dans la question c.

EXERCICES SUPPLÉMENTAIRES

1. Non représentatifs

Dans chacun des cas suivants, quels sont les défauts des échantillons?

- a) Les douanes américaines décident d'ouvrir toutes les boîtes de thon du Nouveau-Brunswick pour vérifier si elles sont avariées.
- b) Un laboratoire prélève 20 échantillons du sang d'un chauffard pour déterminer s'il est en état d'ébriété.
- c) Le bureau national de statistiques profite du recensement pour calculer le nombre de chômeurs.
- d) Un institut de sondage interroge 30 000 citoyens pour connaître leurs intentions de vote.
- e) Allô! Êtes-vous un itinérant?

2. Stratifications

Dans chacun des cas suivants, dites quelles précautions on doit prendre pour s'assurer de la représentativité de l'échantillon.

- a) Allô! Êtes-vous d'avis que le Canada demeure une monarchie?
- b) Allô! Aimez-vous la musique *Heavy Metal*?
- c) Allô! Pensez-vous que le geste du Premier ministre du Canada, qui a saisi Bill Clennet à la gorge le 15 février 1996, est « acceptable dans les circonstances »?
- d) Allô! Pensez-vous que les tribunaux devraient être plus sévères dans le cas des pensions alimentaires?

3. Le taux de réponse

Un sondage de la Sofres indique que 39 % des Français pensent qu'il faut restreindre le droit de grève dans les services publics. Si on y regarde de plus près, on s'aperçoit que 56 % des sympathisants de la droite endossent ce point de vue contre seulement 23 % des sympathisants de la gauche. Le sondage a été effectué entre le 26 et le 28 décembre 1994 auprès d'un échantillon national représentatif de la population française âgée de 18 ans et plus; les personnes ont été interrogées à leur domicile. (*Source* : Le Figaro, 27 février 1995)

- a) Expliquez pourquoi il est intéressant d'utiliser un échantillon stratifié dans une telle enquête.
- b) En supposant que 1024 personnes aient accepté de répondre et que 312 aient refusé ou n'aient pas pu être rejointes, quel est le taux de réponse de l'échantillon?
- c) En supposant que le taux de réponse est de 76,6 % et que 1336 personnes ont été approchées, quel est le nombre de personnes qui ont refusé de répondre?

4. La moyenne des échantillons

Quatre anthropologues possèdent respectivement 1, 4, 5 et 6 crânes de l'homme de Neandertal.

- a) Quel est le nombre moyen de crânes possédé par un anthropologue?

b) En vous inspirant du [tableau 7.3](#) et de la [figure 7.1](#), faites la liste de tous les échantillons de 2 éléments. Calculez le nombre moyen de crânes de chaque échantillon et tracez-en la distribution sur un graphique.

5. Médor ou Minou

Les êtres humains aiment parfois s'entourer d'animaux de compagnie. Un journal vous commande un sondage pour faire un portrait de la situation au Québec et vous demande de faire quelques suggestions sur les variables à mesurer. Ces variables, qui peuvent porter sur le comportement des gens face aux animaux de compagnie ou sur les caractéristiques de ces gens, doivent susciter l'intérêt des lecteurs du journal.

a) Trouvez quelques variables concernant le comportement des gens face aux animaux de compagnie. Si les échelles des variables sont nominales, identifiez leurs catégories.

b) Trouvez quelques variables concernant les caractéristiques des gens qui ont des animaux de compagnie, ou qui n'en ont pas. Si les échelles des variables sont nominales, identifiez leurs catégories.

c) Vous devez sélectionner un échantillon stratifié et tenir compte des variables choisies. Parmi les caractéristiques suivantes, indiquez celles de la population sondée que l'échantillon devrait absolument refléter : sexe, âge, langue maternelle, religion, niveau de revenu, type d'habitat (ville, campagne), région de résidence, orientation politique, poids, niveau de scolarité, état civil.

Note : Dans ce genre de remue-méninges, une bonne façon de trouver des idées consiste à faire appel aux préjugés courants : les grands-mères et les jeunes filles préfèrent les chats, les cultivateurs et les gens de droite préfèrent les chiens, etc. Il sera toujours temps d'infirmier — ou de confirmer — ces préjugés en examinant les résultats du sondage.

6. Le Québec n'a pas la cote... au Canada

En 1996, alors que 500 000 « cousins » français envahissent le Québec comme chaque été, cette destination semble avoir perdu la cote auprès des « frères » canadiens, suite à un référendum très serré sur la question nationale. Un sondage effectué par le CAA indique que les destinations préférées des Canadiens sont, dans l'ordre, la Nouvelle-Écosse, l'Ontario, la Colombie-Britannique, la Floride, l'Alberta, la Caroline du Sud, le Massachusetts, l'État de New York, la Californie et... le Québec. Un second sondage classe, outremer, l'Angleterre première, suivie de la France, de l'Écosse et de l'Allemagne. (Source : *La Presse*, 20 juillet 1996)

a) Quelles sont les deux variables dont il est question dans ce sondage?

b) À quelle échelle appartiennent ces variables? Dites pourquoi le choix de cette échelle est judicieux.

c) Que pensez-vous du fait que les résultats publiés dans *La Presse* ne contiennent aucun chiffre?

7. Laboratoire

a) Représentez graphiquement les données des [tableaux 7.1a et 7.1b](#).

b) Représentez graphiquement les données des deux premières colonnes du [tableau 7.4](#).

c) Représentez chaque colonne du [tableau 7.7](#) sous forme de diagramme en bâton.

DOSSIER 7 QU'EST-CE QU'UN QUÉBÉCOIS?

Dans ce dossier, nous nageons en pleins stéréotypes, terrain miné et passionnant. Qu'est-ce qu'un Québécois ou une Québécoise? Comment se voient-ils? Quelle image d'eux-mêmes voudraient-ils projeter? Et, à l'inverse, pour qui détesteraient-ils qu'on les prenne? Y a-t-il des vrais et des faux Québécois? Parmi les portraits-robots les plus fréquents, nous avons d'ailleurs relevé les types de Québécois suivants : le patenteux (imaginatif mais sans argent), le cultivateur (vaillant et terre à terre), la femme d'affaires (très compétente, mais vulnérable sur le plan sentimental), le harceleur (qui finit par se faire taper sur les doigts), le fêtard (qui aime boire un coup, mais qui n'est pas méchant), le malhonnête (c'est en réalité un individu qui n'a pas peur d'affronter la tyrannie de la société, et que son père battait), etc.

La petite étude qui suit a été réalisée en 1995. Cela permettra aux jeunes étudiants de mieux apprécier la grandeur des générations qui les ont précédés.



Cinq Québécoises et cinq Québécois des années 2000

Pour observer les stéréotypes, rien ne vaut la télévision. Dans les films ou les téléromans, certains rôles sont universels : l'environnement physique et culturel s'efface alors devant l'être profondément humain qui occupe l'écran. On oublie alors que la scène se passe dans le Kalahari ou sur les rives du fleuve Jaune. Dans d'autres cas, au contraire, le personnage ne se déplace jamais sans une panoplie de clichés : expressions locales, « sacres », accents, ceinture fléchée, tabous à ne pas transgresser, gros conformisme, manies. Le personnage est d'abord Québécois avant d'être médecin, secrétaire, financier, artiste, dragueur, cocu, déprimé, perdu, manipulateur, farceur ou bienfaiteur.

Les Québécois vus par eux-mêmes et vus par les autres

Pour en savoir plus, il faut encore une fois faire appel à des méthodes quantitatives : ici, le sondage d'opinion. Nous vous proposons d'examiner les 11 affirmations suivantes et, pour mettre un peu de piquant, nous vous demandons d'estimer vous-même la proportion de Québécois, d'Ontariens et de résidents de la Colombie-Britannique qui auraient été d'accord avec chacune d'entre elles (la 12^e rubrique du sondage est fournie en prime!). D'après vous, par exemple, y a-t-il beaucoup de Québécois qui se définissent d'abord comme Québécois plutôt que comme Canadien? Les réponses

au sondage viendront plus loin : vous aurez alors la chance de mesurer l'écart entre les perceptions des gens et l'idée que vous vous faites de ces perceptions.

- 1. Je me définis d'abord comme résidant de ma province plutôt que comme Canadien.
- 2. Nous devrions manger, boire et être heureux, car nous ne savons pas ce que l'avenir nous réserve.
- 3. Les femmes ayant de jeunes enfants devraient rester à la maison.
- 4. Ça ne me dérangerait pas si un de mes enfants était homosexuel.
- 5. Il est acceptable de fumer de la marijuana.
- 6. Le mois dernier, j'ai lu un livre.
- 7. Les Québécois sont *chialeux*.
- 8. Les Ontariens sont violents.
- 9. J'interdirais à mes invités (à souper) de fumer dans ma maison.
- 10. Le nouveau gouvernement fédéral va privilégier le Québec.
- 11. L'adultère est inacceptable
- 12. Combien de relations sexuelles avez-vous par mois, en moyenne?



Cinq Québécois des années 1950
(Dessins de Renaud Bouret)

Croyez-vous au père Noël?

Les vieilles superstitions sont d'autant plus tenaces qu'elles permettent à certains d'accepter leur sort, et à d'autres d'éviter de regarder les choses en face. Ainsi, une écrasante majorité d'Occidentaux croient encore que l'astrologie est une science exacte.

Jupiter est aligné avec les roues de ma bicyclette.

« Je suis né le 24 septembre à 6 h 12, au moment où Jupiter était aligné avec le Verseau : c'est pourquoi j'aime l'eau, la foudre et le bingo (je joue le 24-6-12) ». Avec de telles précisions, il paraît difficile de mettre en doute le sérieux d'une telle affirmation. Et pourtant, que signifient ces données? S'agit-il de l'heure d'été, de l'heure d'hiver, de l'heure de la capitale, de l'heure moyenne? S'agit-il du calendrier babylonien, julien, grégorien, ou d'une année bissextile? Jupiter serait-elle toujours alignée si on la regardait depuis Mars? Faut-il tenir compte de la précession des équinoxes, qui a fait en sorte que l'apparition des constellations du zodiaque, décalée d'un bon mois depuis l'antiquité, ne correspond plus aux dates officielles des signes astrologiques? Il n'est pas difficile de constater que l'astrologie n'a rien à voir avec la rigueur scientifique que nous prêchons dans ce manuel, ce qui n'enlève rien à son intérêt en tant que sujet d'étude des croyances humaines. Le *New Scientist* de Londres, a interrogé 25 000 personnes dans plus de 20 pays (Source : *Courrier international*, 8 juin 1995). C'est le Canada, où 69 % de gens qui pensent que l'astrologie est une

science exacte, qui obtient la médaille d'or de la superstition. Les autres pays occidentaux suivent. Les Russes, comme la plupart des habitants de l'ancien bloc communiste, demeurent plus sceptiques : 18 % d'entre eux croient à l'astrologie.

Le nouveau pari pascalien.

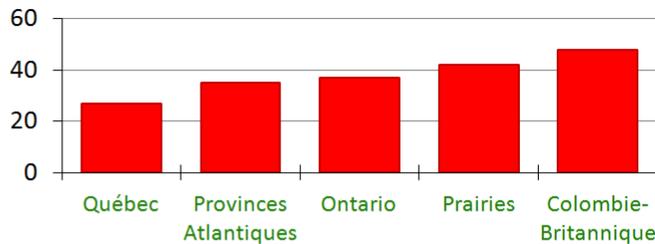
Faut-il croire au père Noël, ne serait-ce que pour ne pas le fâcher au cas où il existerait vraiment? C'est ce que font 37 % des Canadiens adultes (Source : La Presse, 21 décembre 1994, d'après un sondage Gallup). Diriez-vous que les Québécois y croient encore plus que les autres? Vous auriez tort, car c'est au Québec que le scepticisme est le plus élevé : 27 % croient au père Noël, contre 37 % en Ontario et 42 % en Colombie-Britannique (voir figure D7.1). Plus on va vers l'ouest, plus on est crédule : faut-il y voir l'influence de quelque planète?

Figure D7.1 - Qui croit encore au Père Noël?

Proportion de gens qui disent croire au Père Noël (en %)

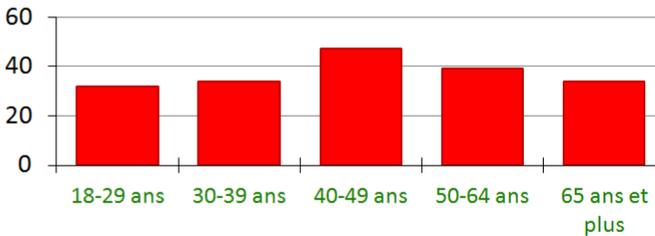
Par région

Québec	27
Provinces Atlantiques	35
Ontario	37
Prairies	42
Colombie-Britannique	48
Canada	37



Par groupe d'âge

18-29 ans	32
30-39 ans	34
40-49 ans	47
50-64 ans	39
65 ans et plus	34



Par sexe

Hommes	39
Femmes	34

	Canada	Québec
Marge d'erreur à 95 %	3,1%	6%
Taille de l'échantillon	1 002	268

Source : La Presse, mercredi 21 décembre 1994. Sondage Gallup effectué auprès des Canadiens de 18 ans et plus.

Des fleurs s'il vous plaît, on aime ça

Le consommateur est roi.

Le dernier sondage que nous vous proposons dans ce dossier porte sur le consommateur québécois. On lit dans *Les Affaires* du 21 janvier 1995 que « le consommateur québécois est un débrouillard qui s'adapte facilement... Le consommateur québécois est devenu entrepreneurial... Le consommateur est devenu astucieux dans ses achats et il a appris à profiter d'une série de situations qui lui permettent d'augmenter sa capacité d'acquérir des biens ». Bien sûr, le sondage ne se limite pas à cela : on y retrouve une série de résultats chiffrés.

Tout ce jargon, digne d'un vendeur d'aspirateurs, n'est pas aussi vide de sens qu'il en a l'air au premier abord. Après tout, les courtisans ne s'y prenaient pas autrement lorsqu'ils s'adressaient au souverain.

Que diriez-vous si les journaux d'affaires publiaient des enquêtes montrant les tares et défauts des consommateurs? Auriez-vous encore envie d'acheter ces journaux? Il faut être conscient que les sondages qui paraissent dans les médias ne sont pas toujours totalement innocents.

QUESTIONS

1. Les Québécois vus par eux-mêmes et vus par les autres

a) D'après vous, quelle est la proportion de Québécois, d'Ontariens et de résidents de la Colombie-Britannique qui ont répondu oui à chacune des 11 premières questions de sondage énoncées un peu plus haut? Indiquez également votre estimation pour la question 12.

b) Comparez vos réponses avec les résultats du sondage qui figurent au [tableau D7.1](#). Calculez l'écart absolu entre votre estimation et le résultat du sondage (pour les 11 premières questions seulement). Calculez le total des écarts par colonne et par ligne. Comparez vos résultats avec ceux de vos camarades.

2. Recherche : portrait-robot du Québécois

Cette étude peut être faite en groupe et présentée ensuite à la classe.

a) En observant les stéréotypes véhiculés à la télévision, dans les journaux ou dans votre milieu, dressez une liste de caractéristiques du Québécois et de la Québécoise typique.

b) Pour chaque caractéristique retenue, formulez une ou plusieurs questions de sondage dans lesquelles le répondant devra exprimer son accord ou son désaccord. La question peut être formulée de façon positive ou négative et elle peut porter à la fois sur l'individu interrogé et son entourage. Voici trois exemples :

- Les Québécois sont honnêtes
- Les Québécois trichent souvent.
- Je triche souvent.

c) Pendant que vous y êtes, formulez aussi des questions pour voir si les Québécois croient vraiment à l'astrologie et au père Noël.

d) Testez vos questions sur un échantillon d'individus. Indiquez dans quelle mesure votre échantillon est représentatif. Compilez les résultats. Commentez-les.

Tableau D7.1 - Réponses au sondage sur les Québécois vus par eux-mêmes... et par les autres

Proportion (en %) des répondants qui ont répondu oui aux questions 1 à 11.

Moyenne des réponses pour la question 12.

<i>Question</i>	Québec	Ontario	Colombie-Britannique
1	49	9	17
2	71	32	17
3	40	30	32
4	85	49	52
5	22	26	32
6	64	78	85
7	38	66	62
8	9	47	25
9	14	38	50
10	7	22	46
11	55	78	67
12	8	6,4	6,9

Source: MacLean's, 3 janvier 1994.