MÉTHODES QUANTITATIVES ET SCIENCES HUMAINES

Pour prendre la mesure du monde

2e ÉDITION

Renaud Bouret

2014

MÉTHODES QUANTITATIVES ET SCIENCES HUMAINES
Deuxième édition
Auteur : Renaud Bouret
Illustrations : Renaud Bouret
Éditeur : Renaud Bouret Éditeur
Gatineau, 2014
ISBN : 978-2-9813201-6-2 (format epub)
(Publié sous format électronique uniquement)

Première édition publiée chez Chenelière Éducation, Montréal, 1997
Auteur : Renaud Bouret (avec la collaboration de François Cauchy)
ISBN : 978-2-8946-1074-9

L'éditeur a fait tout ce qui était en son pouvoir pour retrouver les copyrights. On peut lui signaler tout renseignement menant à la correction d'erreurs et omissions. (NB : La deuxième édition reprend une bonne partie du texte de la première édition.)

DÉNI DE RESPONSABILITÉ

Ce manuel est offert gratuitement, dans un but uniquement pédagogique. Il peut contenir des erreurs. L'auteur ne pourra être tenu responsable de problèmes éventuels causé par l'utilisation de ce manuel et du matériel associé.

DROITS DE REPRODUCTION

AVANT-PROPOS

Il est difficile aujourd'hui d'échapper aux chiffres lorsque l'on étudie la réalité humaine. Les grands penseurs, nous dira-t-on, ne s'encombraient pas de courbes et de tableaux. Cependant, les méthodes et les outils qui se sont généralisés au cours des dernières décennies, comme les enquêtes par sondage, les bases de données ou les chiffriers électroniques, permettent d'aller plus loin dans l'étude des phénomènes humains. Qui plus est, pourvu qu'ils soient bien présentés, ces outils et méthodes sont faciles à utiliser.

Si les chiffres permettent de dépasser la simple intuition, ils reflètent la réalité humaine de façon très approximative. La quantification des concepts humains est souvent ardue. « Qu'est-ce qu'un pauvre? », par exemple? Voilà une notion qui est plus difficile à définir et à mesurer que le taux d'ampoules électriques défectueuses sur une chaîne de production. Les erreurs de mesures sont souvent plus grandes en sciences humaines que dans les sciences « exactes ». S'il est rare que l'ampèremètre mente ou refuse de répondre, il en va tout autrement lorsqu'on cherche à cerner un concept humain. D'autant plus que les définitions de ces concepts peuvent aussi varier d'un endroit ou d'une époque à l'autre.

Par où commencer? La plupart de nos concurrents partent essentiellement de la théorie mathématique et cherchent ensuite à appliquer cette théorie au domaine concret des sciences humaines. Même si ce passage de la théorie à la pratique est généralement fait avec talent, nous croyons que cette démarche est un peu artificielle et qu'elle ne convient pas à la majorité de nos étudiants. En fait, nous nous proposerons de faire exactement l'inverse, un peu comme l'artiste-peintre qui commence par découvrir la richesse et la beauté du monde avant de s'intéresser à la science des pinceaux, des huiles et de la perspective. Tout au long de ce manuel, nous partirons de la réalité humaine et nous chercherons à découvrir, de façon systématique, les outils quantitatifs aidant à mieux comprendre les phénomènes étudiés.

Il ne s'agit donc pas de déballer une panoplie d'outils quantitatifs et de leur chercher une application en sciences humaines. Il s'agit au contraire de faire le tour des phénomènes humains et d'identifier les outils quantitatifs qui permettent de mieux les appréhender.

La presque totalité des chiffres cités dans ce livre sont de « vrais » chiffres et ont une portée suffisamment vaste pour intéresser tous nos lecteurs. Nous essayons de toucher à tous les aspects des sciences humaines, sans nécessairement nous préoccuper des traditionnels découpages disciplinaires. En évitant les exemples fictifs (dans lesquels les humains peuvent le plus souvent être facilement remplacés par des boulons ou des ampoules électriques) et les exemples de portée restreinte (les notes obtenues dans l'école Unetelle), nous poursuivons un double objectif. Le premier est pédagogique : nous pensons que la curiosité que suscitera chez le lecteur la présentation de situations bien réelles lui permettra de s'intéresser véritablement aux outils exposés et de mieux les retenir. Le second est d'ordre plus général et certains nous jugeront un peu présomptueux : nous faisons le pari de développer chez notre lecteur la passion des sciences humaines et de lui faire voir les outils quantitatifs comme des alliés et non comme des ennemis à maîtriser.

Chaque chapitre permettra à l'étudiant d'absorber, de façon progressive, les notions de base des méthodes quantitatives en sciences humaines, tout en suscitant son intérêt pour l'objet de ces méthodes. Nous chercherons aussi à éviter l'écueil classique qui consiste à déballer une série de termes techniques et à essayer ensuite de les rattacher à un concept. À titre d'exemple, nous défions quiconque d'être en mesure de faire une distinction claire entre un *taux*, un *ratio*, un *rapport*, un *pourcentage* ou une *proportion*. Et pourtant tous les ouvrages tentent de définir ces termes de façon précise. Comme dans bien des cas, il s'agit en réalité d'un faux problème, qui résulte d'un douteux amalgame du jargon des diverses sciences humaines et des mathématiques, et du parler de l'homme de la rue. Ainsi que nous le verrons dans ce manuel, la proportion n'est qu'une forme particulière de rapport (la partie divisée par le tout); le pourcentage n'est pas une mesure en soi, mais plutôt une façon de présenter une mesure; quant aux autres termes, leur définition varie énormément selon le contexte et l'usage.

Dans notre ouvrage, nous proposons également une série de dossiers, intercalés entre les chapitres. Ces dossiers touchent à des domaines les plus divers et recouvrent souvent des notions vues dans plusieurs chapitres. Cela permettra à l'étudiant d'aborder des cas bien concrets et de développer son esprit de synthèse. Dans ces dossiers, plus encore que dans les chapitres, nous insisterons sur l'interprétation des chiffres plutôt que sur les techniques de calcul, qui sont d'ailleurs très simples la plupart du temps.

L'ensemble de l'ouvrage, chapitres et dossiers, forme un tout cohérent. Les exercices y sont variés, abondants et riches. Nous revenons constamment sur les concepts déjà vus, tout en cherchant à les approfondir petit à petit. Nous invitons donc le lecteur à se montrer patient : les nuances, les détails, la mise en évidence des liens entre les divers éléments viendront en temps et lieu.

Le lecteur sentira, tout au long de ses pages, la présence d'un ami et d'un guide sûr, qui veut partager avec lui ses modestes connaissances et sa passion des sciences humaines et qui essaiera de temps en temps de le faire sourire. Bonne lecture!

Renaud Bouret

TABLE DES MATIÈRES

1. LES CHIFFRES BRUTS

1. Pourquoi mesurer?
2. Mesurer et dénombrer
3. Unités de mesure
4. Présenter les données

2. LES RAPPORTS

1. Qu'est-ce qu'une proportion?
2. La proportion sous toutes ses formes
3. Les fréquences relatives
4. D'autres rapports : les comparaisons

3. AUTOUR DE LA MOYENNE

1. La moyenne : un équilibre des forces
2. L'écart type : mesurer la dispersion des données
3. La courbe normale et l'écart type
4. D'autres indicateurs de dispersion

4. LES DONNÉES CHRONOLOGIQUES

1. Le taux de variation
2. L'indice de variation
3. Les stocks et les flux
4. Les variations à long terme

5. LES INDICES

1. Qu'est-ce qu'un indice?
2. Les indices synthétiques
3. L'indice du développement humain
4. L'indice des prix à la consommation

6. L'ANALYSE DE TABLEAUX

1. Trois clés : Définir, observer, interpréter
2. Analyser un tableau : faut-il se faire des cheveux blancs?
3. Analyser un graphique : les filles en pantalon
4. Construire ses propres tableaux

7. LES ENQUÊTES PAR SONDAGE

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

8. ESTIMATION ET TEST D'HYPOTHÈSE

1. Estimer des proportions
2. Estimer une moyenne
3. Vérifier une hypothèse à l'aide d'un échantillon
4. Test d'hypothèse sur deux moyennes

9. LES RELATIONS ENTRE VARIABLES

1. Schéma de relations
2. Relation entre deux variables qualitatives
3. Relation entre deux variables quantitatives
4. Les accidents de la route

DOSSIERS

2. Le tour du monde en rapports
3. La réforme de l'écriture chinoise
4. Cinq femmes pour un homme
5. La bombe démographique
6. Le peuple réclame du pain
7. Qu'est-ce qu'un Québécois?
8. Des élections historiques
9. **Un tabou universel**

CHAPITRE 1 LES CHIFFRES BRUTS

TABLE DES MATIÈRES

1. Pourquoi mesurer?

2. Mesurer et dénombrer

3. Unités de mesure

4. Présenter les données

• Exercices supplémentaires

L'être humain est un sujet d'étude inépuisable, et passionnant. Voilà d'ailleurs des millénaires que les philosophes, les moralistes, les écrivains, et enfin les savants se penchent sur son cas.

Pour l'étudiant, il existe deux façons d'aborder l'étude des sciences humaines. L'une consiste à se fier aveuglément à ses propres passions ou au jugement d'autrui. L'autre consiste à rechercher la vérité dans les faits, soit par l'observation directe, soit en puisant son information dans la mine d'or des données existantes. Or, les faits observés font souvent l'objet d'une quantification ou d'une comptabilisation, et les données ainsi récoltées prennent alors la forme de chiffres. Il s'agit ensuite d'interpréter ces chiffres à l'aide de quelques outils efficaces et utilisés *à bon escient*. C'est justement l'objet des *méthodes quantitatives en sciences humaines*.

L'esprit est souvent victime d'intuitions trompeuses lorsqu'il doit évaluer des données chiffrées, c'est pourquoi il est d'autant plus important d'apprendre à analyser ces données de façon intelligente et efficace.

Pour beaucoup d'étudiants en sciences humaines, l'expression « méthodes quantitatives » réveille instantanément de mauvais souvenirs. Ces étudiants croient bientôt être confrontés à des problèmes de haute statistique, qui dépassent largement leurs compétences. Mais rien n'est plus faux. Il n'est pas question d'avoir recours à de savantes formules mathématiques, mais d'aborder les données chiffrées de façon *méthodique*. Le contenu mathématique des méthodes quantitatives se limite essentiellement aux opérations arithmétiques de base.

Contrairement à ce que l'on peut trouver dans la plupart des manuels de méthodes quantitatives, la quasi-totalité des exemples présentés dans ce manuel est tirée de la réalité humaine, dans toute sa richesse et sa diversité. Pour développer une approche méthodique, il est en effet primordial d'aiguiser son esprit devant de véritables problèmes de sciences humaines. Pour acquérir une bonne méthodologie, il s'agit avant tout de résoudre des problèmes concrets, et non d'appliquer tant bien que mal des formules abstraites à la complexité des phénomènes humains.

Étant donné l'ampleur du domaine étudié, les méthodes quantitatives en sciences humaines couvrent des concepts très variés. Le sociologue, le psychologue ou l'anthropologue s'intéresseront, par exemple, à des populations composées d'individus, et aux caractéristiques de ces individus (âge, sexe, opinion politique, consommation d'alcool, etc.). L'historien, le politologue ou l'économiste se pencheront fréquemment sur des données chronologiques (démographie, scores électoraux, production, etc.). Toutes ces données seront ensuite transformées, afin de les rendre plus

« parlantes » : on calculera, par exemple, la moyenne d'âge des motocyclistes, la proportion de femmes à l'université ou le taux de croissance de la production de maïs transgénique. Les six premiers chapitres de ce manuel seront consacrés à ce type de sujets, que l'on peut qualifier d'essentiellement *descriptifs*. Par ailleurs, les données récoltées permettent aussi d'aller plus loin : grâce à des méthodes relativement simples, il est possible, à partir d'un simple échantillon, d'estimer les caractéristiques de toute une population ou d'émettre des hypothèses sur cette même population, tout en mesurant le degré d'influence du hasard. Ce sera l'objet des trois derniers chapitres.

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

- Pourquoi est-il important de savoir bien utiliser les chiffres en sciences humaines?
- Pourquoi est-il faux de prétendre que les chiffres mentent?
- Quelles sont les échelles qui permettent de mesurer ou de classifier les variables?
- Comment peut-on représenter graphiquement les informations recueillies?

1. POURQUOI MESURER?

Cette première section du chapitre explique pourquoi il est important de bien manipuler les chiffres lorsqu'on veut comprendre le monde et agir sur lui. Ceux qui sont déjà convaincus de la chose pourront passer directement à la section suivante.

Lorsqu'ils se trouvent confrontés à un problème, nombreux sont les gens qui se fient à leur intuition, plutôt qu'à leur capacité d'analyse. Or, l'esprit humain a ceci de particulier qu'il se laisse souvent berner par ses intuitions, surtout lorsqu'il est confronté à des données quantitatives.

1.1. Réfléchir avant d'agir?

Un premier individu croira, par exemple, qu'une hausse de prix de 100 % sera annulée par une baisse de prix du même pourcentage. Un second individu estimera qu'une température de 20° est deux fois plus élevée qu'une température de 10°. Et si ce second individu se trouve à être un Québécois « moyen », il y a fort à parier qu'il mesure la température de sa piscine et de son aquarium, non plus en degrés Celsius, mais en degrés Fahrenheit, auquel cas il considèrera qu'une eau à 80°F est deux fois plus chaude qu'une eau à 40°F! Par mesure de prudence et d'économie, un troisième individu se déclarera carrément allergique aux chiffres, ce qui le dispensera, croit-il, d'y avoir recours pour se tirer d'embarras. Or, à moins qu'il ne s'agisse de résoudre des problèmes routiniers, cette méthode instinctive (et peu fatigante) se montre rarement efficace.

Comment devrait réagir, par exemple, l'excursionniste qui s'aperçoit qu'il s'est perdu en forêt? Que devrait faire le premier fabricant mondial de microprocesseurs lorsque la presse s'aperçoit que sa « puce » a un léger défaut? Quelle devrait être l'attitude du conseiller municipal qui est appelé à voter sur le financement de la nouvelle salle de spectacle communautaire?

Ne pas avancer à l'aveuglette.

Nous savons tous que le meilleur moyen de ne plus sortir de la forêt est de courir tête baissée (à l'aveuglette) vers un salut très hypothétique. Il vaut mieux, dans un premier temps, s'asseoir sous un arbre, à l'abri du vent, et réfléchir à la situation : quand ai-je vu le sentier pour la dernière fois? Quelle était ma direction? Où étais-je situé par rapport aux routes ou aux grands cours d'eau et à quelle distance? Y a-t-il un bon poste d'observation à proximité? À quelle vitesse puis-je marcher, et pendant combien de temps? Quelle heure est-il et à quelle heure tombe la nuit? Quels sont les bagages que je possède? En quoi peuvent-ils m'être utiles? Quelle est la quantité d'eau et de nourriture qui s'y trouvent, par exemple? Quelle est la meilleure solution au problème et quelles sont les solutions de rechange? Je fais, en quelque sorte, l'inventaire des informations utiles que je détiens, dont plusieurs sont de nature *quantitative*. J'essaie si possible d'obtenir d'autres informations et je m'arrange pour combiner toutes ces données de façon à faciliter la résolution de mon problème. Bref, j'ai tout intérêt à procéder de façon *méthodique*, quitte à me livrer à quelques calculs élémentaires.

Le fabricant de microprocesseurs, quant à lui, peut instinctivement minimiser le problème de la puce ou même le nier. Il devrait pourtant se poser lui aussi quelques questions s'il veut vraiment sortir du bois : quelle est l'ampleur du défaut? Quelle proportion de ma clientèle est touchée? Quelle proportion de ma clientèle est susceptible de s'inquiéter ou de s'indigner? Quels sont les coûts de la réparation? Quels sont les dommages à ma réputation si je ne répare pas? Quelles sont les différentes manières de réparer?

Le conseiller municipal peut aussi suivre ses impressions personnelles et voter selon qu'il apprécie ou non les arts de la scène ou les artistes. Mais pour se prononcer de façon éclairée sur la question de la salle de spectacle, le conseiller devrait se poser les questions suivantes. Quelle est la clientèle actuelle? Quelle est la clientèle souhaitée? Quelles sont les infrastructures existantes? Quels seront les besoins à plus long terme? Quels sont les coûts du projet? Quelle est l'importance que la population accorde au projet? Quelles sont les retombées favorables au projet (emploi, baisse de criminalité)? En quoi le projet favorisera-t-il le développement des autres branches culturelles et d'autres secteurs (la recherche scientifique)?

Bien utiliser les chiffres, c'est, avant tout, bien les choisir.

Dans les exemples qui précèdent, certaines variables sont chiffrables et d'autres pas. Dans ce manuel, nous nous intéressons plus particulièrement aux variables chiffrables : comment devons-nous les mesurer, les comparer et les combiner afin de prendre une décision éclairée? Et avant tout, nous devons nous demander comment les *choisir* et les *utiliser*, compte tenu du problème que nous avons à résoudre et des moyens dont nous disposons.

Notez comment l'excursionniste perdu ne s'encombre pas la tête de détails inutiles et ne perd pas son temps à chercher des informations utiles, mais difficiles à obtenir. De la même façon, notre conseiller municipal choisira de ne pas se préoccuper de l'évolution en bourse du cours des matières premières. Cette variable, bien que chiffrable, est non pertinente ici vu son peu d'effet sur la prise de décision. Par contre, le coût du projet et ses retombées sont des variables qu'il convient de chiffrer et de comparer pour appuyer le processus de prise de décision.

1.2. Les dangers du « pifomètre »

Nous voilà donc convaincus de l'utilité des chiffres : dans bien des cas, ils permettent une meilleure prise de décision. Il nous reste à développer une méthode pour transformer en données numériques les informations que nous jugeons pertinentes. Afin de mieux illustrer l'importance d'une telle méthode, nous utiliserons un instrument de mesure universellement répandu, mais rarement suffisant : le *pifomètre* (du gaulois *pif* : nez, et du grec *mètre* : mesure).

Construction d'une université au pifomètre:

Avis à la population! L'université qui sera enfin construite dans notre ville devra accommoder *pas mal* d'étudiants (tous ceux qui sont susceptibles de poursuivre leurs études dans la région). Il faudra également équiper *un certain nombre* de laboratoires et les gymnases nécessaires et prévoir un stationnement *adéquat* pour les étudiants et le personnel qui se déplacent en automobile. Le bâtiment principal devra être construit *plutôt* en hauteur, compte tenu du *manque* d'espace. Les entrepreneurs sont priés de soumettre leurs devis de construction dans un délai *raisonnable*. Ces soumissions devront être justes et précises.

Cela irait tellement mieux si on disposait de chiffres concrets pour mesurer les différentes variables du problème. Il est clair qu'on ne peut pas déterminer avec certitude le nombre d'étudiants qui s'inscriront effectivement à l'université. Cependant, certaines données facilement mesurables peuvent nous aider à estimer cette variable essentielle. Il s'agit, entre autres :

- du nombre d'étudiants du secteur préuniversitaire de la région;
- du taux de passage du préuniversitaire à l'université dans les autres régions;

- du nombre d'étudiants originaires de la région qui sont actuellement inscrits dans les autres universités (en tenant compte des programmes qui seront offerts dans la région et de ceux qui ne le seront pas).

Il est également possible d'observer l'évolution des clientèles scolaires depuis plusieurs années afin de constater l'importance des contingents qui atteindront l'âge universitaire au cours de la prochaine décennie. La liste des programmes, des cours et des laboratoires qui en font partie, et les clientèles aideront également à déterminer l'ampleur des investissements en laboratoires et gymnases. On aimerait également obtenir des précisions sur la disponibilité des terrains, leur coût et leur superficie.

Nous n'avons fait qu'effleurer le problème, mais nous pouvons déjà constater que beaucoup de chiffres sont disponibles. Il nous manque toutefois une méthode pour choisir ces chiffres et les utiliser efficacement. Bien sûr, l'élaboration du devis de construction ne repose pas uniquement sur des chiffres, car certaines décisions relèvent de choix éducatifs, sociaux ou politiques. Cependant, le problème ne pourra être résolu sans l'utilisation intelligente de données chiffrées.

1.3. Croire ou comprendre?

Va pour la construction, mais les chiffres et les sciences humaines?

Il fut un temps ou les personnes qui œuvraient dans les domaines des sciences humaines misaient surtout sur l'éloquence de leur discours ou la beauté de leur plume pour convaincre leur public. Bon nombre d'écrivains ont peint avec beaucoup de justesse la nature humaine, sans pour autant utiliser le moindre chiffre. Même de grands économistes comme Marx ou Keynes évitaient d'avoir recours aux chiffres dans leurs démonstrations. Aujourd'hui, tout a changé. On use des chiffres (ce qui peut être fort utile comme nous l'avons remarqué plus haut) et, parfois, on en abuse. Voyons plutôt.

Au plus fort de la campagne électorale, le débat des chefs se déroule dans un déluge de chiffres : crochet du droit au taux de chômage, direct du gauche au déficit, *uppercut* au taux de criminalité. Soudain, un des candidats reçoit un sondage d'opinion à la mâchoire, ce qui a pour effet de lui clouer le bec. Peu après, il va au tapis, assommé par la baisse des mises en chantiers dans la construction. L'arbitre arrête le combat.

Que pensent les spectateurs… pardon, les citoyens? Pour certains, les chiffres constituent des arguments sans appel qu'on accepte sans trop les comprendre, mais avec respect, comme un acte de foi. Pour d'autres, plus sûrs d'eux-mêmes, ces chiffres seront réutilisés, après de légères déformations et réinterprétations, pour discuter avec des collègues le lendemain. Il s'agira alors plutôt de dérouter l'adversaire que de le convaincre. Les plus cyniques verront dans les chiffres un moyen sournois de tromper la population. Pour ces derniers, il est inutile de se fatiguer à comprendre : les chiffres mentent, un point c'est tout!

Il existe évidemment une dernière catégorie de gens : ceux qui veulent conserver leur esprit critique; ceux qui veulent savoir ce qu'un chiffre dit et ce qu'il ne dit pas; ceux qui veulent comprendre le monde et la société dans lesquels ils vivent. Voilà, si cela était encore nécessaire, une excellente raison d'étudier l'utilisation intelligente des chiffres en sciences humaines.

EXERCICES 1

1. Discussion

Il est souvent essentiel de *mesurer* avant de pouvoir décider d'une action à entreprendre. Identifiez, pour chacun des cas suivants, quelques informations chiffrées qui pourraient aider à la prise de décision.

a) Un éditeur doit déterminer le tirage d'un manuel.

b) Un collège doit déterminer le nombre de professeurs à embaucher.

c) Une municipalité doit déterminer le nombre de préposés aux parcomètres.

d) La ville de Los Angeles doit contrôler l'accès aux autoroutes pendant les heures de pointe.

e) Le ministère de l'Environnement veut vérifier l'efficacité de sa campagne anti-mauvaises herbes.

f) La régie de l'eau doit déterminer la capacité que devra avoir la nouvelle station d'épuration.

g) Hydro-Québec veut être en mesure de satisfaire à la demande domestique dans les prochaines années.

Note : Faites d'abord par écrit l'inventaire des variables quantifiables. Partagez ensuite vos trouvailles entre vous. D'autres étudiants pourront ajouter ou retrancher des variables. En cas de litige, les étudiants devront être prêts à justifier leurs réponses.

2. Bulletin météo au pifomètre

Reformulez le bulletin suivant avec des informations chiffrées:

« Aujourd'hui, il fera assez beau dans la région. Ce matin, la température est relativement fraîche et l'atmosphère est très humide. Le vent sera plutôt fort. Les skieurs de fond sont priés de s'habiller en conséquence. »

2. MESURER ET DÉNOMBRER

Nous venons de voir pourquoi il était important, en sciences humaines comme ailleurs, de mesurer les faits, de quantifier les variables d'un problème avant de tenter de le régler. Mesure les choses nous aide à mieux les connaître, à mieux les comprendre et à mieux agir sur elles. Après avoir vu le *pourquoi*, nous abordons maintenant le *comment*.

L'étude de l'être humain peut porter sur les personnes elles-mêmes, mais aussi sur des objets ou des évènements. Intéressons-nous d'abord aux humains et à leurs caractéristiques.

2.1. Définir ce qu'on mesure

La population représente l'ensemble des individus que l'on a choisi d'étudier. Une fois qu'on a identifié cette population, on cherche à mesurer certaines caractéristiques des individus qui la composent.

Choisissons, parmi ces caractéristiques, l'âge, la taille, le salaire, l'état matrimonial, le lieu de naissance et le métier des êtres humains qui peuplent la terre. Portons plus précisément notre attention sur le Québécois et la Québécoise « moyens* ». Notre premier spécimen est un homme, âgé de 47 ans, qui a fumé plus de 100 cigarettes dans sa vie (dont la première à l'âge de 15,7 ans). Il pèse 82,7 kg et mesure 1,76 m, ce qui le classe dans la catégorie « embonpoint » de l'indice de masse corporelle (IMC). Il a accompli quelques travaux extérieurs au cours des trois derniers mois, et il possède un diplôme d'études secondaires. Notre second spécimen est une femme, âgée de 48 ans. Elle a fumé moins de 100 cigarettes dans sa vie (dont la première à l'âge de 16,6 ans). Elle pèse 65,5 kg et mesure 1,63 m. Son IMC est classé comme « normal ». Elle n'a pas accompli de travaux extérieurs au cours des trois derniers mois, et elle possède un diplôme universitaire

Pour dresser les portraits-robots de ces individus fictifs, nous nous sommes basés sur les microdonnées de l'*Enquête conjointe Canada/États-Unis sur la santé* publiée en 2004.

Une variable est une valeur susceptible de changer selon l'individu ou selon les circonstances. Dans ce sens, on peut considérer les caractéristiques d'une population comme des variables.

Toutes ces *caractéristiques* peuvent varier d'une personne à l'autre, c'est pourquoi elles sont souvent appelées des variables. Une fois que l'on a identifié les caractéristiques intéressantes, il reste à les *mesurer*, pour chaque élément de la population que l'on a décidé d'étudier. Cette notion de population ne se limite d'ailleurs pas aux seuls êtres humains. On pourrait aussi étudier la population des films (les chefs d'œuvres ou les navets), des baleines (les grises ou les bleues), ou des jours de l'année (ouvrables ou fériés).

Une variable quantitative est une variable qui prend ses valeurs dans un ensemble de nombres.

L'âge, le poids et la taille de nos spécimens québécois sont exprimés par des chiffres. On dira que ce sont des variables quantitatives. On peut faire bien des calculs avec des variables quantitatives. On pourra, par exemple, calculer le poids moyen des diplômées universitaires et le comparer à celui des autres femmes.

Une variable qualitative est une variable qui prend ses valeurs dans un ensemble de noms ou de catégories.

D'autres caractéristiques décrivant nos phénomènes sont exprimées par des mots ou des catégories : femme, diplôme universitaire, IMC normal. Ce sont des variables qualitatives. On ne peut pas faire de calcul direct sur des variables qualitatives. Personne n'a encore réussi à calculer la moyenne entre deux femmes et trois hommes? Par contre, rien ne nous empêche de compter les diplômés universitaires et les habitants de Montréal. Même lorsque les variables sont qualitatives, on n'échappe pas aux chiffres. D'un côté, on *mesure* les variables quantitatives, de l'autre, on *dénombre* les individus d'une population qui appartiennent à telle ou telle catégorie d'une variable qualitative.

On peut imaginer la variable qualitative comme étant composée de boîtes dans lesquelles on case les individus : on est marié, conjoint de fait, célibataire, divorcé ou séparé; on est un homme ou bien une femme. Ces boîtes constituent les catégories de la variable (appelées aussi *modalités*). La variable quantitative est, quant à elle, comme une ligne sur laquelle on se situe : on possède un certain nombre d'abonnés sur les réseaux sociaux, ou on a une moyenne scolaire qui se situe entre 0 et 100.

Nous disions plus haut que l'être humain se caractérise aussi par certains objets ou certains évènements. En voici, quelques exemples.

Des évènements heureux : 51 953 personnes se sont mariées au Québec en 1973 (le nombre dégringole à 25 021 en 1993, et à 21 138 en 2003). Des évènements tragiques : toujours en 1973, 2209 personnes perdent la vie sur les routes du Québec (ce chiffre descend systématiquement par la suite pour atteindre 824 en 1994 et 436 en 2013). Le lundi 14 novembre 1994, 1 957 000 téléspectateurs ont regardé *La petite vie* à Radio-Canada et le taux directeur de la Banque du Canada était de 6,04 %. Le 21 mai 2014, ils étaient 935 000 à regarder *La Poule aux œufs d'or*, alors que le taux directeur de la Banque du Canada était de 1 %. En 1993, le Japon importait 18 millions de bouteilles de cognac, devançant dans l'ordre les États-Unis et Hong Kong. En 2010, la Chine importait 22,6 millions de bouteilles de cognac, se classant derrière les États-Unis et Singapour. En 1993, le Canada comptait 4 162 000 km^2 de forêts (sur une superficie totale de 9 971 000 km^2) et les Canadiens dépensaient environ 39,9 milliards de dollars pour manger et 18,7 milliards de dollars pour assouvir certains vices (fumer et boire). En 2013, ces dépenses s'élevaient respectivement à 84,1 milliards et de 34,8 milliards, tandis que les forêts canadiennes couvraient 3 969 000 km^2.

Vous avez pu constater sans hésiter que certaines de ces variables sont quantitatives : c'est le cas du taux de la banque centrale, de la superficie des forêts ou des dépenses des Canadiens. La question est plus délicate dans le cas des mariages. On peut considérer qu'il s'agit d'une variable qualitative : le changement d'état civil (avec ses catégories : pas de changement, mariage, divorce). On peut aussi traiter la variable comme une variable quantitative (le nombre de mariages chaque année). À nous de choisir en fonction de nos besoins et de notre point de vue : l'étudiant en sciences humaines décide et les chiffres obéissent.

2.2. Classer ce qu'on mesure : les échelles

Chaque variable prend ses valeurs à l'intérieur d'une *échelle*, c'est-à-dire parmi un éventail de valeurs possibles. On peut classer ces échelles en quatre grands modèles : l'échelle *nominale* et sa variante l'échelle *ordinale* (dans lesquelles les valeurs que prend la variable sont des noms), l'échelle *de rapport* et sa sœur cadette l'échelle *d'intervalle* (dans lesquelles la variable prend des valeurs numériques). Lorsque la variable semble ne pas vouloir se conformer à une de ces quatre échelles, c'est peut-être qu'elle n'a pas été convenablement définie.

Un échantillon représente une partie de la population que l'on veut observer.

Certains films deviennent des classiques à cause de leur qualité esthétique. D'autres marquent un point tournant dans l'histoire du cinéma. *Bedtime for Bonzo* ne remplit peut-être pas ces critères, mais il fait néanmoins l'objet d'un véritable culte. Il faut dire que Ronald Reagan (qui deviendra président des États-Unis 30 ans après le tournage) y joue le rôle d'un très intellectuel professeur d'université. En compilant un certain nombre d'ouvrages de référence sur le cinéma, nous avons établi une liste d'environ 1000 grands classiques (la *population* étudiée ici) avec quelques-unes de leurs caractéristiques. Faute de place, nous ne reproduisons malheureusement qu'une partie de cette liste dans le tableau 1.1. Il s'agit d'un échantillon qui n'a rien de représentatif.

Tableau 1.1 - Les caractéristiques d'une population

Les grands classiques du cinéma (films long métrage de fiction)

	Titre	Metteur en scène	Pays	Genre	Cote	Année	Durée
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	Les Sept Samouraïs	Akira Kurosawa	Japon	Aventure	Excellent	1954	90
2	Aguirre, la colère de Dieu	Werner Herzog	Allemagne	Aventure	Très bon	1972	90
3	Danse avec les loups	Kevin Costner	États-Unis	Western	Excellent	1990	181
4	Johnny Guitare	Nicholas Ray	États-Unis	Western	Très bon	1954	111
5	Frankenstein	James Whale	États-Unis	Fantastique	Bon	1931	71
6	La Belle et la Bête	Jean Cocteau	France	Fantastique	Excellent	1946	90
7	2001: Odyssée de l'espace	Stanley Kubrick	États-Unis	Fantastique	Excellent	1968	141
8	Orange Mécanique	Stanley Kubrick	Royaume-Uni	Fantastique	Très bon	1971	137
9	Ascenseur pour l'échafaud	Louis Malle	France	Policier	Bon	1957	90
10	Psychose	Alfred Hitchcock	États-Unis	Policier	Très bon	1960	109
11	Épouse et concubines	Zhang Yimou	Chine	Drame	Très bon	1994	126
12	Jules César	Joseph Manckiewicz	États-Unis	Drame	Très bon	1953	120
13	Rashomon	Akira Kurosawa	Japon	Drame	Très bon	1951	90
14	Bedtime for Bonzo	Frederick de Cordova	États-Unis	Comédie	Médiocre	1951	83
15	Divorce à l'italienne	Pietro Germi	Italie	Comédie	Très bon	1962	104
16	...			Échelle qualitative		Échelle quantitative	
				nominale	ordinale	d'intervalle	de rapport

Sources : Chronique du Cinéma, Éditions Chronique, 1992; Danny Peary, Cult movies, Delta, 1981; (pour les cotes) Leonard Maltin, TV movies, Signet, Annuel; Pauline Kael, 5001 Nights at the movies, HRW, 1992. Note : la liste des grands classiques a été établie d'après Chronique du Cinéma (les grands thèmes du cinéma) et d'après Danny Peary (Cult movies). faute de place, seul 15 des films de la liste (sur un total de 1000 environ) figurent dans le tableau.

Une échelle représente l'ensemble des valeurs que peut prendre une variable.

Nous regarderons plus particulièrement les quatre caractéristiques (ou variables) correspondant aux quatre colonnes de droite. Les caractéristiques *genre* et *cote* prennent comme valeur des noms ou des adjectifs : elles appartiennent à des échelles qualitatives. Par extension, nous pourrions appeler ces deux caractéristiques des *variables qualitatives*, comme nous l'avons déjà fait un peu plus haut. Les caractéristiques *année* et *durée* prennent comme valeur des nombres : elles appartiennent à des échelles quantitatives (on peut donc les appeler des *variables quantitatives*).

L'échelle nominale est constituée d'un ensemble de catégories exclusives (elles ne se recoupent pas) et exhaustives (elles couvrent toutes les possibilités).

L'échelle nominale s'applique à une variable qualitative. Les valeurs que peut prendre la variable s'expriment par des noms ou des catégories. Les catégories peuvent être délimitées de différentes manières. On peut les regrouper ou les fractionner selon leur pertinence dans le projet à l'étude. Le découpage des catégories doit en tous cas respecter les deux règles suivantes : l'*exclusivité* (un élément ne peut appartenir à plusieurs catégories en même temps) et l'*exhaustivité* (les catégories doivent couvrir tous les cas possibles).

La variable *genre* (colonne 4 du tableau 1.1) appartient à une échelle nominale. Tout film doit appartenir à un genre et un seul. Dans certains cas, la classification du genre peut s'avérer délicate, c'est pourquoi il importe de bien définir les catégories au départ. Si on s'adresse à un public d'initiés, on fera peut-être la distinction entre film policier, film noir et film d'espionnage. Mais dans tous les cas, il faudra s'assurer que chaque film pourra être associé à une catégorie et à une seule.

Dans des enquêtes menées par des étudiants en méthodes quantitatives, il arrive que des individus de la population sondée se voient classés simultanément dans plusieurs catégories de la même échelle. De même qu'un film pourrait être considéré à la fois comme une comédie et un drame historique (variable *genre*), un individu se voit, par exemple, étiqueté doublement comme Canadien et Japonais (variable *nationalité*). Ce qui pose problème, c'est que les caractéristiques d'un tel individu pèseraient alors deux fois plus lourd dans les calculs qu'un individu à étiquette unique. Les résultats de l'enquête seraient alors faussés, et, par conséquent, inutilisables. C'est pourquoi il est indispensable de se plier à la règle de l'exclusivité.

L'échelle ordinale est une échelle nominale dont les catégories peuvent être classées dans un certain ordre.

L'échelle ordinale est une variante de l'échelle nominale. Comme pour l'échelle nominale, la variable prend ses valeurs dans un ensemble de catégories. La seule différence est que, dans l'échelle ordinale, les catégories peuvent être classées dans un certain ordre. Comme pour l'échelle nominale, le découpage des catégories doit être à la fois exclusif et exhaustif.

La variable *cote* (colonne 5) appartient à une échelle ordinale. Chaque film est associé à une catégorie plus ou moins flatteuse (de médiocre à excellent, en passant par bon et très bon). Je t'aime un peu, beaucoup, passionnément, à la folie… pas du tout.

L'échelle d'intervalle est constituée d'un ensemble de valeurs numériques sans point de référence absolu.

Dans l'échelle d'intervalle, on peut comparer les distances entre les valeurs que prend la variable, mais il n'existe aucun point de référence absolu. La variable *année de parution* (colonne 6) appartient à l'échelle d'intervalle. Il s'est écoulé autant de temps entre la sortie de premier et du deuxième film de la liste (1972 – 1954 = 18 ans) qu'entre la sortie du deuxième et du troisième (1990 – 1972 = 18 ans). On ne peut pas pour autant en déduire que *Danse avec les loups* est 1,018 (soit 1990/1954 = 1,018) fois plus récent que *Les Sept Samouraïs*.

L'échelle de rapport est constituée d'un ensemble de valeurs numériques avec un point de référence absolu.

Dans le tableau 1.1, seule la variable *durée* (colonne 7) appartient à une échelle de rapport. En effet, on peut dire non seulement que *2001 : Odyssée de l'espace* dure 70 minutes de plus que *Frankenstein*, mais aussi qu'il est 2 fois plus long (141/70 = 2). Si on se fie à ce petit échantillon, on remarque que les réalisateurs (ou les producteurs) ont une prédilection pour les films d'une heure et demie.

2.3. Dénombrer des individus

Observer une caractéristique qualitative est rarement une fin en soi. Pour le spectateur, qui ne regarde qu'un film à la fois, il est sans doute intéressant de savoir que le film *Les Sept Samouraïs* a été produit par le Japon et a été classé comme excellent par la critique. Le chercheur, par contre, utilisera souvent les variables qualitatives pour faire des *dénombrements*. Il constatera, par exemple, que le Japon a produit 230 films en 1991 (contre 428 aux États-Unis, 146 en France et 948 en Inde) ou que la chaîne 99 ne passe quasiment que des « navets ».

Le nombre d'individus possédant certaines caractéristiques communes (ou fréquence) constitue aussi une variable.

De la même façon, lorsqu'on cherche, par exemple, à connaître la caractéristique *état civil* des individus, c'est soit pour compter la fréquence de chaque catégorie (le nombre de célibataires, le nombre de personnes mariées, etc.), soit pour identifier les éléments de la population qui sont dignes d'être étudiés. Dans le premier cas, on fait un *dénombrement*, ou, si l'on préfère, on mesure une fréquence.

Le nombre d'individus appartenant à une catégorie particulière est aussi une *variable* (sa valeur peut varier selon les circonstances) et cette variable est *quantitative* (elle prend comme valeur un nombre). Ainsi, on pourra observer la quantité de films canadiens produits d'une année à l'autre ou d'une province à l'autre.

On peut même utiliser des caractéristiques *quantitatives* pour faire des dénombrements. Pour en revenir aux classiques du cinéma, on pourrait compter le nombre de films parus chaque année, ou chaque décennie, ou encore le nombre de films « très longs » (deux heures ou plus) ou de longueur « normale » (moins de deux heures).

En somme, il y a deux manières complémentaires d'étudier une population : d'une part on mesure un certain nombre de variables pour chaque élément observé, d'autre part on dénombre les éléments qui possèdent telle ou telle caractéristique.

Pour mieux comprendre toutes ces notions, nous vous proposons quelques exemples que nous rattacherons à chacune des quatre échelles de mesure.

2.4. Quelques échelles nominales : évidentes ou cachées

Lors du dernier recensement officiel de l'ex-Yougoslavie, on demandait aux citoyens de s'identifier à une (et une seule) *catégorie* d'appartenance ethnique. Ces catégories étaient généralement déterminées par la langue (Slovène, Croate), mais parfois par d'autres critères comme la religion (Serbe*, Musulman). Toutefois, ces catégories demeuraient valables dans la mesure où les personnes recensées les reconnaissaient clairement. D'ailleurs, on avait prévu, en cas de doute, une catégorie fourre-tout, les Yougoslaves, dans laquelle pouvaient se reconnaître le « Musulman » athée, le Croate marié avec un Slovène, et le Serbe qui se considérait avant tout comme un citoyen de la fédération. Dans le tableau 1.2, nous nous sommes limités à la Bosnie. On peut y constater qu'aucune des catégories ne l'emportait de façon nette : l'équilibre ethnique y était particulièrement fragile.

Tableau 1.2 - Les communautés en Bosnie	
(milliers de personnes, d'après le recensement de 1981)	
Musulmans	1630
Serbes	1321
Croates	758
Yougoslaves	326
Total	**4035**

Source : Atlas des peuples d'Europe centrale, La Découverte, 1991.

Les Serbes constituent un groupe ethnique, mais aussi un groupe religieux orthodoxe, sous-ensemble de la chrétienté.

Dans le tableau 1.2, les règles de l'exclusivité et de l'exhaustivité sont respectées. En Bosnie, on ne pouvait pas se déclarer en même temps Serbe *et* Croate (même si papa était Serbe et maman, Croate). Il fallait obligatoirement choisir. D'autre part, si on ne se considérait ni Musulman, ni Serbe, ni Croate, c'est qu'on était nécessairement un simple Yougoslave.

Des catégories qui s'ignorent

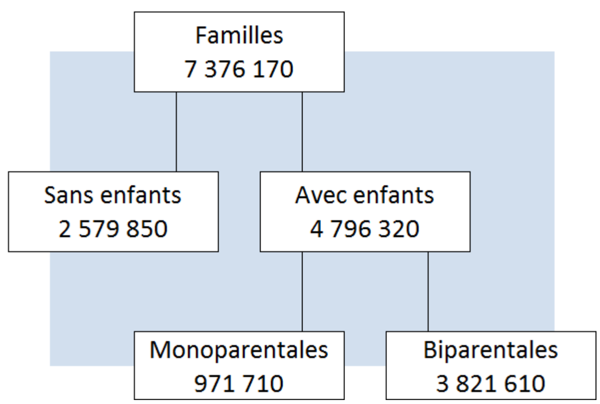
Parfois, les chercheurs, les statisticiens, ou peut-être les responsables de la mise en page des études, prennent un malin plaisir à compliquer la présentation des données pour économiser de l'espace. Observez le tableau de la figure 1.1 sans regarder le schéma qui se trouve sur sa droite (en ouvrant seulement l'œil gauche). Êtes-vous en mesure d'affirmer que les catégories sont bien exclusives et exhaustives (c'est-à-dire que toutes les possibilités sont couvertes, sans chevauchement)?

FIGURE 1.1 - Structure des familles canadiennes

Type de famille	**Nombre de familles**
Sans enfants	2 579 850
Biparentale avec enfants	3 821 610
Monoparentale avec enfants	974 710
Total	**7 376 170**

Source : Annuaire du Canada 1994, Ottawa.

Données du recensement de 1991.



Nous avons reconstitué la structure qui sous-tend le tableau. Il y a, au départ, deux sortes de familles : sans enfants et avec enfants, et, parmi ces dernières on trouve soit des familles monoparentales, soit des familles biparentales*. Les trois catégories figurant dans le tableau de la figure 1.1 correspondent aux trois cases terminales du schéma (celles dont il ne part aucune flèche). Pour y voir clair dans les chiffres, il suffit souvent d'un peu d'ordre et d'une simple opération arithmétique (ici, une addition ou une soustraction).

Apparemment, il n'existe pas encore de famille triparentale, et les enfants sans parents ne sont pas considérés comme appartenant à une famille.

2.5. Quelques échelles ordinales : authentiques et trafiquées

Vous sentez-vous en sécurité lorsque vous vous promenez tout seul dans leur quartier, après la tombée de la nuit? C'est la question que l'on a posée à quelques milliers de Canadiens âgés de 15 ans ou plus lors des Enquêtes sociales générales (voir le tableau 1.3). On reconnaît dans ce tableau une échelle ordinale tout à fait typique (très, assez, pas très, pas du tout en sécurité). Pour rendre l'étude plus intéressante, nous avons rajouté la variable nominale *sexe*.

TABLEAU 1.3 - Sentiments de sécurité éprouvés par les personnes dans leur quartier

Vous sentez vous en sécurité lorsque vous marchez seuls dans votre quartier après la tombée de la nuit?

	Hommes	Femmes	Total
	(milliers de personnes)		
	Enquête 1993		
Très en sécurité	2,4	0,8	**3,2**
Assez en sécurité	2,1	2,0	**4,1**
Pas très en sécurité	0,3	1,2	**1,5**
Pas du tout en sécurité	0,2	1,0	**1,2**
Total	**5,0**	**5,0**	**10,0**
	Enquête 2009		
Très ou plutôt en sécurité	9,5	8,5	**18,0**
Pas très ou pas du tout en sécurité	0,5	1,5	**2,0**
Total	**10,0**	**10,0**	**20,0**

Source : Juristat, janvier 1995; ESG 1993; ESG 2009, Statistique Canada.

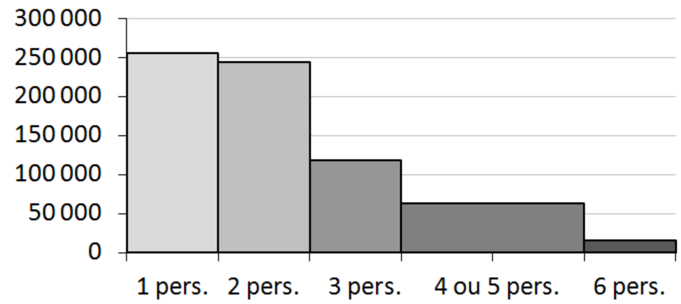
Note : Enquête auprès de 10 000 Canadiens âgés de 15 ans et plus effectuée en 1993 et 2009. Nous avons légèrement arrondi les chiffres.

Dans la figure 1.2 figurant ci-après, on reconnaît les caractéristiques de l'échelle ordinale (qui n'est après tout qu'une échelle nominale à laquelle on a ajouté une hiérarchie). Le ménage* montréalais fait nécessairement partie de l'une (et une *seule*) des cinq catégories énumérées. D'autre part, les ménages sont classés en *ordre* croissant selon le nombre de membres qu'ils comptent. Observez les deux dernières catégories. On a regroupé les ménages de 4 et 5 personnes (après tout, ça revient à peu près au même) et on a mis dans le même panier tous les ménages de 6 personnes ou plus (ces ménages sont peu nombreux et il serait inutile, pour la plupart des observateurs, d'entrer dans les détails).

Le ménage est un groupe de personnes vivant sous le même toit.

Figure 1.2 - Répartition des ménages familiaux selon le nombre de membres à Montréal en 1991

Nombre de personnes par ménage	Nombre de ménages
1 personne	255 485
2 personnes	243 320
3 personnes	118 445
4 ou 5 personnes	124 565
6 personnes ou plus	15 710
Total	757 525



Source : Statistique Canada, Cansim. Données de 1993 pour la Communauté urbaine de Montréal.

Un histogramme est un graphique où les catégories sont représentées par une série de rectangles dont la surface reflète l'importance relative des fréquences de chaque catégorie.

Nous avons représenté les données sur les ménages sous forme d'histogramme (partie droite de la figure 1.2). Chaque colonne rectangulaire de l'histogramme représente une catégorie de la variable. Les colonnes se succèdent dans l'ordre croissant des catégories : si la variable appartenait à une simple échelle nominale, nous n'aurions pas pu tracer d'histogramme.

Dans l'histogramme, la surface de chaque colonne est proportionnelle à la valeur de la catégorie correspondante. La deuxième colonne, par exemple, est deux fois plus haute que la troisième. Nous avons élargi la base de la quatrième colonne, puisque la catégorie dont il est question a une amplitude plus grande que les précédentes (deux personnes au lieu d'une seule), tout en ajustant sa hauteur en conséquence : on remarque que la troisième et la quatrième colonne ont à peu près la même surface, puisque les catégories qu'elles représentent comptent à peu près le même nombre de ménages. Enfin, pour la dernière catégorie, nous avons dû nous résigner à fixer une limite arbitraire.

L'histogramme est moins précis que le tableau dont il est issu, mais il permet d'avoir un bon coup d'œil d'ensemble sur la façon dont les données sont réparties.

Certains, plus égaux que d'autres.

Dans le tableau 1.4 figurant ci-après, comme dans le précédent, nous avons adapté une variable *quantitative* (le revenu, qui peut prendre un nombre à peu près infini de valeurs) à l'échelle ordinale (la valeur du revenu n'est alors plus un chiffre, mais une catégorie). Cette astuce permet d'obtenir un portrait efficace de la répartition (ou distribution) des revenus. Les catégories (appelées ici des *classes*) ont été découpées de façon simple : chiffres ronds, intervalles réguliers (sauf pour les deux dernières catégories) et nombre raisonnable de catégories (suffisamment, mais pas trop). Remarquez aussi qu'on a tout prévu : les familles dont le revenu se situe entre 0 et l'infini n'auront aucun mal à identifier leur catégorie. Par bonheur, on n'a pas encore inventé le revenu négatif!

Tableau 1.4 - Répartition des familles canadiennes en fonction du revenu

Tranches de revenu familial	Nombre de familles
moins de 20 000 $	944 150
20 000 $ à 39 999 $	2 043 199
40 000 $ à 59 999 $	1 895 676
60 000 $ à 74 999 $	1 032 664
75 000 $ et plus	1 460 482
Total	**7 376 170**

Source : Annuaire du Canada 1994, Ottawa.

Données du recensement de 1991.

2.6. Une échelle d'intervalle : les grands voyageurs

Dans les échelles qui suivent (échelles d'intervalle et de rapport), les variables prennent des valeurs numériques. Par exemple, l'année de la mort d'Étienne I^{er}, fondateur de l'Église serbe, a pour valeur *1228* (qui est un chiffre appartenant à l'échelle d'intervalle) et la population de baleines bleues s'élevait à 200 000 (autre chiffre, appartenant à l'échelle de rapport) au début du XX^{e} siècle*.

Cette population avait baissé à 2000 en 1994.

Grâce au tableau 1.5, on peut affirmer, par exemple, qu'il s'est écoulé 492 ans (environ) entre la visite de Leif Ericson en Amérique (en l'an 1000) et celle de Christophe Colomb (en 1492), mais qu'il a suffi ensuite d'un délai beaucoup plus court pour que Jacques Cartier se pointe lui aussi dans le coin (42 ans après Colomb). Avec de tels chiffres, on peut faire des comparaisons : tout le monde sait que 492 ans, c'est plus long que 42 ans (c'est même, à peu près, 12 fois plus long).

Tableau 1.5 - Les aventuriers sans moteur

Année	Voyage
-400	Xénophon atteint la Mer Noire avec ses compagnons vaincus à la bataille de Cunaxa (près de la Bagdad actuelle).
-218	Parti d'Espagne, Hannibal traverse les Pyrénées et les Alpes avec son armée et ses éléphants.
1000	Leif Ericson quitte son Groenland natal avec ses Vikings pour explorer les côtes de l'actuelle Nouvelle-Angleterre.
1275	Marco Polo, accompagné de son père et de son oncle, arrive à Pékin en provenance de Venise.
1353	Ibn Battuta revient au Maroc après un périple de 120 000 km (en 28 ans) qui l'a conduit jusqu'en Inde.
1492	Christophe Colomb découvre l'Amérique (oui, d'accord, elle existait déjà mais les autres ne le savaient pas).
1521	Magellan et son équipage atteignent les Philippines en provenance de l'Espagne via le Pacifique.
1534	Jacques Cartier découvre et explore le Canada.
1877	Stanley atteint l'Atlantique après avoir traversé l'Afrique d'est en ouest, à pied et en pirogue.
1935	Mao Zedong arrive au Shaanxi avec les restants de l'Armée Rouge, après 12 000 km de marche à travers la Chine.
2000	George W. Bush est élu président des États-Unis.

Sources : Xénophon, Polybe, l'auteur anonyme de la saga d'Éric le Rouge, Marco Polo, Ibn Battuta, Christophe Colomb, Antonio Pigafetta, Jacques Cartier, Henry Morton Stanley, Mao Zedong, Maison Blanche.

Par contre, on ne peut pas dire que George W. Bush a mis deux fois plus de temps à accéder à la présidence des États-Unis que Leif Ericson à toucher aux rivages de l'Amérique. En effet, l'année 0 (qui n'a d'ailleurs jamais existé) n'est pas un point de départ absolu, mais une convention arbitraire (comme le degré zéro des thermomètres, qui est placé à des endroits différents selon les pays).

Les valeurs indiquées dans la colonne de gauche du tableau 1.5 appartiennent donc à une échelle d'intervalle. C'est-à-dire qu'on peut calculer le nombre d'années qui se sont écoulées entre deux évènements, mais qu'il est impossible de diviser une valeur par une autre dans le but de comparer leur grandeur. Il en va de même pour la mesure de la température : non, il ne fait pas deux fois plus chaud à 20 qu'à 10 degrés!

Plusieurs aventuriers mentionnés dans le tableau 1.5 ont laissé des récits de leur voyage dans l'inconnu, à une époque où l'automobile et le Club Med n'étaient pas encore inventés. Vous prendrez grand plaisir à lire ces récits : ils illustrent une facette surprenante* et passionnante de l'être humain.

Xénophon, *L'Anabase*; Marco Polo, *Le devisement du monde*; Ibn Battuta, *Voyages*; Pigafetta (compagnon de Magellan), *Navigation et découvrement de l'Inde supérieure*; Henry Morton Stanley, *Comment j'ai retrouvé Livingstone*, etc.

2.7. Une échelle de rapport : les dernières baleines

Nous parlions un peu plus haut de la baleine bleue : la revoici en compagnie de quelques collègues dans le tableau 1.6. La variable *population de baleines* appartient bien à une échelle de rapport. D'une part elle prend des valeurs chiffrées (en fonction de l'espèce ou de l'époque considérée). D'autre part, on peut faire un rapport entre deux valeurs de la variable, car il existe un point de référence absolu (la valeur 0, que, espérons-le, nous n'atteindrons jamais pour ce qui est des baleines bleues). Nous pouvons dire, par exemple, que les baleines bleues étaient 100 fois plus nombreuses au début du XXe siècle qu'au début du siècle suivant.

Tableau 1.6 - Les dernières baleines

Population des baleines (en milliers)

	Début du XXe siècle	**Début du XXIe siècle**
Baleine grise du Pacifique Nord-est	10	26
Baleine franche	120	6
Rorqual boréal	200	11
Baleine bleue	200	2
Rorqual commun	470	110
Baleine des Basques	200	3
Rorqual à bosse	125	65
Baleine grise de l'Atlantique	0	0

Sources : Courrier international, 10 novembre 1994. Commission baleinière internationale.

À titre de consolation, notons que la situation est moins tragique en ce qui concerne le rorqual commun. Il en reste 110 000 contre 470 000 au début du XXe siècle. Là encore, on peut comparer l'ordre de grandeur des deux chiffres.

Les plus observateurs d'entre vous sont peut-être perplexes devant ce tableau : sa légende indique qu'il s'agit d'une variable quantitative (la population de baleines). Et pourtant, on y trouve des catégories de baleines (variable appartenant à une échelle nominale) et des années (variable appartenant à une échelle d'intervalle). Vous avez deviné juste, ce tableau illustre le rapport qui existe entre trois variables : 1) la population de baleines, 2) l'espèce de baleine et 3) le temps. On reverra très souvent ce type de croisement. Si on veut mettre le doigt sur les différentes variables, on se demande, pour chacune d'entre elles, « qu'est-ce qui varie? »

EXERCICES 2

1. Des échelles connues

Pour chacun des exemples qui suivent, (1) identifiez la variable, (2) précisez s'il s'agit d'une variable qualitative ou quantitative, (3) dans ce dernier cas, énumérez les catégories, (4) indiquez le type d'échelle utilisé.

a) Le grade des membres d'un club de karaté.

b) La quantité d'eau absorbée par un antialcoolique.

c) L'heure d'ouverture des discothèques.

d) Les programmes offerts par un collège.

2. À vos échelles!

a) Selon vous, les deux variables suivantes appartiennent-elles à la même échelle de mesure : la hauteur d'un immeuble, l'altitude d'une montagne?

b) Donnez un exemple détaillé et original pour chacune des quatre échelles de mesure.

3. Histo-quoi?

Tracez un histogramme à partir des données du tableau 1.4.

3. UNITÉS DE MESURE

Toutes les variables que nous avons énumérées jusqu'ici contenaient implicitement une unité de mesure. Pour les variables qualitatives, le problème était relativement simple : on faisait le décompte de chaque catégorie. Il y avait par exemple 255 485 ménages constitués d'une personne à Montréal et 2 579 850 familles sans enfants au Canada. Parfois, pour simplifier la présentation, les décomptes étaient exprimés en milliers (1321 milliers de Serbes en Bosnie), ou en millions.

3.1. Diversité des unités de mesure

Pour les variables quantitatives, les unités de mesure peuvent être très diverses : la date (en *années*) du voyage de Marco Polo, le revenu (en *dollars*) de l'Américain le plus riche, le poids des baleines bleues (en *tonnes*) et des nouveau-nés (en *grammes*), la production de pétrole à Terre-Neuve (en *barils*).

Comment doit-on choisir l'unité de mesure que l'on utilise?

Le choix de l'unité de mesure dépend souvent de l'objectif recherché. Le producteur de pommes ne veut pas compter le nombre de pommes de son verger, mais il se contente de savoir le poids de la récolte par catégories (Lobo, McIntosh, Golden, etc.). Le camionneur veut connaître le volume (12 m^3 de Lobo et 6 m^3 de McIntosh = 18 m^3 de conteneur), peu importe la catégorie. Le douanier se préoccupe plutôt de la valeur de la cargaison (5000 $). Le pique-niqueur s'intéresse au nombre (8 pommes pour 8 personnes : une chacun) tandis que l'excursionniste s'intéresse au nombre et au poids. Quant à la personne qui suit un régime, elle mesurera plutôt les fruits en calories : une pomme + une orange = 85 calories. Cette dernière addition vous semble peu orthodoxe? Nous y reviendrons très bientôt.

Avant d'analyser des chiffres, il est essentiel de bien identifier les unités de mesure utilisées. Il en va de même lorsque l'on veut faire des calculs à partir de ces chiffres. Il faut alors veiller à convertir dans la même unité, si nécessaire, les valeurs calculées en fonction d'unités différentes. Si le Québec (et ses 7,4 millions d'habitants en 1995) fusionnait avec Terre-Neuve (et ses 573,6 milliers d'habitants) pour former une république, quelle serait la population du nouvel État? La conversion elle-même ne présente pas de difficulté : ce qui compte, c'est de ne pas l'oublier lorsqu'elle est nécessaire. Notons que plus l'unité de mesure est grande (au Québec, on compte en *millions* et à Terre-Neuve en *milliers*), plus le chiffre est petit.

3.2. On n'additionne que des éléments homogènes

Comment préparer une salade de fruits sans faire de la compote.

S'il est facile d'additionner des Québécois et des Terre-Neuviens, il est par contre impossible d'en faire autant avec des pommes et des oranges. C'est du moins ce que nous avons tous appris à la petite école. Il s'agissait alors d'un dogme irréfutable. Le moment est venu, chers lecteurs, de remplacer cet acte de foi en raisonnement un peu plus subtil.

Si le tiroir de mon réfrigérateur peut contenir 20 fruits de calibre moyen, peu importe pour moi que je doive y placer 15 pommes + 10 oranges ou 10 pommes + 15 oranges. Ce qui est certain, c'est qu'il restera 5 fruits qui n'entreront pas dans le tiroir. Si je me préoccupe du stockage des fruits, j'ai tout à fait le droit d'additionner des pommes et des oranges. Par contre, s'il s'agit de préparer un canard à l'orange suivi d'une tarte aux pommes, il n'est plus question de mettre les deux sortes de fruits dans le même chaudron.

Des éléments sont homogènes s'ils sont exprimés dans la même unité de mesure.

Les éléments ne peuvent être additionnés que lorsqu'ils sont exprimés dans la même unité de mesure. On dit alors que ces éléments sont homogènes. Comme nous venons de le voir, l'homogénéité dépend parfois du point de vue. Lorsque le ministère de l'Industrie annonce qu'il y a eu 273 203 mises en chantiers au Canada durant l'année 1976 (année record), il est clair qu'on a additionné des bungalows à des appartements et à des maisons en rangées. Il en va de même lorsqu'on indique qu'au Canada 2,7 millions de personnes étaient des travailleurs autonomes indépendants en 2014, dont 1,7 million d'hommes. Dans ce cas, on peut en effet considérer que 3 menuisiers + 2 écrivains sont bien égaux à… 5 travailleurs autonomes. Mais ce genre de gymnastique a ses limites.

3.3. Une unité de mesure commode, mais capricieuse : la valeur monétaire

Quand on veut évaluer un ensemble d'éléments disparates, on a souvent recours à une unité de mesure partagée par beaucoup de variables : la valeur monétaire. Grâce à la valeur monétaire, on peut additionner 2 douzaines d'œufs à 10 tranches de fromage et une caisse de bière : en tout, ça peut valoir 50 $, par exemple. On utilisera la monnaie pour mesurer les revenus des individus, la production d'un pays, la dette extérieure, les taxes municipales, le coût de construction d'une maison, la valeur d'une récolte, d'un troupeau ou d'une terre, etc.

Cependant, si la monnaie est bien commode (elle rend *homogènes* les divers éléments), elle diffère de toutes les autres unités de mesure sur le point suivant : sa valeur peut varier n'importe quand. Dieu merci, le mètre et le gramme, deux unités de mesure bien établies, n'ont pas changé depuis leur création. Par contre, dès que les prix montent, la monnaie perd de la valeur. Alors, ne vous laissez pas impressionner par votre grand-père qui prétend s'être débrouillé avec 100 $ par mois dans sa jeunesse. Nous verrons plus tard comment faire les ajustements nécessaires lorsque la valeur de la monnaie change. Pour l'instant, notons seulement que l'unité monétaire n'a de valeur que par rapport à une date et un lieu donné.

3.4. Un degré de précision adapté

Youpi! J'ai perdu 3,2 grammes!

Avez-vous constaté à quel point les enfants sont précis sur leur âge (« j'ai 5 ans et 3/4 ») alors que les adultes sont plutôt évasifs (« je suis dans la quarantaine »). Certains parents vous communiqueront avec fierté l'âge de leur nourrisson en mois et même en jours (mais jamais en heures). Et lors d'une naissance, la première chose qu'on vous annonce, c'est le poids du bébé, à une once ou dix grammes près. Le niveau de précision dépend, encore une fois, du point de vue.

Si on veut s'assurer que le petit bébé a bien absorbé son biberon dans la pouponnière de l'hôpital, un degré de précision de 10 ou 20 grammes dans son poids sera intéressant. Dans le cas contraire, la précision n'a aucune valeur en tant que chiffre. Elle sert seulement à avoir l'air scientifique ou à impressionner. D'ailleurs, si le poids du bébé semble jouer un rôle de prestige dans notre société (« plus le bébé est gros, plus on est fier »), il faut bien se rendre compte qu'un bébé de 3,520 kg n'est peut-être pas plus gros que son collègue de 3,500 kg : sur quelle balance le bébé a-t-il été pesé, quelle était la pression atmosphérique, le bébé avait-il les cheveux mouillés, etc.?

En ce qui nous concerne, étant donné que nous voulons traiter les chiffres de façon *objective* et non comme des instruments magiques ou des boîtes de poudre aux yeux, il nous faut adapter le degré de précision à l'objectif de l'étude. Les nombreux exemples tirés de la réalité que nous donnons dans cet ouvrage illustrent, nous l'espérons, ce principe de simple bon sens.

3.5. Une « société distincte »

Nous l'avons vu un peu plus haut, les Québécois constituent un des rares peuples à utiliser simultanément deux systèmes de poids et mesures. Voilà un excellent sujet d'enquête pour un travail de fin de session, voire pour une thèse de doctorat en anthropologie. Pour le moment, nous nous contenterons d'explorer brièvement ce thème afin de récapituler les principales notions vues jusqu'ici dans ce chapitre.

Au Québec, nous l'avons dit, la température de l'air se mesure souvent en degrés Celsius alors que celle des liquides dans lesquels on trempe les êtres vivants (baigneurs ou poissons rouges) se mesure plutôt en degrés Fahrenheit. L'échelle Fahrenheit est par ailleurs largement utilisée pour la température de la dinde rôtie et des malades. Le poids des êtres humains se mesure en livres (et en onces pour les bébés), alors que la capacité maximum des ascenseurs est indiquée en kilogrammes. La taille de ces mêmes humains se mesure en pieds et en pouces, alors que les distances sur les routes se mesurent en kilomètres. L'essence, le vin et l'eau se mesurent en litres, ou en millilitres, alors que le « gros gin » se mesure en onces.

L'enquête en question consisterait à interroger ou observer un échantillon de la population du Québec et de relever, pour chaque individu, une série de caractéristiques. On s'intéresserait notamment aux différents systèmes de mesure utilisés par les individus selon les circonstances de la vie courante, ainsi qu'à d'autres *caractéristiques* plus générales, telles que le sexe, la langue maternelle (échelle nominale); l'âge, le niveau de scolarité, ou le revenu de ces mêmes individus (échelle de rapport). On pourrait même examiner leur capacité (échelle ordinale) à convertir des onces en livres, des pouces en pieds et des pieds en milles, ou à orthographier correctement le mot « Fahrenheit »!

Ces caractéristiques, qui constituent nos fameuses « variables », seraient par la suite comptabilisées ou transformées. On pourra, par exemple, évaluer la proportion d'individus qui usent (et abusent) des kilogrammes, ou la moyenne d'âge des individus allergiques au système métrique. On pourra également étudier les différentes relations entre ces variables, ainsi que l'intensité de ces relations. Tous ces sujets font justement l'objet de ce manuel et seront abordés progressivement dans les chapitres à venir.

EXERCICES 3

1. Conversions massives

Répondez aux questions suivantes en vous aidant du tableau 1.7.

a) Convertissez la production québécoise de quadrupèdes (mentionnés dans le tableau) en millions d'unités.

b) Convertissez la production de lait du Québec en millions de litres.

c) Que pensez-vous des unités de mesure choisies dans le tableau?

Tableau 1.7 - Production agricole d'origine animale dans certaines provinces canadiennes

Produit	Bovins	Porcs	Moutons	Volaille	Œufs	Lait	Êtres humains
	(milliers)	(milliers)	(milliers)	(milliers de tonnes)	(millions de douzaines)	(milliers de kilolitres)	(milliers)
Québec	1368	2994	124	176,6	82,7	2793	6813
Ontario	2210	3112	209	207,1	179,2	2463	9846
Alberta	4403	1702	247	50,6	37,3	581	2502

Source : Annuaire du Canada 1994, Ottawa. Données de 1991.

2. Soyons précis… mais pas trop.

Pour chacune des situations suivantes, indiquez l'unité de mesure à utiliser et dites quel degré de précision vous parait souhaitable.

a) L'âge des bébés inscrits à la pouponnière de l'hôpital;

b) L'âge des participants à un marathon;

c) La température d'une salle d'urgence;

d) La température d'un malade;

e) Le nombre de bénévoles pour un souper de fèves au lard;

f) Le nombre de Québécois qui visitent la France chaque année;

g) Le nombre de jours fériés dans l'année;

h) La durée des vacances annuelles;

i) La durée de la semaine de travail;

j) Le temps de cuisson d'un œuf à la coque;

k) Le prix d'une maison, d'une voiture, d'un kilo de viande hachée.

4. PRÉSENTER LES DONNÉES : UNE IMAGE VAUT MILLE CHIFFRES

Les chiffres bruts sont souvent présentés sous forme de tableaux ou de graphiques. Encore une fois, la façon de présenter les données dépend du but recherché. Dans les paragraphes qui suivent, nous vous proposons quelques situations bien réelles que des tableaux ou des graphiques permettent d'illustrer de façon efficace.

4.1. Les jours fériés

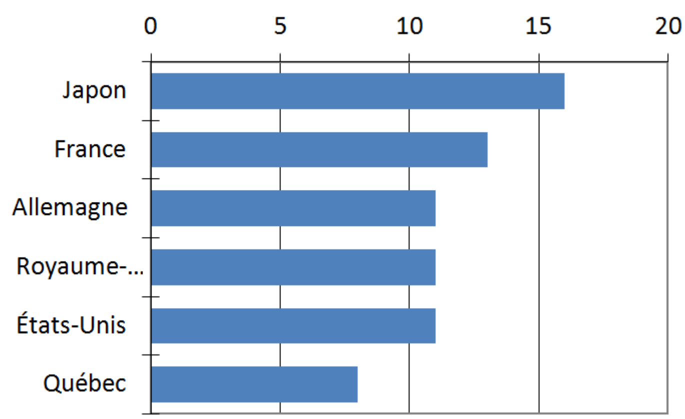
Cher Océan, c'est à ton tour…

D'après vous, qui, des Japonais, des Américains, des Français et des Québécois bénéficient du plus grand nombre de jours fériés (figure 1.3)? N'allez pas plus loin : avant de lire la réponse, faites travailler vos préjugés! Eh bien, les Japonais, qui sont réputés être des bourreaux de travail viennent d'obtenir, en 1996, un seizième jour de congé officiel (la Fête de l'Océan). Ils sont maintenant deux fois plus gâtés que les Québécois.

Figure 1.3 - Jours fériés officiels

Nombre de jours fériés par année	
Japon	16
France	13
Allemagne	11
Royaume-Uni	11
États-Unis	11
Québec	8

Sources : *Holland Herald*, février 1995 et *Code du travail du Québec*, 1995.



Nous serions donc les travailleurs les plus zélés au monde?

Quelques remarques s'imposent toutefois sur ces données. Le tableau ne constitue pas un classement exhaustif, car il ne s'agit pas, par exemple, des 6 pays du monde qui offrent le plus de jours fériés à leurs travailleurs. Le choix s'est limité de façon délibérée aux 5 plus gros pays industrialisés et au Québec. Ainsi, la France est la deuxième du groupe des cinq et non la deuxième au monde. D'autre part, le tableau ne dit pas toute la vérité sur les congés, car il ne tient pas compte des vacances annuelles des employés. Selon le *Holland Herald*, les Japonais prennent 3 semaines de vacances par an. C'est moins que les Français et les Allemands (6 semaines), mais c'est déjà plus que les Américains (2 semaines). Si on considère l'évolution à plus long terme, on peut même noter un renversement de tendance depuis 1975, puisque seuls les Américains prennent de moins en moins de vacances. On estime même que ces derniers ont sacrifié en 20 ans l'équivalent de 4 semaines de vacances (jours de congé en moins et heures supplémentaires en plus) dans le but de maintenir leur niveau de vie.

Dans un diagramme en bâton, les différentes valeurs que peut prendre une variable sont représentées par des bâtons plus ou moins grands.

Le graphique présenté à la figure 1.3 est un diagramme en bâtons. Les bâtons sont horizontaux, parce qu'on voulait écrire le nom des pays lisiblement, mais le graphique aurait très bien pu être construit dans l'autre sens sans violer une règle sacrée. On y retrouve deux variables : le nombre de jours fériés et le pays. Vous avez remarqué que la variable *pays* est qualitative et qu'on retrouve à la base des bâtons les catégories, c'est-à-dire les différentes valeurs que peut prendre cette variable. La variable *nombre de jours fériés* est une variable quantitative qui appartient à une échelle de rapport et les valeurs qu'elle prend se reflètent dans la longueur des bâtons.

4.2. Les députés de la mafia

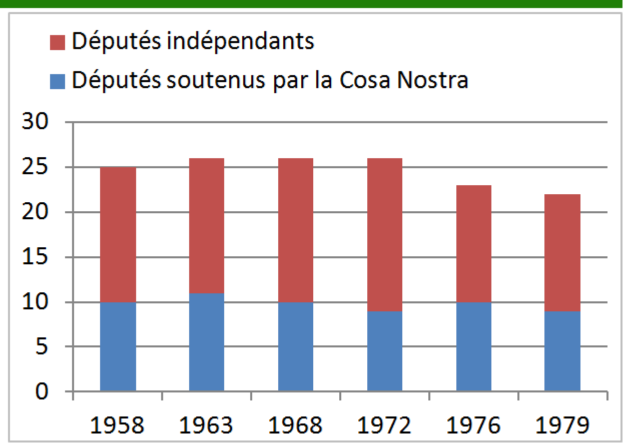
Entre 1958 et 1979, un député sicilien sur trois est élu grâce à l'appui de la Mafia. C'est ce que prétend le très sérieux hebdomadaire italien *L'Espresso* du 3 mars 1995. L'article est publié au moment où Giulio Andreotti, sept fois président du conseil, est mis en accusation. L'Italie est en pleine opération « mains propres » : de nombreux dirigeants sont accusés de corruption, et quelques juges courageux tombent sous les balles de la Cosa Nostra (nom de la Mafia sicilienne). Les procès donnent lieu à des dénonciations et plusieurs députés sont identifiés comme « hommes d'honneur » de la Cosa Nostra.

Le tableau accompagnant la figure 1.4, publié par l'hebdomadaire *L'Espresso*, illustre la présence massive de la mafia dans la députation (le diagramme en bâtons que nous avons tracé à côté du tableau reprend les mêmes données). Les chiffres concernent la Sicile occidentale, c'est-à-dire la partie de l'île où la Cosa Nostra est la plus présente (région de Palerme).

Figure 1.4 - Les députés mafieux en Sicile

Année d'élection	Députés soutenus par la Cosa Nostra	Total des députés
1958	10	25
1963	11	26
1968	10	26
1972	9	26
1976	10	23
1979	9	22

Source : *L'Espresso* , 3 mars 1995.
Les données concernent la Sicile occidentale.



Les liens entre la mafia et la politique touchent particulièrement le parti de monsieur Andreotti (la Démocratie chrétienne). La plupart des mafieux se retrouvent dans ce parti et une bonne proportion des démocrates chrétiens siciliens sont des mafieux. La Sicile est une région très peuplée (un Italien sur douze y vit) et la Démocratie chrétienne tenait sans doute à y conserver son emprise, quitte à pactiser avec le diable.

Qu'est-ce qui se cache derrière les chiffres?

Derrière les quantités se cachent des éléments de nature qualitative. Il faut souligner que les députés de la Cosa Nostra se réservaient autant que possible les portefeuilles clés du gouvernement italien : ministère de l'Intérieur (police), de la Défense (trafic d'armes), du Commerce extérieur (contrebande), de la Marine marchande (re-contrebande) et de l'Agriculture (subventions aux riches vergers siciliens). Les particularités du système électoral italien (à la proportionnelle avec report possible de votes sur certains candidats) permettaient d'autre part aux mafieux d'écarter de la course des candidats jugés trop hostiles à leur cause.

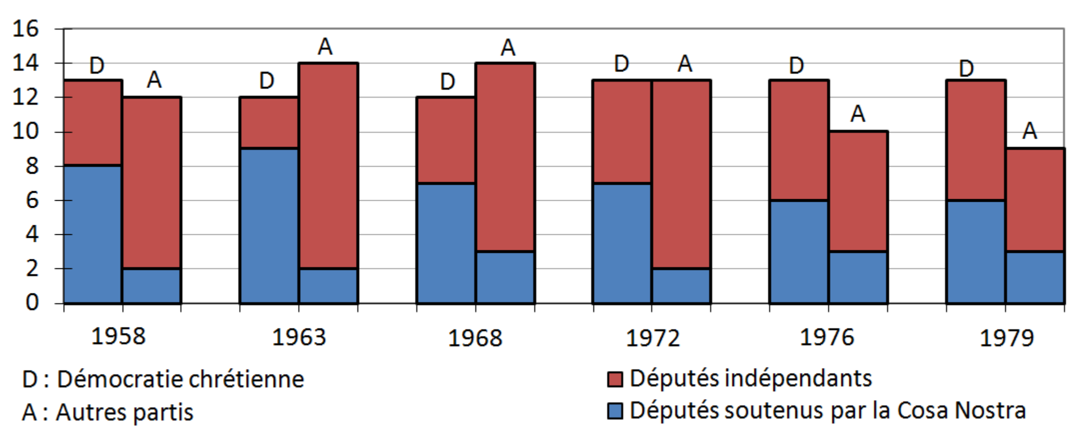
La figure 1.4 présente l'évolution, à travers le temps, du nombre de députés soutenus par la mafia, et du nombre de députés qui ne le sont pas. Ces deux variables appartiennent à une échelle de rapport, même si, au départ, elles ont été construites à partir d'une échelle nominale (le fait pour un député d'être affilié ou non à la mafia). Les bâtons du diagramme sont divisés en sections, ce qui permet d'illustrer simultanément les deux variables en présence. Les bâtons sont orientés verticalement, ce qui rend les comparaisons plus commodes. Le diagramme est d'ailleurs assez clair pour que l'on puisse se passer du tableau. À titre de comparaison, le tableau des jours fériés était, par contre, indispensable : dans un cas comme celui-là, les lecteurs sont curieux de connaître le nombre *exact* de jours de congés fériés (*revoir la figure 1.3*).

La figure 1.5 ajoute une nouvelle variable au tableau précédent. En plus de l'année d'élection et du lien avec la mafia des députés, on tient compte de leur affiliation aux partis politiques. On vise ainsi trois buts : le lien avec la mafia montre l'influence de cette dernière sur la vie politique, l'année d'élection montre la constance de cette influence et l'affiliation politique montre la relation privilégiée entre le parti au pouvoir (la Démocratie chrétienne) et la mafia. Le diagramme en bâton fait encore l'affaire, mais il commence à être encombré. Si la situation se compliquait (si, par exemple, on tenait compte de plus de deux partis, ou de plus de deux types de liens avec la mafia), il faudrait peut-être tracer plusieurs figures séparées.

Figure 1.5 - Les députés mafieux en Sicile selon l'affiliation politique

	DÉMOCRATES CHRÉTIENS		AUTRES PARTIS	
	Députés soutenus par la Cosa Nostra	Total des députés	Députés soutenus par la Cosa Nostra	Total des députés
1958	8	13	2	12
1963	9	12	2	14
1968	7	12	3	14
1972	7	13	2	13
1976	6	13	3	10
1979	6	13	3	9

Source : L'Espresso , 3 mars 1995. Les données concernent la Sicile occidentale.



4.3. La planète du vin

La France et l'Italie se partage généralement la place de premier producteur de vin. En 1993, la France assurait encore près du quart de la production mondiale. On constate par ailleurs que la vigne se répand progressivement dans toutes les zones tempérées de la planète. Il y a du vin partout… mais toujours un seul Château-Margot. En 2013, la production mondiale de vin était de 281 millions d'hectolitres (soit sensiblement le même niveau que 20 ans plus tôt), et la consommation s'élevait à 245 millions d'hectolitres. L'écart entre production et consommation s'explique par les usages industriels du vin (fabrication de brandy, vermouth, de vinaigre, etc.). Étant donné qu'un hectolitre contient 100 litres, et que les bouteilles de vin (piquette mise à part) contiennent généralement 0,75 litre (ou 750 millilitres), on peut facilement estimer la consommation mondiale de 2013 à environ 32,9 milliards de bouteilles (pour 7,140 milliards d'habitants sur la terre).

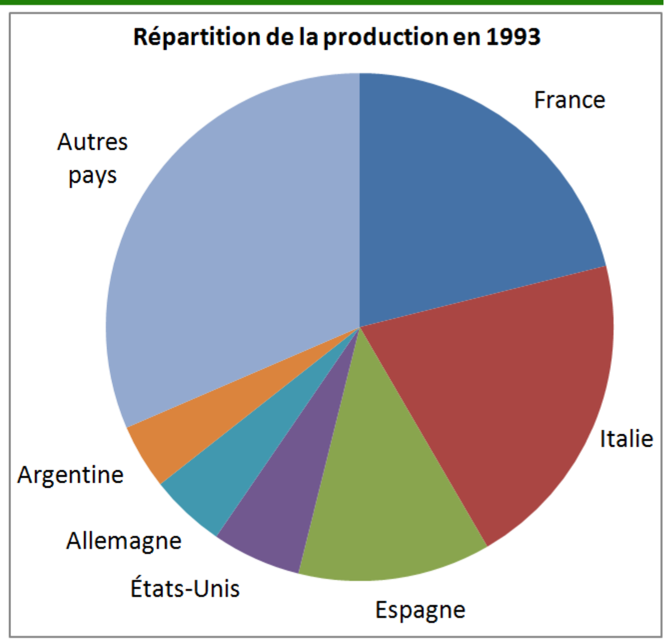
Notez les ressemblances et les différences entre la figure 1.6 et les trois figures précédentes. Quelle sorte de variable y retrouve-t-on? Pourquoi avoir choisi un type de diagramme différent pour l'illustrer?

Comme le tableau précédent (les députés mafieux), la figure 1.6, figurant ci-après, contient une variable de type quantitatif à échelle de rapport : le volume de vin produit. Cette fois, la variable peut prendre des valeurs continues (et pas seulement des valeurs entières). On observe la relation entre la quantité de vin produite et le lieu de production (espace), et ce à travers les années (temps).

Figure 1.6 - Principaux producteurs de vin

(en millions d'hectolitres)

	1993	2013
France	58,7	44,1
Italie	57,0	44,9
Espagne	34,1	40,0
États-Unis	15,8	22,0
Allemagne	13,4	9,0
Argentine	11,5	15,0
Afrique du Sud	9,3	11,0
Portugal	9,0	6,7
Roumanie	8,0	5,9
Grèce	5,0	3,7
Australie	4,7	13,5
Hongrie	4,4	2,6
Brésil	3,7	2,7
Chine	3,5	..
Chili	3,2	12,8
Autres pays	36,7	47,1
Total mondial	**278,0**	**281,0**



Sources : Courrier international du 10 novembre 1994; Organisation internationale de la vigne et du vin (OIV) 2014.

Le diagramme circulaire est découpé en tranches dont la grosseur reflète la valeur de chaque catégorie d'une variable.

La particularité de ce tableau (par rapport aux trois tableaux précédents) vient du fait que chaque valeur représente la partie d'un tout. Le diagramme circulaire est tout à fait approprié pour ce type d'information, puisqu'il permet à la fois de comparer les pays entre eux et de les comparer au total. Pour ne pas encombrer le diagramme, il a fallu cependant simplifier les données et regrouper les pays les moins significatifs. Si le diagramme circulaire offre un meilleur coup d'œil, le tableau fournit une plus grande précision.

EXERCICES 4

1. Faites parler les données

Tracez un graphe pour représenter de façon « parlante » les données des tableaux suivants :

a) Tableau 1.2.

b) Tableau 1.3.

c) Tableau 1.6.

2. Recherche

Trouvez des données qui se prêteraient bien aux représentations graphiques suivantes :

a) Diagramme circulaire.

b) Diagramme en bâton.

c) Histogramme.

EXERCICES SUPPLÉMENTAIRES

1. Au pays des centenaires

a) Identifiez les variables concernant la population japonaise qui ont servi de base à la création du tableau 1.8.

b) Dites à quelle échelle de mesure appartiennent ces variables.

c) Quelles sont les précautions à prendre avant de comparer les lignes du tableau?

d) Tracez le graphique qui, selon vous, illustrerait le mieux ce tableau.

Tableau 1.8 - La pyramide des âges au Japon le 1er janvier 2014

		Hommes	Femmes	Total
			(en milliers)	
[1]	0 à 9 ans	5 410	5 150	**10 560**
[2]	10 à 19 ans	6 030	5 740	**11 770**
[3]	20 à 39 ans	14 900	14 360	**29 260**
{4]	40 à 59 ans	16 900	16 780	**33 680**
[5]	60 à 79 ans	15 250	17 020	**32 270**
[6]	80 à 99 ans	3 320	6 190	**9 510**
[7]	100 ans et plus	10	50	**60**
	Total	**61 820**	**65 290**	**127 110**

Source : Office statistique du Japon (総務省統計局).

2. Échelle de valeurs

a) Identifiez quelques variables utilisées dans le tableau 1.9.

Tableau 1.9 - Nombre annuel d'infractions au Canada selon le type d'infraction

Types d'infraction	1998	2005	2012
Total des infractions avec violence prévues au Code criminel	**405 567**	**447 857**	**415 119**
Homicide	558	663	543
Agression sexuelle grave	219	176	130
Voies de fait graves	2 625	3 095	3 514
Autres infractions avec violence	402 165	443 923	410 932
Total des crimes contre la propriété	**1 717 754**	**1 574 808**	**1 190 972**
Total des affaires d'introduction par effraction	350 774	261 362	175 712
Total des vols de véhicules à moteur	165 920	160 014	77 939
Total des vols de moins de 5 000 $ (sauf les véhicules à moteur)	713 632	638 684	496 781
Autres crimes contre la propriété	487 428	514 748	440 540

Source : Statistique Canada, Cansim 2502-0051.

b) Pour chaque variable, précisez si elle est qualitative ou quantitative, et à quelle échelle elle appartient.

c) Si une des variables appartient à l'échelle nominale, vérifiez qu'elle en possède bien les deux caractéristiques. Dites pourquoi cette échelle ne peut pas être qualifiée « d'échelle ordinale ».

3. Quelques arpents de neige

a) Complétez le tableau 1.10. Si vous avez de la difficulté, essayez d'abord la question b.

Tableau 1.10 - Les terres forestières au Canada

Catégorie	(millions d'hectares)	**Catégories cumulées**	(millions d'hectares)
Terres forestières boisées	219,6	Terres forestières boisées et productives autre que les réserves	219,6
Terres forestières non boisées	17,1	Terres forestières productives autres que les réserves	236,7
Terres forestières productives en réserves	_____	Terres forestières productives	245,4
Terres forestières non productives	170,8	Terres forestières	416,2
Terres non forestières	505,3	Superficie des terres	_____
Eau	75,6	Superficie totale du Canada	997,1

Source : Comptes nationauc des revenus et dépenses, 1982-1993, Statistique Canada.

Note : 1 hectare (ha) = 100 m × 100 m = 10 000 m²; 100 ha = 1 km².

b) Présentez les catégories et leurs valeurs sous forme de structure arborescente (voir la figure 1.1 sur la structure des familles canadiennes).

c) Convertissez la superficie du Canada en km² et vérifiez l'adage « multiplier l'unité c'est diviser la mesure et réciproquement ».

4. Recherche : la Deuxième Guerre mondiale

a) Identifiez 4 pays ayant participé à la Deuxième Guerre mondiale et pour chacun d'entre eux trouvez : le camp auquel ils appartenaient au début de 1943, leur date d'entrée en guerre, le nombre de soldats mobilisés et une autre variable de votre choix.

b) Identifiez l'échelle de mesure de chacune des variables.

c) Représentez une des variables sous forme de graphique.

5. Pourquoi parlent-ils notre langue?

Avant la chute du bloc de l'Est, le français était, en dehors du russe, la principale langue étrangère parlée en Roumanie, en Moldavie et en Bulgarie.

a) En Moldavie, parmi les 259 300 élèves des écoles secondaires de langue moldave qui étudiaient une langue étrangère en 1992, 219 181 avaient choisi le français (contre 24 543 l'anglais, 4185 l'allemand et 5175 l'espagnol). Vérifiez que *la langue étudiée* est bien une variable nominale.

b) À l'aide d'un chiffrier électronique, tracez un diagramme circulaire représentant la distribution des quatre catégories de la variable mentionnée dans la question précédente (le français en Moldavie).

c) Aux États-Unis, en 1990, le français est la deuxième langue étrangère (après l'espagnol) étudiée dans les écoles. Au niveau primaire, le français a une clientèle 2 fois plus élevée que toutes les autres langues réunies (en dehors de l'espagnol). Le même phénomène se retrouve dans l'enseignement secondaire. Dans l'enseignement supérieur, la domination du français sur les langues autres que l'espagnol est moins forte, mais le français distance encore largement l'allemand, l'italien, le russe et le japonais (dans l'ordre). Quelles sont les trois variables dont il vient d'être question et à quelle échelle de mesure appartiennent-elles?

d) Recherche : Obtenez des données récentes à propos des phénomènes cités dans les questions ci-dessus.

(Sources : Atlas de la langue française, Bordas, 1995. Données de 1992 pour la Moldavie et de 1990 pour les États-Unis.)

CHAPITRE 2 LES RAPPORTS

TABLE DES MATIÈRES

1. Qu'est-ce qu'une proportion?

2. La proportion sous toutes ses formes

3. Les fréquences relatives

4. D'autres rapports : les comparaisons

• Exercices supplémentaires

Que seraient les journaux si on n'y retrouvait pas quelques chiffres provocateurs? En feuilletant ce matin notre gazette favorite, nous avons été fortement impressionnés par les titres suivants : le 35e meurtre de l'année (ils sont tous numérotés) sur le territoire de la Capitale nationale; le Conseil du Trésor annonce des coupures de plus de 100 000 $; Elvis Gratton a perdu 10 kilos; enfin, ce qui porte notre angoisse à son comble, le ministre Untel prévoit qu'un Québec indépendant perdra 2 millions d'emplois.

Réflexion faite, ces chiffres tout nus ne nous disent pas grand-chose. Combien de meurtres y avait-il eu l'an dernier à pareille date? Quelles sont les dépenses *totales* du Conseil du Trésor? Combien pesait Elvis avant sa dépression? Combien y a-t-il de Québécois et de Québécoises au travail? Nous ne pouvons pas vraiment comprendre toute cette soi-disant information sans utiliser des points de repère. Après avoir constaté, par exemple que le déficit du gouvernement était de 35 milliards de dollars, nous pouvons relativiser la nouvelle des coupures : 100 000 $, c'est moins qu'une goutte d'eau dans la mer des dépenses publiques. Quant au ministre Untel, le voilà complètement discrédité à nos yeux, car à ce rythme-là il y aura bientôt plus de chômeurs que d'habitants au Québec.

Au chapitre 1, nous nous sommes limités aux chiffres bruts, c'est-à-dire des chiffres qui n'avaient subi aucun traitement. Avec ce chapitre, nous commencerons à faire *travailler* les chiffres. Nous verrons toute l'information supplémentaire que nous pouvons tirer d'une donnée brute en la mettant en *rapport* avec une autre ou, tout simplement, en divisant une donnée par une autre.

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

- Comment les rapports peuvent-ils nous être utiles à mieux comprendre le sens des données brutes?
- Quels sont les principaux types de rapports que l'on utilise?
- Quelles sont les différentes formes sous lesquelles les rapports sont présentés?

1. QU'EST-CE QU'UNE PROPORTION?

En 2013, au Canada, 22 % des travailleurs ne détenaient aucun diplôme et 29 % d'entre eux étaient syndiqués; le taux de chômage des femmes atteignait de 6,7 % et celui des jeunes de 15 à 24 dépassait les 13 %; un tiers des voitures particulières vendues provenaient de l'extérieur de l'Amérique du Nord, tandis que la part de marché de Toyota était de 10 %.

Malgré la diversité des formulations utilisées dans la phrase précédente, toutes les données chiffrées citées possèdent un point commun. Ce sont toutes des *proportions*. Mais qu'est-ce qu'une proportion, exactement?

1.1. Un petit se faufile entre les géants

Nous sommes en 1992, une année cruciale pour les producteurs de microordinateurs. Dix ans après le lancement du premier PC, la compagnie IBM, qui n'a cessé de céder du terrain à ses imitateurs (les fabricants de « compatibles »), voit sa première place menacée. Une nouvelle stratégie est alors mise en place sous l'égide d'un nouveau président. Si, en 1992, IBM ne vend plus que 10 % des microordinateurs dans le monde, sa part de marché remonte à 10,8 % dès l'année suivante. C'est suffisant pour garder quelques distances avec l'ennemi de toujours (aux yeux du public) : Apple. Ce dernier réduit cependant légèrement son écart en passant de 8,5 % à 9,5 % du marché pendant la même période.

Qui va l'emporter : IBM ou Apple? Aucun des deux, car un troisième larron va venir se faufiler entre les géants. Compaq, qui était devenu le premier parmi les imitateurs d'IBM dans les années 1980 grâce à ses innovations audacieuses, traine de la patte. Lui aussi est en perte de vitesse, après des années de prospérité, et sa part de marché n'est plus que de 4,8 % en 1992, loin derrière les grands. Les autres compagnies américaines (Dell, Packard-Bell, HP), les Japonais (Nec, Toshiba) et les Taiwanais (AST, Acer) se rapprochent dangereusement. Qu'importe, car Compaq vient de se dénicher un nouveau président en Europe : Eckhard Pfeiffer. Ce dernier établit une stratégie très ambitieuse.

Deux ans plus tard, en 1994, sur les 48 millions de microordinateurs vendus* dans le monde au cours de l'année, 4,8 millions ont été fabriqués par Compaq. De l'autre côté, après une légère remontée en 1993, les ventes d'IBM et d'Apple sont retombées à 4 millions chacune pour 1994. Apple peut se consoler en constatant qu'elle a maintenant « rattrapé » IBM en reculant moins vite qu'elle.

Source : Fortune, 17 avril 1995. Le marché mondial atteindra un sommet de 341,3 millions de microordinateurs vendus en 2011, comme nous le verrons en exercice.

La question qu'il faut se poser est la suivante : Compaq a-t-il augmenté sa part de marché en 1994? Nous ne pouvons y répondre sans effectuer un petit calcul. Il est clair que Compaq détient 1/10 du marché (4,8/48). On peut dire aussi que la proportion du marché détenue par Compaq est de 10/100 ou 10 %. Ce simple calcul nous permet de mieux situer cette entreprise dans le temps (elle a doublé sa part de marché en deux ans) et dans l'espace (elle a rattrapé les autres grands).

La proportion est égale à la partie divisée par le tout.

Pour comparer l'évolution des trois compagnies, nous avons utilisé un type de rapport très répandu : la proportion, c'est-à-dire la partie rapportée au tout. Même si la proportion se cache parfois sous un autre nom (ici, elle s'appelle « part de marché »), il faut savoir la reconnaître. Il faut aussi être en mesure d'identifier ses deux composantes : de quelle *partie* et de quel *tout* est-il question? Dans notre exemple, le tout représente la valeur des ventes totales de microordinateurs et la partie représente les ventes de Compaq. Cette compagnie a accaparé 10 % du gâteau.

1.2. Une seule vérité, plusieurs visages

Il n'y a qu'une proportion (c'est toujours la partie divisée par le tout), mais il y a plusieurs manières de la présenter. Calculons la proportion occupée par IBM tout en gardant à l'esprit que la manière de présenter ou de nommer des chiffres dépend de l'utilisation que l'on veut en faire.
La part de marché détenue par IBM est de 4 millions par rapport à un total de 48 millions d'ordinateurs vendus dans le monde en 1994.

Proportion = Partie/Tout

On peut présenter cette proportion sous forme de :

- nombre décimal : 4/48 = 0,083
- pourcentage : (4/48) × 100 % = 8,3 %
- fraction : 4/48 = 1/12

Pourquoi utiliser une forme ou l'autre?

Le *nombre décimal*, résultat brut de l'opération effectuée (4 / 48), risque de semer la confusion chez le lecteur ou l'interlocuteur, du moins dans ce contexte. Il est cependant bien commode : nous aurons l'occasion de nous en apercevoir à maintes reprises. D'ailleurs, pour transformer ce résultat en pourcentage, nous n'avons qu'à déplacer la virgule de deux crans vers la droite.

Le *pourcentage* est particulièrement approprié pour présenter de façon éloquente le poids d'IBM sur le marché : on peut dire que 8,3 % des microordinateurs vendus en 1994 dans le monde ont été fabriqués par IBM.

La *fraction* peut présenter des avantages et des inconvénients. La phrase « un microordinateur sur douze est fabriqué par IBM » sonne bien. On peut facilement « voir » la proportion en pensant à une douzaine d'œufs, une douzaine de bouteilles de bière, ou une douzaine d'apôtres. Mais ici, nous avons eu doublement de la chance : le numérateur ne peut pas être plus simple et le dénominateur est un beau chiffre rond. Cela n'arrive pas toujours (et pourtant nos chiffres sont authentiques!). D'autre part, la fraction n'est pas très commode pour faire des comparaisons. Compaq obtient 1/10 et IBM 1/12. Qui est le meilleur et quel est l'écart qui les sépare? Pas évident pour tout le monde. Il est certainement plus simple de comprendre que Compaq vaut 10 (pour cent) et IBM, 8,3 (pour cent).

Notons également que, contrairement à un mythe très répandu, la fraction, le pourcentage et le nombre décimal ne sont pas ici trois concepts différents, mais bien trois manières de présenter un même concept : la proportion.

La proportion est toujours comprise entre 0 et 1.

Supposons qu'en faisant des crêpes, nous laissions échapper sur le plancher la douzaine d'œufs dont nous parlions tantôt. Dans le pire des cas, la proportion d'œufs encore entiers est de 0/12. Dans le meilleur des cas, cette proportion est de 12/12. La proportion est donc toujours comprise entre 0 et 1 (ou 0 % et 100 %). De même, la part de marché d'un fabricant de microordinateurs est théoriquement comprise entre 0 (le fabricant ne vend rien) et 1 (il détient un monopole).

1.3. Obtenir un chiffre brut à partir d'une proportion

Sachant qu'IBM contrôle 8,3 % d'un marché de 48 millions de microordinateurs, on devrait être en mesure de retrouver le nombre total d'appareils vendus par cette firme.

8,3/100 × 48 millions = 3, 984 millions, soit approximativement 4 millions d'unités.

Comme nous avions arrondi notre pourcentage (de 8,3 %), nous pouvons bien arrondir notre chiffre de vente (à 4 millions).

Partie = Proportion × Tout

Sachant que 8,3 % = 8,3/100 = 0,083, nous pouvons accélérer le calcul en multipliant directement 48 millions par 0,083.

0,083 × 48 (millions) = 3,984 (millions), soit approximativement 4 (millions).

De façon similaire, si une proportion nous était présentée sous forme de fraction, nous pourrions quand même reconstituer la donnée correspondante. Lorsqu'on dit, par exemple, que 4 microordinateurs sur 15 sont produits par les trois « grands » (sur un total de 48 millions), cela fait combien d'ordinateurs en tout?

$$\frac{4}{15} \times 48 = \frac{4 \times 48}{15} = 12{,}8 \text{ (millions de microordinateurs)}$$

Ici encore, nous n'avons fait que multiplier la *proportion* (4/15) par le *tout* (48 millions) pour obtenir la *partie* (12,8 millions). Cette manière de présenter les choses n'est autre que la célèbre règle de trois.

Un dernier mot sur la proportion que nous venons d'utiliser : 4/15 = 0,266 = 26,6 %. C'est la part combinée des trois grands : Compaq (10 %), IBM (8,3 %) et Apple (8,3 %). En effet : 10 % + 8,3 % + 8,3 % = 26,6 %.

EXERCICES 1

1. La proportion, à l'endroit, à l'envers

a) En 1990, la consommation de pétrole dans le monde est de 2,8 milliards de tonnes sur une consommation totale d'énergie de 8,7 milliards de tonnes (mesurées en équivalent-pétrole). Quelle est la proportion de pétrole dans la consommation d'énergie?

b) En 1990, le bois représentait 10 % de la consommation totale d'énergie dans le monde. Le total s'élevait à 8,7 milliards de tonnes d'équivalents pétrole (TEP). Quelle était alors la quantité de bois consommée (en TEP)?

c) En 1993-94, le nombre d'entrées au cinéma au Canada était de 78,8 millions. Sachant que 2,3 millions de spectateurs ont fréquenté les cinémas en plein air, quelle est leur proportion par rapport à tous ceux qui sont allés au cinéma?

d) En 1985, 97,8 % de la population mondiale vivait dans son pays natal. Le reste, soit 105,5 millions de personnes vivait dans un autre pays. Quelle était la population mondiale en 1985?

e) En 1995, la population active chinoise était composée de 406,660 millions d'hommes et 316,623 millions de femmes. Les chiffres pour le Canada étaient respectivement de 8,353 et 5,607 millions. Quelle était la proportion de femmes dans la population active de chaque pays?

Sources : Natur Munich cité dans le *Courrier international* du 22 septembre 1994 (a et b). Statistique Canada, no 87-211 (c). Banque mondiale 1995 (d et e).

2. La proportion et ses formes

Relevez dans le texte suivant (tiré de la revue *Un coup d'œil sur l'agriculture canadienne* d'octobre 1994) les chiffres qui sont des proportions. Pour chacune des proportions, indiquez sous quelle forme elle se présente (nombre décimal, pourcentage ou fraction) et identifiez la *partie* et le *tout* qui la constituent.

« Selon la base de données sur la population agricole du recensement de 1991, parmi les 391 000 exploitants agricoles au Canada 241 000, ou 62 %, étaient des agriculteurs principaux (ceux dont l'emploi principal est l'agriculture par opposition aux agriculteurs secondaires). Le quart des exploitants étaient des femmes. L'Ontario comptait 30 % des agriculteurs secondaires du Canada. 43 % des agriculteurs secondaires avaient obtenu un grade ou un diplôme supérieur à un diplôme d'études secondaires, comparativement à un quart des agriculteurs principaux. Le tiers des agriculteurs principaux ayant terminé leurs études postsecondaires ont étudié dans un domaine relié à l'agriculture. Les trois légumes les plus cultivés (maïs sucré, pois verts, tomates) représentaient 55 % de la superficie consacrée à la culture maraîchère. En 1990, près de 13 % du produit intérieur brut de la Saskatchewan provenait de l'agriculture. »

3. Le bien-portant imaginaire

En 1995, quelques jours après la mort de François Mitterrand, son médecin personnel publiait un livre (aussitôt interdit) décrivant les cachotteries du président français sur son état de santé. Ce dernier, qui avait le cancer de la prostate depuis une douzaine d'années, prétendait être en parfaite santé pour assumer les devoirs de sa charge. La censure, cette fois, ne fut pas assez rapide : le jour même de l'interdiction du livre, une copie numérisée de ce dernier avait quitté la France pour les États-Unis, via l'Écosse. Par la magie du réseau internet, qui prenait alors son essor, le livre secret était diffusé dans le monde entier.

Complétez les cases vides du tableau 2.1.

Tableau 2.1 - Les décès par cancer au Canada en 1995

Type de cancer	Fréquence	Proportion en %	Proportion en nombre décimal
Cancer du poumon		33	
Cancer colorectal	3 700		
Cancer de la prostate			0,12
Autres types de cancer	14 835		
Total	**33 700**	100	1

Source : Tendances sociales canadiennes, n° 39, hiver 1996.

4. Le déclin relatif des ordinateurs personnels

Au début de cette section, il était question des parts de marché des principaux fabricants d'ordinateurs au début des années 1990. Comme on peut s'en douter, les ventes totales de microordinateurs (48 millions d'unités en 1994) n'ont cessé de progresser par la suite, pour atteindre un premier plafond en 2011 (avec 355,2 millions d'unités). Bon nombre des protagonistes des années 1990 sont encore présents en 2011. Parmi les principaux fabricants, on retrouve HP (fusionné avec Compaq depuis 2002, avec 17,2 % du marché), Dell (12,1 %), Acer (11,2 %), Apple (10,7 %) et Lenovo (successeur d'IBM PC, avec 9,3 % du marché). (Source : Gartner.)

Quel était le nombre total de microordinateurs vendus respectivement par HP et par Apple en 2001?

2. LA PROPORTION SOUS TOUTES SES FORMES

Les proportions sont souvent utiles pour traiter les données en sciences humaines. C'est ce que nous voulons illustrer en vous présentant diverses situations. Mais avant d'aller plus loin, rappelez-vous qu'il n'y a qu'une seule proportion (c'est la partie divisée par le tout), même s'il y a plusieurs manières de l'habiller.

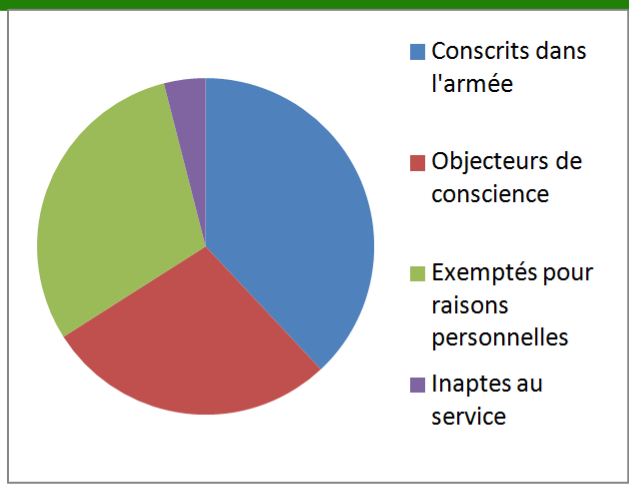
2.1. Où sont passés les soldats?

Après la fin de la Guerre froide, l'armée allemande (la *Bundeswehr*) se cherche un nouveau rôle (les missions internationales). Mais voilà, les jeunes Allemands ont, quant à eux, perdu la vocation. Bien que le service militaire soit obligatoire, seulement 38 % d'entre eux finissent par le faire. Parmi ceux qui ne le font pas, il y a d'abord les pacifistes (28 %) qui, pour des raisons de conscience, refusent d'entrer dans quelque armée que ce soit. Viennent ensuite les soutiens de familles et autres exemptés pour raisons personnelles (30 %). Il ne reste déjà plus que 42 % du contingent. Sur 42 soldats qui se rendent effectivement dans les casernes, 4 sont rejetés pour inaptitude (voir la figure 2.1). On notera que tous les rapports qui viennent d'être mentionnés sont des proportions.

Figure 2.1 - Les jeunes Allemands et le service militaire

	(en % des appelés)
Conscrits dans l'armée	38
Objecteurs de conscience	28
Exemptés pour raisons personnelles	30
Inaptes au service	4
Total	**100**

Source : Die Presse - Vienne, dans *Courrier International* , 30 mars 1995.



Voici que deux députés se disputent à coups de proportions dans l'enceinte du parlement (le Bundestag*. Pour le premier, il y a 4 % d'inaptes (sur 100 appelés, 4 sont jugés inaptes). Pour le second, il y a près de 10 % d'inaptes (sur 42 jeunes qui se rendent jusqu'à la caserne, 4 sont jugés inaptes). Pour nous, le désaccord n'a rien de mystérieux : les deux députés comparent la *partie* (les inaptes) à deux *touts* différents. Lorsqu'on a une idée claire de ce qu'est une proportion, on est déjà moins vulnérable aux trafiquants de chiffres.

D'après vous, comment s'appelle la Banque centrale d'Allemagne?

Il y a d'ailleurs une bonne façon de réconcilier les députés, c'est de leur donner tort à tous les deux. La proportion totale de personnes réellement inaptes au service (parmi tous les jeunes appelés) dépasse sans doute 4 %, car l'armée n'a pas eu l'occasion de tester les objecteurs et les exemptés.

2.2. Personne au bout du fil

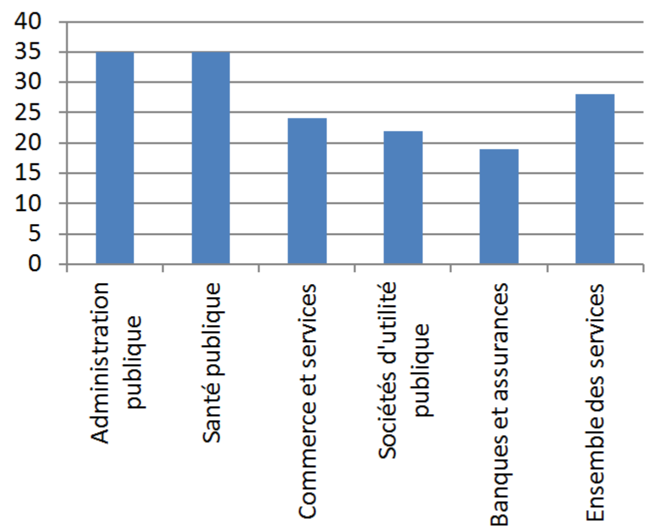
Les données brutes.

Un bureau d'étude des Pays-Bas* a essayé de rejoindre par téléphone divers services publics et privés. Sur 750 000 appels effectués par le biais d'un standardiste, 210 000 se sont avérés infructueux, c'est-à-dire qu'il a été impossible de rejoindre l'interlocuteur désiré. On peut déduire de ces données brutes que la proportion d'appels infructueux (que nous pourrions baptiser *taux d'échec*) est de 28 % (210/750). La figure 2.2 donne plus de détails.

La firme Quentel Services. Les données de cet exemple proviennent de NRC Handelsblad citée dans le *Courrier international* du 2 mars 1995.

Figure 2.2 - Taux d'échec des communications téléphoniques aux Pays-Bas

(Nombre d'échecs sur 100 tentatives pour obtenir un interlocuteur)	
Administration publique	35
Santé publique	35
Commerce et services	24
Sociétés d'utilité publique	22
Banques et assurances	19
Ensemble des services	28



Source : NRC Handelsblad, dans *Courrier International*, 2 mars 1995.

En passant, pourquoi croyez-vous qu'on a choisi de représenter ces données avec un diagramme en bâton plutôt qu'avec un diagramme circulaire comme dans le cas précédent (figure 2.1)?

Quelques nuances s'imposent.

Même si le problème est généralisé (aucun secteur n'y échappe), le taux d'échec varie en fonction du secteur appelé. Il est presque deux fois plus grand dans l'administration publique que dans le secteur des banques et assurances.

Pourquoi est-ce un problème?

Un appel infructueux coûte du temps et mobilise du matériel sans rien rapporter. Avec un taux d'échec aussi élevé, les pertes sont considérables. Le Nederlands Economisch Instituut évalue ces pertes à 500 millions de florins (450 millions de $ CAN). De plus, la personne dont on ne prend pas en compte la demande se sent frustrée et développe, à la longue, une aversion face au pourvoyeur de service impliqué.

Comment s'y prendre?

Le problème est généralisé, mais il prend des proportions alarmantes dans le secteur public. C'est sans doute par là qu'il faudrait commencer. En fait, il doit y avoir des solutions générales pour l'ensemble du pays et des solutions particulières à chaque secteur.

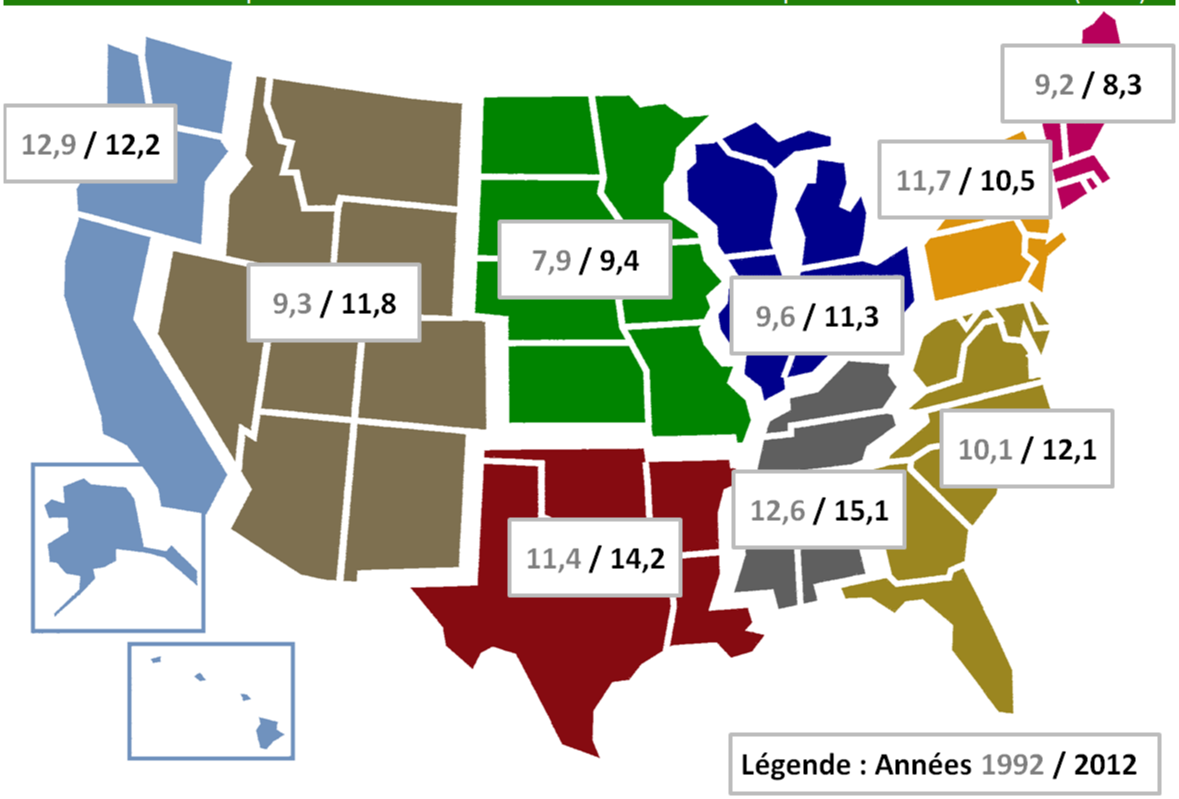
Vive les proportions!

Chacun des chiffres de la figure 2.2 est une proportion indépendante. Nous ne connaissons, par exemple, ni le nombre d'échecs ni le nombre de tentatives dans le secteur de l'Administration publique. Nous ne voulons d'ailleurs pas les connaître : seul le *rapport* entre ces deux chiffres nous intéresse ici.

2.3. La pauvreté aux États-Unis

La figure 2.3 présente la *proportion* de familles vivant sous le seuil de la pauvreté aux États-Unis pour chacune des neuf divisions de recensement. Étant donné que la population varie d'une division à l'autre, il aurait été difficile de se contenter de simples données brutes (le nombre de pauvres) pour faire des comparaisons. La carte peut être une façon originale et très « parlante » de représenter des proportions.

FIGURE 2.3 - Proportion des familles vivant sous le seuil de la pauvreté aux États-Unis (en %)



Source des données brutes : US Census Bureau, *American Community Survey*.

Note : L'unité de mesure choisie est le revenu familial corrigé en fonction du coût de la vie.

Il est peut-être étonnant de constater que la plus grande proportion de familles vivant sous le seuil de la pauvreté se trouve sur la côte ouest du pays en 1992. On aurait pu penser que le Sud serait plus gravement touché (on y trouve quand même 12,6 % de pauvres). C'est que les chercheurs qui ont produit cette carte ne se sont pas bornés à comparer le niveau de revenu des familles : ils ont apporté une correction à ces données en tenant compte du coût de la vie. C'est peut-être cet ajustement au coût de la vie qui fait grimper la proportion de familles vivant sous le seuil de la pauvreté dans la région ouest du pays. Par ailleurs, un État riche et tempéré tel que la Californie est plus susceptible d'attirer les pauvres chez lui que des régions moins favorisées.

Vingt ans plus tard, qu'en est-il de la pauvreté dans le plus riche des grands pays du monde? Au lieu de s'améliorer, la situation n'a fait qu'empirer, à l'exception notable du nord-est du pays, et, dans une moindre mesure, de la côte Pacifique.

EXERCICES 2

1. Le pétrole, nerf de la guerre

Le 2 août 1990, l'Irak envahit le Koweït, qu'il annexe quelques jours plus tard. Le 16 janvier 1991, les États-Unis entrent en guerre contre l'Irak. Pour comprendre le nœud du conflit, quelques données sur les réserves pétrolières pourraient s'avérer plus utiles que des dissertations sur les valeurs démocratiques.

a) Vérifiez que les chiffres présentés dans la figure ci-dessous représentent bien des proportions.

b) Personnellement, quel type de graphique auriez-vous choisi pour représenter les données? Justifiez votre réponse.

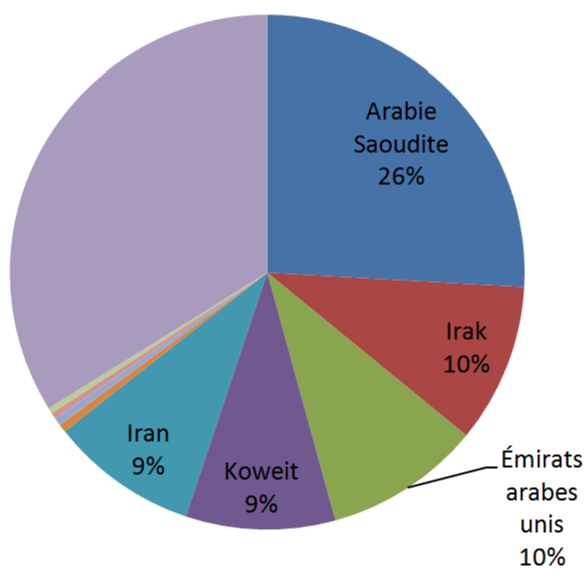
c) Quelle est la proportion des réserves mondiales de pétrole situées dans les pays riverains du Golfe Persique?

d) Évaluez le poids qu'aurait eu l'Irak dans le domaine pétrolier s'il avait annexé le Koweït. Commentez brièvement.

Figure 2.4 - Réserves pétrolières prouvées (au 1er janvier 1994)

	(en % du total mondial)
Arabie Saoudite	25,9
Irak	10,0
Émirats arabes unis	9,8
Koweit	9,4
Iran	9,3
Zone neutre	0,5
Oman	0,5
Yémen	0,4
Qatar	0,4
Autres pays	33,8
Total	**100,0**

Source : Financial Times dans *Courrier international*, 30 mars 1995.



2. Le spectre de Tchernobyl

En 1995, 10 ans après la catastrophe de Tchernobyl, 40,5 % de l'électricité produite en Ukraine provenait encore de l'énergie nucléaire (4,5 % pour les deux réacteurs encore en activité à Tchernobyl et 36 %, soit 71 TWH, pour les 14 autres réacteurs). À titre de comparaison, la proportion d'électricité produite par des centrales nucléaires représentait alors 0 % en Biélorussie (la région la plus affectée par la catastrophe), 20 % en République tchèque, 75 % en France (soit 358,6 TWH) et 2 % aux Pays-Bas.

a) Tracez un diagramme en bâton illustrant les proportions dont il est question dans le texte.

b) Calculez la production totale d'électricité en Ukraine et en France.

c) Recherche : quelles sont les proportions correspondantes pour le Québec et l'Ontario?

Note : Un TWH (térawatt/heure) = 1 milliard de KWH (kilowattheure).

3. Recherche : La carte électorale

Trouvez ou construisez une carte qui illustre les différentes valeurs que prend une proportion à travers un territoire (suggestions : la proportion de votes obtenus par un parti, la proportion de personnes parlant une langue particulière).

3. LES FRÉQUENCES RELATIVES

Nous avons déjà discuté des fréquences dans le chapitre précédent. Nous disions par exemple qu'on dénombrait 1 630 000 Musulmans en Bosnie lors du dernier recensement effectué dans la défunte Yougoslavie. Ce chiffre est une fréquence (nombre d'individus possédant une caractéristique donnée), c'est-à-dire une simple donnée brute. Il serait intéressant de *relativiser* cette *fréquence* brute, en comparant le nombre de Musulmans (la partie) à la population totale (le tout) de la Bosnie. On obtiendrait alors une fréquence relative.

La fréquence relative est donc une proportion comme une autre, si ce n'est qu'elle est appliquée systématiquement à chacune des catégories d'une variable qualitative. Nous reviendrons bientôt sur la Bosnie, après avoir fait un petit crochet du côté du Québec.

3.1. La famille « monoparentale sans enfant »

Qui sait, ce terme (en instance de brevet) est peut-être destiné à remplacer un jour le mot « célibataire » dans le dictionnaire de la rectitude politique. Selon le tableau 2.2, on dénombrait, à Montréal, 255 485 ménages qui répondaient à ce critère en 1993. Nous avons déjà présenté ces chiffres au chapitre précédent, mais cette fois, nous avons ajouté deux autres régions pour procéder à des comparaisons.

TABLEAU 2.2 - Répartition des ménages familiaux selon le nombre de membres

Fréquences absolues			
	Montréal	**Québec**	**Outaouais**
(nombre de membres)	(nombre de ménages)		
1 personne	255 485	57 825	**15 910**
2 personnes	243 320	63 015	**23 825**
3 personnes	118 445	35 480	**15 855**
4 ou 5 personnes	124 565	40 135	**18 480**
6 personnes ou plus	15 710	2 605	**1 400**
Total	**757 525**	**199 060**	**75 470**

Source : Statistique Canada, Cansim. Données de 1993.
Note : il s'agit des Communautés urbaines de Montréal, de Québec et de l'Outaouais.

La fréquence relative mesure la proportion occupée par chaque catégorie d'une variable.

Réflexion faite, il n'est pas très facile de comparer directement ces données brutes. C'est pourquoi nous allons calculer la proportion, ou fréquence relative, de chaque catégorie de ménage. Nous ferons successivement ces calculs pour chacune des régions.

$$\boxed{\text{Fréquence relative d'une catégorie} = \text{Fréquence de la catégorie/Fréquence totale}}$$

Fréquence relative de ménages d'une personne à Montréal = 255 485/757 525 = 0,337 = 33,7 %

Nous retrouvons ce chiffre dans le tableau 2.3 ci-dessous, ainsi que toutes les autres fréquences relatives. Sous cette forme, les données sont plus intéressantes. Elles « parlent » plus. En effet, on se rend compte, par exemple, que dans la région de Montréal, la proportion de ménages occupée par la catégorie *une personne* est plus élevée que dans l'Outaouais. Inversement, on compte un nombre *relativement* plus élevé de « grosses » familles dans l'Outaouais qu'à Montréal. La région de Québec, quant à elle, se situe entre ces deux extrêmes.

TABLEAU 2.3 - Répartition des ménages familiaux selon le nombre de membres

Fréquences relatives (en proportion du nombre total de ménages de la ville)

	Montréal	**Québec**	**Outaouais**
(nombre de membres)		(en %)	
1 personne	33,7	29,0	21,1
2 personnes	32,1	31,7	31,6
3 personnes	15,6	17,8	21,0
4 ou 5 personnes	16,4	20,2	24,5
6 personnes ou plus	2,1	1,3	1,9
Total	100,0	100,0	100,0

La fréquence cumulée d'une catégorie est égale à somme des fréquences de cette catégorie et des catégories précédentes.

Étant donné que la variable *nombre de membres du ménage* appartient à une échelle ordinale, nous pouvons grouper des données successives. Cela nous permettra de tirer encore plus d'information du tableau. On pourra se demander par exemple si les ménages de 2 personnes ou moins représentent la majorité. C'est ce que nous avons fait dans le tableau 2.4 ci-dessous, où nous nous sommes limités à une seule région. On y verra notamment que Montréal compte 498 805 ménages de 2 personnes ou moins (dans la colonne 2 qui représente les fréquences cumulées. En termes relatifs, les ménages de 2 personnes ou moins sont nettement en majorité, puisqu'ils représentent 65,8 % du total (c'est la fréquence relative cumulée : colonne 4). Ce dernier chiffre peut être obtenu de deux façons différentes :

1) on divise la fréquence cumulée par la fréquence totale : 498 805 / 757 525 = 0,658 = 65,8 %;

2) on additionne les fréquences *relatives* des deux premières catégories : 33,7 % = 32,1 % = 65,8 %.

TABLEAU 2.4 - Répartition des ménages familiaux à Montréal selon le nombre de membres

Fréquences cumulées

	Fréquence absolue		**Fréquence relative**	
	[1] Simple	**[2] Cumulée**	**[3] Simple**	**[4] Cumulée**
(nombre de membres)	(nombre de ménages)		(en %)	
1 personne	255 485	255 485	33,7	33,7
2 personnes	243 320	498 805	32,1	65,8
3 personnes	118 445	617 250	15,6	81,5
4 ou 5 personnes	124 565	741 815	16,4	97,9
6 personnes ou plus	15 710	757 525	2,1	100,0
Total	**757 525**		**100,0**	

Est-ce à dire que la plupart des Montréalais vivent dans des ménages de deux personnes ou moins? Loin de là! Le chiffre de 65,8 %, cité au paragraphe précédent, représente la proportion des ménages, et non celle des individus. On compte en réalité 742 125 individus dans les ménages de deux personnes ou moins (car [255 485 × 1] + [243 320 × 2] = 742 125). Pour les ménages de plus de deux personnes, le calcul est plus approximatif, à cause d'une moins grande précision sur le découpage des classes, mais on peut estimer le nombre total d'individus à au moins 1 000 000.

3.2. La mosaïque des peuples : un tableau à double entrée

Nous vous présentions dans le dernier chapitre un tableau sur la répartition des différentes nationalités en Bosnie. Nous ajoutons maintenant une seconde dimension à ce tableau en l'élargissant à l'ensemble des républiques de l'ex-Fédération de Yougoslavie.

Dans un tableau de fréquences à double entrée, les colonnes représentent les catégories d'une première variable et les lignes représentent les catégories d'une seconde variable.

Le tableau 2.5 ci-dessous indique le nombre d'individus (la fréquence) en fonction de la nationalité à laquelle ils s'identifient et de la république où ils demeurent au moment du recensement. Il s'agit donc d'un tableau à deux dimensions ou à double entrée. La variable *Nationalité* est divisée en quatre catégories mutuellement exclusives (les quatre lignes du tableau). Nous avons regroupé dans la dernière catégorie les autres Slaves (ceux qui se considèrent comme Yougoslaves tout court, les Slovènes, les Macédoniens, les Bulgares et les Slovaques) et les non-Slaves (Hongrois, Roumains, Albanais et Turcs). Nous avons également découpé la variable *république* en quatre catégories (les quatre colonnes du tableau). Le tableau 2.5 compte donc 16 cases (4 lignes × 4 colonnes) en plus des totaux. Comme on le voit, il s'agit d'une véritable « macédoine » de peuples.

TABLEAU 2.5 - Les nationalités dans l'ancienne Yougoslavie

Fréquences absolues

(en milliers de personnes, d'après le recensement de 1981)

	Croatie	Bosnie	Serbie	Autres républiques	Yougoslavie
Serbes	532	1 321	6 747	86	**8 686**
Croates	3 455	758	140	56	**4 409**
Musulmans	24	1 630	289	40	**1 983**
Autres nationalités	429	326	1 984	3 455	**6 194**
Total	**4 440**	**4 035**	**9 160**	**3 637**	**21 272**

Source : Atlas des peuples d'Europe centrale, La Découverte, 1991.

Note : Les Serbes, les Croates et les Musulmans partagent la même langue écrite, aussi certains se déclarent-ils Yougoslaves sans autre précision. Les Monténégrins sont inclus avec les Serbes.

Pour mieux comprendre la situation, il vaut mieux transformer ces données en fréquences relatives. C'est ce que nous avons fait de deux façons différentes. Dans le premier cas (tableau 2.6), nous avons indiqué pour chaque république (chaque colonne) la proportion occupée par les différentes nationalités. Ainsi, la Croatie compte, à l'époque du recensement, 12 % de Serbes (car 532/4440 = 0,12 = 12 %) et 77,8 % de Croates (3455/4440 = 0,778 = 77,8 %). On remarque aussi que les Serbes sont deux fois plus nombreux que les Croates dans l'ensemble de la Yougoslavie (40,8 % contre 20,7 % dans la colonne *Total Yougoslavie*). Le total de chaque colonne donne évidemment 100 %.

TABLEAU 2.6 - Les nationalités dans l'ancienne Yougoslavie					
Fréquences relatives en colonne					
(en % de la république)					
	Croatie	**Bosnie**	**Serbie**	**Autres républiques**	**Yougoslavie**
Serbes	12,0	32,7	73,7	2,4	**40,8**
Croates	77,8	18,8	1,5	1,5	**20,7**
Musulmans	0,5	40,4	3,2	1,1	**9,3**
Autres nationalités	9,7	8,1	21,7	95,0	**29,1**
Total	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**

Dans le tableau 2.7, nous avons voulu mettre en évidence la répartition des communautés à travers le territoire. Ainsi, on peut voir à la deuxième ligne du tableau que 78,4 % des Croates habitent en Croatie proprement dite et que 17,2 % d'entre eux vivent en Bosnie.

TABLEAU 2.7 - Les nationalités dans l'ancienne Yougoslavie					
Fréquences relatives en ligne					
(en % de la nationalité)					
	Croatie	**Bosnie**	**Serbie**	**Autres républiques**	**Yougoslavie**
Serbes	6,1	15,2	77,7	1,0	**100,0**
Croates	78,4	17,2	3,2	1,3	**100,0**
Musulmans	1,2	82,2	14,6	2,0	**100,0**
Autres nationalités	6,9	5,3	32,0	55,8	**100,0**
Total	**20,9**	**19,0**	**43,1**	**17,1**	**100,0**

D'une façon générale, les deux tableaux de fréquence relative (en colonnes ou en lignes) illustrent le fait que nationalité et territoire étaient loin de coïncider dans l'ex-Yougoslavie. On peut, par exemple, affirmer que 32,7 % des Bosniaques sont des Serbes (tableau 2.6) ou encore que 23,3 % des Serbes vivent à l'extérieur de la Serbie (tableau 2.7). Tout dépend du point de vue. Et il ne s'agit là que d'un exemple : les tableaux de fréquences relatives donnent un éclairage particulièrement riche à la situation.

EXERCICES 3

1. Québec, juste milieu entre deux extrêmes

a) Calculez les fréquences relatives cumulées pour la région de Québec en vous servant uniquement des données du tableau 2.2.

b) Calculez les fréquences relatives cumulées pour la région de l'Outaouais vous servant uniquement des données du tableau 2.3.

c) Tracez un histogramme pour la région de Québec en vous servant des proportions comme unité de mesure sur l'axe vertical.

d) Commentaire : comparez les trois régions.

e) Recherche : mettez les données à jour pour les trois régions.

2. Tableau à double entrée

a) À partir des données du tableau 2.8, construisez deux autres tableaux contenant les fréquences relatives selon le sexe (en ligne) et selon la situation sur le marché du travail (en colonne).

b) Corrigez les mensonges suivants : *38,9 % des femmes sont au chômage et les hommes monopolisent 86,2 % des emplois.*

TABLEAU 2.8 - Les chômeurs et les chômeuses au Québec

	Hommes	Femmes	Total
	(en milliers de personnes)		
Personnes employées	1 726	1 421	**3 147**
Personnes au chômage	277	176	**453**
Total (population active)	**2 003**	**1 597**	**3 600**

Source : Statistique Canada, La population active, mars 1996.

3. Des preuves SVP

On affirme un peu plus haut, dans les commentaires des tableaux 2.6 et 2.7, que 32,7 % des Bosniaques sont des Serbes et que 23,3 % des Serbes vivent à l'extérieur de la Serbie. Prouvez-le en montrant comment ces proportions ont été obtenues.

4. D'AUTRES RAPPORTS : LES COMPARAISONS

Les proportions, que nous venons d'observer sous différentes coutures, sont un type de rapport particulier. Nous allons maintenant examiner les autres types de rapports.

4.1. Des rats et des hommes

À Cheremietievo (Russie), un avion a été dans l'impossibilité de décoller (en 1994). Motif : les rats avaient dévoré les fils de la cabine de pilotage. Le rat a été le compagnon (et l'ennemi) de l'être humain depuis plusieurs siècles, grâce à ses stupéfiantes capacités d'adaptation. Il survit par -40° aussi bien que par +40°. Ses dents peuvent exercer une pression de 1700 kg/cm², ce qui est suffisant pour percer des blocs de béton ou des tuyaux de plomb. En un an, une femelle adulte a entre 2 et 8 portées de 8 à 12 petits capables de se reproduire dès l'âge de trois mois.

Alors que la proportion représente la partie divisée par le tout, d'autres rapports sont du type variable 1/variable 2.

Il y a dans le monde environ un rat par habitant. À Moscou, ce taux a nettement augmenté, après l'effondrement du régime communiste : on comptait, en 1994, cinq rats pour un Moscovite. Ces chiffres représentent un *rapport* : de même que l'on disait qu'un microordinateur sur douze était vendu par IBM, on peut affirmer que Moscou comptait cinq rats par habitant. Il y a cependant une différence notable entre ces deux rapports. Alors que le premier est une proportion (les ordinateurs IBM font partie de l'ensemble des ordinateurs vendus), le deuxième met en relation deux variables différentes : l'effectif des rats et l'effectif des humains (les rats ne font pas partie de la population moscovite!). Il existe ainsi de nombreux rapports qui permettent de relier des variables différentes (la pression de 1700 kg/cm² en était un exemple). Pour bien les interpréter, il convient de les distinguer des proportions, même s'ils se cachent souvent sous le même déguisement.

Saviez-vous que les rats adorent les câbles électriques (et particulièrement les câbles d'ordinateur)? Aux États-Unis, on a estimé que 20 % des incendies étaient provoqués par des rats qui avaient rongé les gaines des câbles électriques*.

Les données de cet exemple proviennent de Moskovskie Novosti cité dans le *Courrier international* du 17 novembre 1994.

4.2. Comparer deux variables apparemment bien distinctes

Lorsqu'il existe une relation entre deux variables, il peut être intéressant d'établir un rapport. On parlera ainsi d'habitants au km² (c'est le territoire qui permet à l'être humain de vivre), de lignes téléphoniques par habitant (ce sont les gens qui utilisent le téléphone), de quintaux de blé à l'hectare (c'est la terre qui fait pousser le blé).

Le coût horaire du travail est un bon exemple de rapport mettant en jeu deux variables différentes : d'une part les dépenses effectuées par l'entreprise pour utiliser un travailleur et, d'autre part, le temps travaillé. On estime qu'en Allemagne (patrie de Mercedes, Volkswagen et Audi), le coût horaire du travail est de 38,22 $US (en 1994). Ce coût n'est que de 18, 06 $ en Grande-Bretagne (patrie de… Toyota et Honda, les fleurons de l'industrie automobile anglaise ayant été sacrifiés sur l'autel du néolibéralisme thatchérien).

Les travailleurs allemands de l'automobile gagneraient plus du double de leurs collègues britannique?

Il y a toutefois une différence importante entre le coût payé par l'entreprise et le salaire reçu par l'ouvrier. L'entreprise doit aussi tenir compte, dans ses coûts horaires, des congés payés, des cotisations à l'assurance chômage et à l'assurance maladie, des avantages marginaux, des primes, etc. On ne peut donc se fier au seul coût horaire pour déduire le salaire réellement reçu par l'employé.

Par ailleurs, il est possible que les travailleurs allemands méritent de meilleurs salaires parce qu'ils sont plus productifs que les Anglais. La *productivité* est aussi un rapport entre deux variables, puisqu'elle représente ici la valeur de la production (en dollars) divisée par la quantité de travail (en heures). Comment les Allemands pourraient-ils être plus productifs? Les hypothèses sont légion : ils pourraient être mieux équipés, mieux formés, mieux encadrés, plus motivés, plus vaillants que leurs collègues britanniques. Mais il s'agit là d'un autre sujet d'étude.

4.3. Comparer deux variables qui n'en font qu'une

Lorsqu'une variable comporte exactement deux catégories on utilise souvent le rapport Catégorie 1/Catégorie 2.

Sur mille naissances au Québec, on compte, bon an mal an, 513 garçons pour 487 filles, ce qui donne en gros 105 garçons pour 100 filles. Si on veut vraiment être précis (pourquoi pas?), on peut affirmer que pour chaque petite Québécoise, il naît 1,053 (513/487) petit Québécois. Curieuse façon de présenter les choses, n'est-ce pas? Mais somme toute, l'information est passée. On désigne ce rapport par l'expression *ratio de masculinité*.

Dans cet exemple, on ne peut pas parler de proportion : les 513 garçons ne font pas partie des 487 filles. Ces garçons et ces filles ne sont que les deux catégories d'une même variable : les bébés mis au monde. Nous aurions pu utiliser ici une proportion : 51,3 % des naissances sont masculines (car 513/ (513 + 487) = 0,513 = 51,3 %). Nous avons pourtant préféré présenter les choses sous un jour différent. Le ratio de masculinité (1,053 ou 105,3 pour 100) met mieux en évidence l'écart entre garçons et filles.

Le tableau 2.9 illustre ce genre de rapport. Dans la première partie du tableau (colonnes 1 et 2), nous avons choisi de diviser le nombre de filles par le nombre de garçons pour mieux mettre en évidence l'écart entre les deux seules catégories que compte la variable (car on s'attend à des écarts, du moins pour certains pays). Par contre, dans la deuxième partie du tableau (colonnes 3 et 4), nous nous sommes contentés d'une simple proportion : dans ce cas-ci, on se préoccupe moins de mesurer l'écart que de déterminer la distance qui reste à parcourir pour atteindre le seuil de 50 %.

TABLEAU 2.9 - Les femmes à l'école et au travail en Scandinavie				
	Nombre de filles pour 100 garçons à l'école secondaire (Filles/Garçons)		**Proportion de femmes dans la population active (Femmes actives/Total actifs)**	
	1991	**2012**	**1992**	**2012**
	[1]	**[2]**	**[3]**	**[4]**
Finlande	111	105	47	48
Norvège	105	98	41	47
Danemark	106	191	45	47
Suède	109	98	45	47
Canada	96	98	40	47
Égypte	76	98	10	24

Source : Banque mondiale, Rapport sur le développement dans le monde 1994; IDM 2014.

Comme toujours, il faut interpréter les chiffres avec précaution. Il se pourrait, par exemple, que beaucoup de femmes *adultes* suivent des études secondaires, ou que la différence de taux d'activité entre sexes ne concerne que la génération la plus âgée.

4.4. Des petites boutiques aux grandes surfaces

On sait que le développement des commerces à grande surface a bouleversé les habitudes de consommation et entraîné la disparition de nombreux petits détaillants. Le tableau 2.10 montre que ce phénomène n'a pas touché tous les membres de l'Union européenne (en 1994) de la même façon. Les pays les moins avancés (Portugal, Grèce) conservent relativement plus de détaillants que les pays où le développement des grandes surfaces est plus ancien.

TABLEAU 2.10 - Les petits commerces en Europe	
	Nombre de magasins pour 10 000 habitants
Portugal	175
Grèce	174
Italie	161
Belgique	128
Espagne	117
Danemark	94
Luxembourg	93
Irlande	84
France	82
Allemagne (ouest)	70
Pays-Bas	64
Royaume-Uni	61

Source : Die Welt dans Courrier international, 17 mars 1994.

Est-ce vraiment le seul facteur qui permette d'expliquer l'évolution du phénomène?

On y retrouve aussi une dimension culturelle. La Belgique, dont le niveau de vie est semblable à celui des Pays-Bas, a tenu à garder ses petits commerces. Les chiffres du tableau représentent le rapport entre deux variables (le nombre de magasins et le nombre d'habitants). Pour rendre les données plus lisibles, et compte tenu du fait qu'il y a beaucoup plus de clients que de commerçants (merci pour le commerce!), on a choisi de calculer les magasins par tranche de 10 000 habitants. On obtient ainsi un éventail de chiffres ni trop gros ni trop petits. Cela dit, on aurait pu utiliser une autre unité et dire par exemple que l'Allemagne compte 7 magasins pour 1000 habitants, ce qui est pareil à 70 magasins pour 10 000 habitants.

4.5. Une *forme* et un *nom* adaptés au sujet étudié

Nous vous proposons, pour conclure, une série de rapports entre deux variables. Nous verrons que les noms et les formes donnés à ces rapports varient en fonction de l'usage qu'on en fait et de certaines traditions. On ne peut donc pas se fier uniquement au *nom* pour comprendre la *nature* d'un *rapport* et on doit être en mesure de passer d'une forme à l'autre. Le tableau 2.11 permet de comparer, à travers quelques rapports, le Canada à divers pays.

TABLEAU 2.11 - Comparaison entre pays : divers rapports

	Taux de natalité	**Habitants par médecin**	**Taux de pénétration du téléphone**	**Taux de couverture des importations de biens**
	[1]	**[2]**	**[3]**	**[4]**
	Naissances / Population (pour 1000 habitants)	Habitants / Médecins	Lignes / Population (pour 1000 habitants)	Exportations / Importations (en %)
	1992	1990	1990	1992
Niger	52	34 850	1	93,1
Égypte	28	1 320	33	36,8
Italie	10	210	388	96,7
Canada	15	450	577	108,1
Japon	11	610	441	147,0

Source : Banque mondiale, Rapport sur le développement dans le monde 1994.

Les rapports qui mettent en relation deux variables, ou deux catégories d'une même variable sont souvent appelés ratios, taux ou même coefficients, mais il s'agit d'une simple question d'usage. Fiez-vous au concept et non aux mots.

Dans la première colonne de chiffres (taux de natalité), on note que le rapport concerne deux variables différentes : les naissances ne constituent pas une des catégories de la population*. Les deux variables sont néanmoins reliées, car les naissances alimentent la population et une partie de cette dernière (les parents) est à l'origine des naissances. L'unité choisie est le nombre de naissances pour 1000 habitants, ce qui permet de présenter des chiffres ni trop gros ni trop petits. Le rapport naissance/population a été baptisé taux, comme beaucoup de rapports qui ne sont pas des proportions. Cependant, il ne faut pas se laisser impressionner par cette appellation. Dans des cas plus ou moins similaires, on entendra parler de coefficients ou de ratios. Parfois même, le mot *taux* sera employé pour une proportion (par exemple pour le taux de chômage qui est la proportion de personnes actives sans emploi).

Par contre, si on divisait le nombre de nourrissons, de citadins, de femmes ou d'athées) par la population, on aurait là des proportions : chacune de ces catégories forme en effet une partie de la population.

On observe dans la deuxième colonne du tableau (habitants par médecins) de grands écarts entre les chiffres, surtout si on compare ces rapports à ceux de la colonne précédente. Cela ne signifie pas nécessairement que la disparité soit plus marquée pour ce deuxième rapport. S'il est vrai que l'accès à un médecin peut être très inégal d'un pays à l'autre (le rapport est 57 fois plus élevé au Niger qu'au Japon), il faut reconnaître que les taux de natalité eux aussi sont aussi très éloignés les uns des autres (au Niger, il est proche du maximum physiologique, et au Japon, il est proche du minimum).

La colonne 3 illustre ce que l'on a nommé le « taux de pénétration du téléphone ». Il ne s'agit évidemment pas de la profondeur à laquelle on peut enfoncer les poteaux téléphonique selon les pays. On parle plutôt de lignes téléphoniques par habitant (par 1000 habitants en fait). Selon ce tableau, on trouverait, au Canada, moins de deux habitants par ligne (1000/577 = 1,73). Serait-ce à cause de la proportion élevée de « familles monoparentales sans enfants »MD dans la région montréalaise? Il ne faut pas oublier non plus que certaines personnes possèdent plusieurs lignes de téléphone alors que, dans le même temps certaines lignes appartiennent plutôt à des institutions qu'à des individus.

La colonne 4 illustre ce qu'on a désigné comme étant le taux de couverture des importations de biens. Dans la mesure où les recettes provenant des exportations peuvent servir à financer les importations, il est intéressant d'associer les deux variables dans un tel rapport. Ainsi, l'Égypte, avec un taux inférieur à 100 %, doit sans doute faire appel à d'autres sources de financement (tourisme, emprunts) pour payer ses importations, tandis que le Japon, de son côté, ne sait plus quoi faire de ses devises étrangères. Que pensez-vous de la position du Canada : le surplus est-il suffisant pour payer les voyages en Floride et les intérêts sur la dette extérieure?

EXERCICES 4

1. Conversions volontaires

Utilisez les données des tableaux 2.10 et 2.11.

a) Quel est le nombre de magasins pour 1000 habitants au Portugal et aux Pays-Bas?

b) Quel est le nombre de naissances pour 100 habitants en Égypte?

c) Quel est le nombre de médecins pour 10 000 habitants au Canada?

d) Quel est le nombre de médecins pour 1000 habitants au Canada?

e) Montrez que les rapports figurants dans les deux tableaux ne sont pas des proportions.

2. 2. La bataille d'Angleterre

a) Complétez les cases vides du tableau 2.12 figurant ci-après.

TABLEAU 2.12 - Coût horaire du travail dans l'industrie automobile

Pays	(en FF)	en % des ÉU	en $ US
Allemagne	182	144	**38**
Japon	145		**30**
États-Unis	126	100	
Suède	125	99	**26**
Belgique	121		**25**
Pays-Bas	103	82	**22**
France	95	75	**20**
Italie	87	69	
Espagne	86	68	**18**
Grande-Bretagne	86	68	**18**
Total	**757 525**	**199 060**	**75 470**

Sources : Fédération de l'industrie automobile allemande; The Independent on Sunday dans Courrier international, 30 mars 1995.

Note : 1 FF (franc français) = 0,21 $US en mars 1995.

b) Toyota et Ford-Mazda viennent d'annoncer de nouveaux investissements en Grande-Bretagne. Justifiez leur décision en utilisant les données du tableau.

c) Avec ou sans l'aide d'un tableur, représentez les salaires sous forme d'un graphe qui vous semble approprié pour illustrer les différents coûts horaires dans l'industrie de l'automobile.

EXERCICES SUPPLÉMENTAIRES

1. 1. Les formes de la proportion

a) Dans le texte suivant, identifiez les chiffres qui sont des proportions, indiquez sous quelle forme (nombre décimal, pourcentage, fraction) ces proportions se présentent et convertissez-les dans une des autres formes.

« Dans une expérience* menée en collaboration avec une station de télévision, des personnes ayant regardé un film de 12 secondes relatant une attaque à main armée, présentée durant le bulletin de nouvelles, ont eu l'occasion d'identifier l'assaillant sur une rangée de 6 suspects. Des 2000 spectateurs qui ont appelé après l'émission, seulement 15 % ont été capables d'identifier la bonne personne, un résultat semblable à ce qu'aurait prédit le simple hasard (1 chance sur 6 = 0,166 = 16,6%). »

Source : BUCKOUT, R. dans *Introduction à la psychologie*, McGraw-Hill, 1994.

b) Recherche : trouvez un texte où les rapports sont présentés sous leurs différentes formes.

2. Des taux : en veux-tu? En voilà!

Pour chacun des exemples suivants, dites s'il s'agit d'une *donnée brute*, d'une *proportion* ou d'un *rapport entre deux variables*. Dans ce dernier cas, identifiez les deux variables concernées. (*Note* :toutes les données sont authentiques.)

a) Le taux d'abstention au référendum de 1980 au Québec était de 16,0 %.

b) Le taux d'imposition provincial au Québec est de 23 % pour la tranche de revenu comprise entre 28 900 $ et 31 770 $ (pour l'année fiscale 1994).

c) Il y a une salle de cinéma aux États-Unis pour 10 000 habitants en 1993.

d) La radio internationale la plus écoutée dans le monde est la BBC avec un taux d'écoute de 130 millions d'auditeurs réguliers de ses programmes diffusés en 39 langues en 1993.

e) Le taux de suicide au Sri Lanka est un des plus élevés au monde : il est de 47 pour 100 000 habitants en 1991.

f) Avant son éclatement, l'Union soviétique avait un taux de nuptialité de 9,8 pour mille (en 1987).

g) En 1941, les Québécois constituaient 29 % de la population canadienne.

h) Le taux d'élucidation des crimes a été de 36,75 % en France en 1992.

i) En 1971, la ville de Montréal comptait 6,5 % de citoyens ayant l'italien pour langue d'usage.

j) Selon l'association allemande des propriétaires de chiens, l'Allemagne ne compte que 5,5 chiens pour 100 habitants contre 21,6 pour les États-Unis, 16,9 pour la France et 13 pour le Canada (en 1993).

k) Entre 1978 et 1992, 13,4 % des bébés-éprouvettes (fécondation in vitro) sont nés en France, 11,2 % aux États-Unis et 2 % en Allemagne.

l) Au Canada, les exploitants agricoles de sexe masculin sont proportionnellement à peu près 4 fois plus nombreux que les exploitants de sexe féminin à avoir fait des études en agronomie (en 1991).

3. Quelques petits calculs peut-être?

a) Calculez le nombre de crimes avec violence au Canada en 1987 et 1992, sachant que le Canada avait un taux de 856 crimes avec violence pour 100 000 habitants en 1987 et de 1081 pour 100 000 en 1992, et que la population du pays est passée de 26 549,7 milliers d'habitants en 1987 à 28,542 en 1992.

b) Calculez le nombre de quotidiens pour 100 000 habitants aux États-Unis en 1910 et 1989, sachant que les États-Unis comptaient 2202 quotidiens en 1910 pour une population de 81,7 millions d'habitants et 1642 quotidiens en 1989 pour une population de 248,8 millions d'habitants.

b) En Chine, le taux de natalité est de 19 pour mille et la population s'élève à 1 178,4 millions d'habitants (en 1993). Quel est le nombre de naissances à l'heure? Quel intervalle de temps moyen sépare deux naissances?

4. Le sirop d'érable, caviar du Québec

Les producteurs de sirop d'érable québécois ont recueilli 54,5 millions de livres de cet élixir en 1994, ce qui représente 61,3 % de la production mondiale. La majeure partie la production (70 %) a été vendue à l'extérieur de la province (*voir également le tableau 2.13 figurant ci-après*).

TABLEAU 2.13 - Exportations et production de sirop d'érable du Québec

A. Exportations selon les principaux pays

Pays de destination	**1991**	**1992**
	(en millions de $CAN)	
États-Unis	30,1	49,6
Japon	1,2	2,5
Allemagne	1,8	1,5
France	0,6	1,1
Autes pays	2,6	3,3
Total	**36,3**	**58,0**

Source : Les Affaires, 15 avril 1995.

B. Production totale du Québec et du Canada

	2009	**2010**	**2012**	**2013**
		(en millions de $CAN)		
Québec	346,3	252,7	268,2	346,1
Canada	395,2	291,1	304,5	408,1

Source : Statistique Canada, Cansim, tableau 001-0008.

a) Quelle est la production mondiale de sirop d'érable en 1994?

b) Quelle est la quantité de sirop d'érable vendue hors du Québec en 1994?

c) Quelle est la proportion de sirop d'érable exportée (à l'extérieur du Canada) qui est achetée par le Japon? Par les États-Unis?

d) Autrefois, la feuille de l'érable à sucre était le symbole des Québécois. Il pourrait d'ailleurs sembler étonnant, en examinant la partie B du tableau 2.13, de voir ce symbole figurer aujourd'hui sur le drapeau du Canada. Commentez ce point de vue, en vous appuyant sur des chiffres dérivés du tableau.

5. Plus de jugements, autant d'exécutions

a) À partir des données du tableau 2.14, calculez les fréquences relatives pour chacune des catégories de sentences en 1936. Faites le même exercice pour l'année 1954.

b) Pourquoi est-il impossible ici de calculer des fréquences cumulées? Pourquoi est-il impossible de tracer un histogramme?

TABLEAU 2.14 - Sentences des jugements pour meurtres au Canada

	1934	**1954**
Acquittement	7	13
Détention pour aliénation mentale	1	9
Peine capitale	10	10
Accusations	18	32

Source : Statistiques historiques du Canada, Ottawa, 1983.

6. La fréquentation des écoles

Le tableau 2.15 indique la répartition des effectifs scolaires au Québec entre 2000 et 2010.

TABLEAU 2.15 - Effectifs scolaires à plein-temps selon le niveau au Québec

	2000-2001	2009-2010
	(en milliers)	
Primaire (1re à 6e année)	575,9	463,2
Secondaire (à partir de la 7e année)	447,9	459,6
Collégial	213,4	213,8
Universitaire	233,6	275,5
Total	**1 470,7**	**1 412,1**

Source : Ministère de l'Éducation du Québec, Statistiques de l'éducation, 2011.

a) Calculez les fréquences relatives pour chaque niveau.

b) Calculez les fréquences relatives cumulées.

7. Le roi de l'aloyau

Le tableau 2.16 nous révèle quelques petits secrets sur les mangeurs.

TABLEAU 2.16 - Consommation des trois viandes reines au Canada

(en kg par année par habitant)

	1981	1991
Bœuf	40	33
Porc	31	27
Volaille	23	29

Source : Statistique Canada, Un coup d'oeil sur l'agriculture canadienne, octobre 1994.

a) Calculez la proportion de chaque catégorie en 1981 et en 1991.

b) Tracez un diagramme circulaire pour chaque année (suggestion : la surface de chaque « tarte » pourrait être proportionnelle au total de chaque année)

c) Faites un commentaire du tableau sans citer aucun chiffre. En quelques mots, qu'est-ce qu'on observe en 1991? Quels sont les changements par rapport à 1981?

DOSSIER 2 LE TOUR DU MONDE EN RAPPORTS

Les tableaux qui suivent nous permettront d'observer quelques facettes de l'être humain. On y retrouve *divers* rapports reflétant certains comportements et habitudes à travers le monde. Nous pourrons ainsi mieux nous connaître en faisant connaissance avec les autres. Nous avons choisi une brochette de huit pays de taille, de culture et de niveau de développement économique très variés.

Comment lire les tableaux

Les questions proposées un peu plus loin devraient vous aider à tirer le maximum d'information des tableaux, grâce aux notions acquises dans le chapitre. Avant de passer à ces questions, voici quelques observations sur les chiffres fournis. Vous pourrez ainsi vérifier si vous lisez correctement les données contenues dans les tableaux.

TABLEAU D2.1 - Proportion des femmes dans la population active

	1970	1992	2012
		(en %)	
Burundi	50	47	51
Côte-d'Ivoire	38	34	38
Tunisie	12	25	27
Mexique	18	27	39
Canada	32	40	47
France	36	40	47
États-Unis	37	41	46
Japon	39	38	42

Source de tous les tableaux : Banque mondiale, Rapport sur le développement dans le monde 1994; Indicateurs du développement dans le monde 2014.

Observation : Au Burundi, sur 100 personnes qui participent au marché du travail en 1992, 51 sont des femmes et 49 sont des hommes (en 2012).

TABLEAU D2.2 - Population urbaine

	Population totale 1992 (en millions)	Toutes les villes 1992	Villes d'au moins un million d'habitants 1992	Capitale 1990	Toutes les villes 2012
		(en % de la population totale)			
Burundi	6	6	0	4	11
Côte-d'Ivoire	13	42	19	18	52
Tunisie	8	57	23	20	67
Mexique	85	74	30	25	78
Canada	27	78	30	3	81
France	54	73	21	15	86
États-Unis	255	76	38	1	83
Japon	124	77	37	15	92

Observation : Au Japon, 77 % des gens vivent dans des villes en 1992. Plus précisément, 37 % des Japonais vivent dans des villes d'au moins un million d'habitants et 40 % vivent dans des villes de moins d'un million d'habitants. Pour être encore plus précis, 15 % des gens vivent à Tokyo et 22 % vivent dans les autres villes d'au moins un million d'habitants.

TABLEAU D2.3 - Déboisement net annuel, 1981-1990

	(en milliers de km²)	(en % de la superficie totale boisée)
Indonésie	12,1	1,0
Côte d'Ivoire	1,2	1,0
Thaïlande	5,2	2,9
Brésil	36,7	0,6
Canada	...	...
France	-0,1	-0,1
États-Unis	3,2	0,1
Suisse	-0,1	-0,6
Japon	0,0	0,0

Observation : L'Indonésie a perdu 12 100 km² de forêts par an (en moyenne) dans les années 1980. Ce chiffre tient compte à la fois des surfaces qui ont été livrées aux bûcherons et des surfaces qui ont été reboisées. Le pays a ainsi perdu annuellement 1 % de ses forêts. Note : dans ce tableau, nous avons inclus trois pays supplémentaires qui nous paraissaient particulièrement intéressants au chapitre des forêts.

TABLEAU D2.4 - Épargne intérieure		
	1970	**1992**
	(en % du PIB)	
Burundi	4	-2
Côte-d'Ivoire	29	14
Tunisie	17	21
Mexique	19	17
Canada	24	18
France	27	21
États-Unis	18	15
Japon	40	34

Note : l'épargne intérieure représente la part du PIB non consommée.

Observation : Au Canada, la population a consommé 82 % du PIB (produit intérieur brut) en 1992, que ce soit en achetant ces produits ou en les obtenant sous forme de services publics. Les 18 % restants ont été mis à la disposition des entreprises pour que celles-ci puissent investir dans leurs équipements. Au Japon, l'épargne est relativement plus élevée qu'au Canada : non seulement les entreprises n'ont aucune difficulté financer leurs projets d'investissements, mais il est même possible qu'une partie de cette épargne soit prêtée aux autres pays.

TABLEAU D2.5 - Nombre d'élèves par maître au primaire			
	1970	**1991**	**2012**
Burundi	37	66	47
Côte-d'Ivoire	45	37	42
Tunisie	47	26	17
Mexique	46	30	28
Canada	23	15	15
France	26	12	18
États-Unis	27	..	14
Japon	26	21	17

Observation : Il y a en 1991, dans le réseau scolaire burundais, 66 fois plus d'élèves que de professeurs.

TABLEAU D2.6 - Nombre de filles pour 100 garçons au primaire			
	1970	**1991**	**2012**
Burundi	49	84	99
Côte-d'Ivoire	57	71	85
Tunisie	64	85	98
Mexique	92	94	100
Canada	95	93	99
France	95	94	100
États-Unis	95	95	98
Japon	96	95	100

Observation : En 1991, sur 195 petits Américains fréquentant l'école primaire, on compte en moyenne 100 garçons et 95 filles.

TABLEAU D2.7 - Consommation d'eau douce						
	Usage domestique	**Usages industriels et agricoles**	**Usage domestique**	**Uage industriel**	**Usage agricole**	**Total**
	Moyenne 1970-1992		**2011**			
	(en m^3 par habitant)		(en m^3 par habitant)			
Burundi	7	13	5	2	23	**30**
Côte-d'Ivoire	15	52	28	14	31	**73**
Tunisie	41	276	34	10	203	**247**
Mexique	55	865	94	62	513	**669**
Canada	304	1384	262	919	158	**1339**
France	125	654	88	335	60	**484**
États-Unis	244	1624	210	708	618	**1536**
Japon	125	607	136	124	445	**704**

Observation : Pendant la période 1970-1992, les Français consomment en moyenne 125 m^3 d'eau douce par an à leur domicile. Par ailleurs, à la même époque, la consommation d'eau de l'industrie et de l'agriculture française équivaut à 654 m^3 multipliés par le nombre d'habitants.

TABLEAU D2.8 - Lignes téléphoniques par habitant				
	1970	1990	2012	Population 2012
	(lignes pour 1000 habitants)			(en millions)
Burundi	1	2	2	9,8
Côte-d'Ivoire	3	5	13	19,8
Tunisie	9	38	101	10,8
Mexique	16	66	170	120,8
Canada	304	577	507	34,8
France	84	495	621	65,7
États-Unis	329	545	435	313,9
Japon	158	441	505	127,6

Observation : Il y a environ 136,5 millions de lignes téléphoniques aux États-Unis en 2012, car 313,9 millions x (435/1000) = 136,5 millions. On note également un déclin du nombre de lignes par habitant en Amérique du Nord entre 1990 et 2012, suite à la généralisation du téléphone portable.

QUESTIONS

1. Questions générales

a) Vérifiez que les rapports qui figurent dans les tableaux D2.1 à D2.4 sont des proportions.

b) Les rapports qui figurent dans les tableaux D2.5 à D2.8 sont de type [Variable 1/Variable 2]. Trouvez ces variables pour chacun des tableaux.

c) Indiquez la forme sous laquelle sont présentés les rapports dans chaque tableau (*pourcentage* ou *autre forme*).

2. Les femmes actives (tableau D2.1)

a) Comment expliquez-vous l'évolution de la situation au Canada?

b) Quelle pourrait être, selon vous, l'évolution future au Canada?

3. La ville et ses attraits (tableau D2.2)

a) Combien de villes d'au moins un million d'habitants y a-t-il au Burundi en 1992?

b) Quelle est la proportion de Burundais qui vivent dans les autres villes que la capitale en 1992?

c) Pourquoi est-il probable que la Tunisie ne compte qu'une ville d'un million d'habitants ou plus en 1992?

d) Quels sont les pays qui correspondent le plus à cette description : « beaucoup de grosses villes et une métropole surdéveloppée »?

4. Les bûcherons (tableau D2.3)

a) Calculez la superficie totale boisée du Brésil.

b) D'après les chiffres, le déboisement de la forêt brésilienne vous semble-t-il alarmant?

5. Séraphin Poudrier (tableau D2.4)

Reliez les affirmations suivantes au tableau (année 1992).

a) Les entreprises japonaises ont plus de facilité à financer de gros investissements que les entreprises canadiennes.

b) Les Japonais se serrent relativement plus la ceinture que les Américains.

c) Non seulement les Burundais n'épargnent plus, mais ils utilisent l'épargne des pays étrangers.

6. Les classes surchargées (tableau D2.5)

a) On prétend souvent que les classes sont de plus en plus surchargées au primaire. Qu'en est-il d'après les chiffres?

b) Au Burundi, les voyageurs peuvent constater que la plupart des classes primaires comptent entre 30 et 40 élèves. Comment conciliez-vous cela avec les données du tableau?

7. Les filles à l'école (tableau D2.6)

a) Commentez l'évolution du rapport filles/garçons dans les trois premiers pays.

b) Comment expliquez-vous la baisse du rapport filles/garçons dans certains pays en 1991?

8. L'eau douce (tableau D2.7)

a) Combien de litres d'eau douce consomme le Canadien moyen chez lui chaque jour pendant la période 1970-1992?

b) Comment expliquez-vous les écarts de consommation industrielle et agricole entre les pays du tableau?

9. Allô, j'écoute! (tableau D2.8)

a) Parmi les éléments suivants, lesquels pourraient faire monter le taux de lignes téléphoniques par habitant au Canada?

- 1. Diminution de la taille des familles.
- 2. Augmentation de la population.
- 3. Baisse des frais pour les lignes d'affaires.
- 4. Baisse des frais pour les lignes résidentielles.

b) Le faible taux du Burundi signifie-t-il que les individus ont plus de difficultés à communiquer entre eux qu'au Canada?

10. Faisons mentir les chiffres

Montrez que les affirmations suivantes sont fausses et découlent d'une mauvaise lecture des chiffres présentés dans les tableaux.

a) Au Japon, 38 % des femmes sont actives en 1991 (tableau D2.1).

b) En Côte d'Ivoire, 79 % (42 + 19 + 18) des gens vivent dans des villes en 1992 (tableau D2.2).

c) En Indonésie, il ne se fait aucun reboisement (tableau D2.3).

d) En moyenne, un Tunisien épargne un montant d'argent plus élevé qu'un Canadien en 1992 (tableau D2.4).

e) Il y a en moyenne 66 élèves par classe au Burundi en 1991 (tableau D2.5).

f) Aux États-Unis, 95 % des élèves du primaire sont des filles en 1991 (tableau D2.6).

g) Les Canadiens sont plus propres que les Japonais (tableau D2.7).

h) Un Français sur deux (environ) a le téléphone chez lui en 1991 (tableau D2.8).

CHAPITRE 3 AUTOUR DE LA MOYENNE

TABLE DES MATIÈRES

1. La moyenne : un équilibre des forces

2. L'écart type : mesurer la dispersion des données

3. La courbe normale et l'écart type

4. D'autres indicateurs de dispersion

• Exercices supplémentaires

On dit parfois qu'une image vaut mille mots. De la même façon, un ou deux chiffres peuvent remplacer avantageusement plusieurs pages de tableaux. Le tout est de bien choisir ces chiffres. Dans ce chapitre, nous apprendrons justement à bien choisir et à bien interpréter ces quelques chiffres, qui résument d'un coup d'œil, l'ensemble d'une situation.

Avant de commencer, prenons un exemple tiré de l'hebdomadaire économique polonais *Gazeta Bankowa*, peu après l'effondrement du mur de Berlin. Les Polonais se trouvent à l'étroit, et à juste titre : ils ne disposent, en moyenne, que de 18 m² habitables par personne. C'est encore moins que les Japonais (28 m² en moyenne) qui n'ont pourtant pas la réputation d'occuper des logements spacieux. On voit ici qu'un simple chiffre, même s'il ne dit pas tout (il doit y avoir des Polonais plus à l'étroit que d'autres), nous dispense de consulter la liste des 40 millions d'habitants de cette république d'Europe de l'Est, si tant est qu'une telle liste existe. À titre de comparaison, l'Américain moyen dispose alors de 62 m² de surface habitable et l'Européen de l'Ouest de 37 m² en moyenne.

Comme on le dit vulgairement, les Américains se « pilent » moins sur les pieds que les Polonais. Il faut cependant faire quelques nuances. Même si l'Amérique est grande, tous les Américains ne logent pas à Beverly Hills. Il doit y avoir des écarts en fonction notamment du milieu social. Les disparités entre les gens sont encore plus élevées en Europe de l'Ouest, du moins en termes relatifs. Même si la moyenne d'espace de logement occupé par ces derniers est de 37 m² par habitant, la plupart des Européens de l'Ouest ont soit beaucoup plus, soit beaucoup moins. Les plus favorisés sont les gens de la campagne, ceux qui vivent dans des maisons de banlieue, ceux qui possèdent une résidence secondaire et plus généralement les personnes fortunées. Dans ce chapitre, nous verrons aussi comment évaluer rapidement l'importance de ces écarts autour de la moyenne.

Nous en profiterons également pour établir la position d'un individu à l'intérieur d'un groupe. Où se situe par exemple l'Américain qui occupe 100 m² de surface habitable? Fait-il partie de l'élite? Est-il dans les 10 % meilleurs? Dans les 40 %? Fait-il partie de la moitié la moins bien lotie de la population? Toutes ces situations étant possibles (même si la moyenne générale du pays est de 62 m²), il est parfois bon de connaître non seulement la moyenne, mais de posséder également des informations sur les écarts entre les éléments du groupe, ainsi que sur la position d'un élément particulier.

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

- L'individu « moyen » est-il celui qu'on risque le plus de rencontrer dans le groupe qui est le sien?
- Quel est le rapport entre la moyenne et les données qu'elle veut résumer?
- Comment peut-on situer une valeur (et le sujet qu'elle désigne) par rapport à la moyenne? Par rapport aux autres valeurs du même groupe?
- Comment les autres mesures comme la médiane, le mode ou les quintiles peuvent-elles nous être utiles à rendre la moyenne plus signifiante?

1. LA MOYENNE : UN ÉQUILIBRE DES FORCES

Tout le monde a entendu parler des moyennes. Tout le monde s'en sert. Comme on pouvait s'y attendre, notre but principal consiste davantage à bien interpréter les moyennes qui nous sont servies régulièrement dans les livres et les journaux que de nous livrer à de savants calculs. Cependant, nous ne sommes pas toujours de simples récepteurs d'information : nous avons parfois nous aussi quelque chose à dire. C'est pourquoi nous verrons comment bien choisir et bien calculer les moyennes. Comme toujours, nous prendrons nos exemples dans la réalité, car l'humanité a ceci de particulier : mieux on la connaît et plus on la trouve intéressante.

1.1. Le jour J : des péniches et des hommes

Jour J. Le 6 juin 1944. 745 navires, groupés en 38 convois, approchent des côtes normandes. En tout, 4006 péniches de débarquement sont mises à l'eau, au large des plages, et 185 000 hommes débarquent sur le territoire français. Regardons les choses à une échelle un peu plus humaine : combien d'hommes y avait-il en moyenne sur chaque péniche de débarquement? Essayez d'imaginer la réponse sans faire de calcul, mais plutôt en visualisant ces péniches. Et rappelez-vous qu'il ne s'agit pas d'un film hollywoodien, mais bien d'une histoire vécue par des jeunes gens tels que vous.

La moyenne est une mesure qui correspond à la somme des valeurs que prend une variable divisée par le nombre de ces valeurs.

Notons que la variable que l'on cherche à évaluer ici est la capacité, en nombre d'hommes, des péniches de débarquement. Nous ne disposons pas de la liste des péniches avec leur nombre respectif de passagers, mais nous connaissons le nombre de péniches et le total des passagers. Chaque péniche peut être associée à une *valeur* (le nombre d'hommes qu'elle contient). Le *nombre des valeurs* (le nombre de péniches) et la *somme des valeurs* (le nombre total de passagers) nous suffisent pour obtenir la moyenne des individus présents dans les péniches.

185 000 hommes/4006 péniches = 46 hommes en moyenne par péniche

Moyenne = Somme des valeurs/Nombre de valeurs

Les péniches de débarquement contenaient donc en moyenne 46 hommes chacun. Nous vous demandions plus haut d'essayer de visualiser ce résultat avant de le calculer. Si votre intuition correspondait au résultat, félicitations. Si vous aviez imaginé des péniches avec beaucoup moins d'hommes, ou beaucoup plus (nous sommes dans ce cas), cela prouve qu'il n'est pas toujours inutile de faire un petit calcul lorsque l'on veut avoir une idée *objective* de la réalité.

C'est donc dire que la plupart des péniches contenaient 46 hommes?

La moyenne ne prétend pas que tous soient égaux (il y aurait alors eu 46 hommes dans chaque péniche). Elle ne fait que nous dire quelle valeur aurait eu chacune des données si celles-ci avaient été toutes pareilles. Cela dit, rien ne prouve que la plupart des péniches contenaient 46 hommes. Certaines étaient peut-être très petites, d'autres très grandes. L'exemple qui suit nous démontrera que la moyenne est une valeur *autour* de laquelle les autres se situent et qu'elle n'est pas nécessairement celle qu'on retrouve le plus souvent.

1.2. César et ses Romains

Jules César, malgré ses nombreuses conquêtes, son immense talent et tous ses autres mérites, n'a jamais réussi à se faire couronner empereur. Sa carrière tardive fut brusquement interrompue lors d'une visite au Sénat, le 15 mars de l'année 44 avant notre ère. Le Sénat de l'époque était plus fringant que le nôtre! Auguste, fils adoptif de César (et son *vrai* fils selon les mauvaises langues) devait cependant fonder une dynastie qui dura près d'un siècle. Pour consoler les étudiants qui traînent encore dans les classes malgré leur âge avancé, signalons que Jules César n'a véritablement amorcé sa carrière qu'à 40 ans, après avoir réalisé de façon dramatique « qu'il n'avait encore rien fait de mémorable à un âge où Alexandre (le Grand) avait déjà soumis toute la terre ».

Si César et Auguste furent de grands hommes d'État (dont l'héritage est toujours vivant), on ne peut pas en dire autant de ceux qui leur ont succédé. Tous furent célèbres pour leurs débauches et leurs excès. Selon Suétone (l'auteur de la *Vie des douze Césars*), Tibère s'était fait installer un « jardin des plaisirs* » dans sa retraite de Caprée. Caligula, son successeur — et sans doute son meurtrier —, entretenait des relations incestueuses avec ses sœurs. Caligula nomma même son cheval sénateur (notez l'évolution du Sénat avec le temps). Il fut assassiné, tout comme ses successeurs Claude (homme plutôt lâche et influençable) et Néron (qui se prenait pour une grande vedette).

Avis aux étudiants qui n'aiment pas l'histoire : c'est le moment ou jamais de vous y intéresser!

Les présentations étant faites, revenons à nos chiffres. À quel âge, en moyenne, devenait-on empereur sous les Césars? Combien de temps restait-on sur le trône? À quel âge mourait-on?

Pour calculer ces moyennes, nous utiliserons la même formule que dans le cas précédent. La seule différence ici est que nous devons calculer nous-mêmes la somme (à l'aide des données du tableau 3.1 figurant ci-après).

TABLEAU 3.1 - La dynastie des Césars

	Naissance	Début du règne	Fin du règne et décès	Âge au début du règne	Durée du règne	Âge au décès
Auguste	-63	-24	**14**	39	37	**76**
Tibère	-42	14	**37**	55	23	**78**
Caligula*	12	37	**41**	25	4	**29**
Claude*	-10	41	**54**	50	13	**63**
Néron*	37	54	**68**	17	14	**31**
	Échelle d'intervalle			Échelle de rapport		

Source des trois premières colonnes de chiffres : Petit Robert 2.

* Mort assassiné.

Âge moyen au début du règne : (39 + 55 + 25 + 50 + 17)/5 = 37,2 ans
Durée moyenne du règne : (37 + 23 + 4 + 13 + 14)/5 = 18,2 ans
Âge moyen du décès : (76 + 78 + 29 + 63 + 31) = 55,4 ans

Il faut reconnaître qu'une bonne compréhension des échelles de mesure s'avère ici fort utile. Puisque les données figurant dans les trois premières colonnes du tableau appartiennent à une échelle d'intervalle, il était évidemment impossible d'en tirer des moyennes. Par contre, les trois dernières colonnes du tableau — que nous avons déduites des trois premières — appartiennent à une échelle de rapport. C'est pourquoi nous avons été en mesure de calculer les trois moyennes correspondantes.

Voici maintenant la preuve que les méthodes quantitatives requièrent avant tout un esprit *méthodique* plutôt que des connaissances poussées en mathématiques. Pour calculer les valeurs des trois dernières colonnes du tableau, nous avons dû tenir compte du fait que la vie et le règne de certains empereurs chevauchaient le premier millénaire avant Jésus-Christ et le premier millénaire après Jésus-Christ (il va de soi que les Romains utilisaient une échelle d'intervalle dont le point de départ était différent). Or, le premier millénaire après Jésus-Christ a de toute évidence commencé en l'an 1* et non en l'an 0, qui n'a d'ailleurs jamais existé. Ainsi, pour calculer la durée du règne d'Auguste, nous ne pouvons pas simplement additionner 24 (années avant J.-C.) et 14 (années après J.-C.). En réalité, il s'est écoulé 37 ans — et non 38 — entre l'an –24 et l'an 14.

Notons que le problème aurait été évité si nous avions utilisé le calendrier romain, qui part de la fondation de Rome (l'an 753 av. J.-C. correspondant à l'an 1 de l'ère romaine). Selon le calendrier romain, Auguste devint empereur en 730 et mourut en 767, ce qui donne bien 37 ans de règne : ouf, le compte est bon!

Si le premier millénaire a commencé en l'an 1, il est clair que le troisième millénaire a commencé en l'an 2001, et non en l'an 2000, n'en déplaise aux « millénaristes », qui considèrent plus les chiffres comme des objets de culte que comme une source de connaissance.

Cela dit, une telle erreur de méthode ne porterait pas tellement à conséquence ici, puisque nous utilisons des chiffres arrondis à l'année près (les dates considérées peuvent se situer entre le 1er janvier et le 31 décembre d'une année), et que le phénomène étudié n'exige pas un degré de précision extrême. Par contre, dans d'autres circonstances (dossier criminel, médical ou bancaire), une erreur de ce genre serait inacceptable. Il est donc important de bien porter attention au point de départ de l'échelle lorsque celle-ci est une échelle d'intervalle.

1.3. La moyenne et les fréquences

Combien d'individus compte, en moyenne, un ménage montréalais? Il n'y eut que 5 empereurs dans la dynastie des Césars, mais on a dénombré 756 000 ménages dans la communauté urbaine de Montréal en 1993. Le principe de la moyenne demeure toujours le même (la somme des valeurs divisée par le nombre de valeurs), mais la masse imposante des données nous oblige ici à trouver une méthode plus efficace.

On ne va quand même pas additionner la valeur (en nombre de personnes) des 756 000 ménages et diviser par 756 000 pour connaître la moyenne!

Dans le tableau 3.2 ci-après (tiré d'un exemple du chapitre précédent), la variable est la *taille du ménage*. Cette variable peut prendre diverses valeurs déterminées par le nombre d'individus par ménage (colonne 1). Étant donné que les ménages de plus de 4 personnes sont rares, notre tableau met dans le même paquet tous les ménages comptant 4 personnes et plus. Pour simplifier, nous estimerons que les ménages de 4 personnes et plus comptent en moyenne 4,5 personnes. La colonne 2 contient le nombre de ménages dénombrés (la fréquence, en milliers) pour chaque valeur possible. La colonne 3 reprend la colonne 2 sous forme de proportions (les fréquences relatives).

TABLEAU 3.2 - Répartition des ménages familiaux à Montréal selon le nombre de membres

	Fréquence (nombre de ménages en milliers)	**Fréquence relative**	**Fréquence relative cumulée**
[1]	[2]	[3]	[4]
1 personne	255	33,7	**33,7**
2 personnes	243	32,1	**65,9**
3 personnes	118	15,6	**81,5**
4 personnes ou plus	140	18,5	**100,0**
Total	**756**	**100,0**	

Source : Statistique Canada, Cansim. Données de 1993.

Note : Il s'agit de la Communauté urbaine de Montréal.

Pour obtenir la moyenne, nous pourrions aligner 756 chiffres (255 chiffres *1*, suivis de 243 chiffres *2*, etc.), les additionner et diviser le tout par 756. Nous pourrions encore, ce qui reviendrait au même, effectuer l'opération suivante :

$$[(255 \times 1) + (243 \times 2) + (118 \times 3) + (140 \times 4{,}5)]/756 = 1725/756 = 2{,}28$$

On notera que nous avons attribué, comme convenu, 4,5 personnes aux 140 ménages de 4 personnes et plus. On constatera également que la somme des fréquences (255 + 243 + 118 + 140) est égale au dénominateur (756). En fin de compte, le ménage montréalais moyen compte 2,28 individus.

$$\boxed{\text{Moyenne} = [(\text{Fréquence 1} \times \text{Valeur 1}) + (\text{Fréquence 2 x Valeur 2}) + \ldots]/\text{Fréquence totale}}$$

Cette manière de calculer la moyenne n'est qu'une généralisation de la formule précédente (dans laquelle les fréquences de chaque élément sont toutes égales à 1). Avez-vous remarqué, par ailleurs, que les chiffres du tableau 3.2 représentent en réalité des *milliers* de ménages? Il y a en effet 756 000 ménages à Montréal et non 756. Cette particularité ne change rien au résultat, puisque les proportions restent les mêmes. C'est un peu comme si on mesurait le poids moyen des poulets en kilogrammes plutôt qu'en grammes : ça ne rendrait pas les poulets plus légers ni plus lourds.

1.4. La moyenne pondérée : des poids relatifs

Dans l'exemple qui précède (la taille moyenne des ménages), nous aurions pu obtenir le même résultat en utilisant les fréquences relatives. Ce qui compte, après tout, c'est de respecter les proportions de chaque catégorie.

Si on utilise ces proportions sous leur forme décimale, le calcul de la moyenne se fait de la façon suivante :

(0,337 × 1) + (0,321 × 2) + (0,156 × 3) + (0,186 × 4,5) = 2,28 personnes par ménage

Moyenne = (Pondération 1 × Valeur 1) + (Pondération 2 × Valeur 2) + …

Si on préfère utiliser les proportions sous forme de pourcentage, il faudra diviser le résultat final par 100 pour obtenir la moyenne : n'oublions pas que 100 % = 100/100 = 1.

[(33,7 × 1) + (32,1 × 2) + (15,6 × 3) + (18,5 × 4,5)]/100 = 2,28 personnes par ménage

Les fréquences relatives représentent le poids relatif (ou *pondération*) de chaque composante. La somme de ces pondérations est bien sûr égale à 100 % ou 1.

33,7 % + 32,1 % + 15,6 % + 18,5 % = 99,9 % = 100 % = 1
0,337 + 0,321 + 0,156 + 0,185 = 0,999 = 1

Lorsque les pondérations sont connues, il n'est même plus nécessaire de connaître les diverses fréquences (ou effectifs) pour calculer la moyenne.

Parfois, les pondérations sont déterminées à partir d'autres critères que les fréquences. La moyenne des notes accumulées par un étudiant dans un cours est souvent une moyenne pondérée : tel exercice peut compter pour 10 % de la note finale, tel examen pour 30 %, etc. L'enseignant peut fixer ces pondérations en fonction de l'importance qu'il accorde personnellement à chaque évaluation, pourvu que la somme des pondérations soit égale à 100 %.

1.5. Les limites de la moyenne

Peut-être avez-vous déjà entendu l'expression « la loi de la moyenne » dans la bouche d'un mordu du casino : « Le numéro 13 n'est quasiment pas sorti hier. Aujourd'hui, il devrait se rattraper. J'en suis sûr, mon vieux, c'est la loi de la moyenne. Mise tout sur le 13! ».

Devant la majesté de la loi, l'homme n'a plus qu'à s'incliner, n'est-ce pas? C'est oublier que la moyenne est une mesure qui *traduit* la réalité, elle ne la *commande* pas. Le hasard n'a pas de mémoire, et la soi-disant « loi de la moyenne » n'est qu'une superstition héritée de nos croyances primitives.

Même si la moyenne est un concept relativement simple, il arrive parfois qu'on l'utilise de façon abusive, et pas seulement pendant les parties de roulette ou les performances sportives. C'est pourquoi nous allons mentionner quelques erreurs typiques liées au calcul ou à l'utilisation des moyennes.

Erreur de calcul : la moyenne des moyennes

Alors que chaque Américain consomme *en moyenne* l'équivalent de 6794 kg de pétrole sous forme d'énergie diverse par an, le Mexicain n'en consomme en moyenne que 1588 kg, et le Canadien, 7270 kg*. Quelle est la moyenne d'énergie consommée en Amérique du Nord?

On ne peut additionner ici les 3 chiffres et diviser le total par 3 (cela donnerait [6784 + 1588 + 7270]/3 = 5217 kg). Il faut tenir compte de la population de chaque pays (314 millions pour les États-Unis, 121 millions pour le Mexique et 35 millions pour le Canada, soit une population totale de 470 millions).

Source : Banque mondiale, *Indicateurs de développement dans le monde*. Données de 2012. À titre de comparaison, la consommation par habitant était, en 1992, de 7662 kg aux États-Unis, 1525 au Mexique, et 7912 au Canada. Les populations respectives de ces trois pays étaient de 255, 85 et 27 millions. La consommation moyenne d'énergie pour l'ensemble de l'Amérique du Nord était donc d'environ 6260 kg d'équivalent pétrole.

Dans ce cas, la véritable moyenne serait de :

[(6794 × 314) + (1588 × 121) + (7270 × 35)]/470
= [2 133 316 + 191 148 + 254 450]/470
= 2 579 914059/470 = 5489 kg d'équivalent pétrole consommé en moyenne par habitant

Moyenne = [Consommation totale des États-Unis + Consommation totale du Mexique + Consommation totale du Canada]/Population totale de l'Amérique du Nord

Erreur d'utilisation

Il ne suffit pas de faire de bons calculs. Il faut encore que la moyenne veuille dire quelque chose.

Saviez-vous que l'on parle 7 langues locales différentes en Italie : l'italien, bien sûr, mais aussi l'occitan (dans le Piémont), le français (dans le Val d'Aoste), l'allemand (dans le Tyrol), le frioulan (en Vénétie), le slovène (en Vénétie aussi) et le sarde (en Sardaigne). Tout cela sans compter les multiples dialectes de l'italien aux différences beaucoup plus grandes qu'entre, par exemple, le français parlé au Québec ou à Paris. Étant donné que l'Italie comptait 57,8 millions d'habitants en 1992, on pourrait en déduire que chacune des 7 langues parlées en Italie comptait en moyenne 8 257 000 locuteurs et des poussières (soit 57,8/7).

Le résultat précédent ne présente non seulement aucun intérêt, mais il est même trompeur. En réalité, 56 millions d'Italiens parlent l'italien standard ou un de ses dialectes. Le reste (1,8 million) se partage les 6 autres langues, soit en moyenne 0,3 million ou 300 000 individus par langue (car 1,8/6 = 0,3). On peut donc dire que les langues secondaires comptent en moyenne 300 000 locuteurs : ça au moins, c'est une moyenne qui veut dire quelque chose.

Avant de faire une moyenne, il faut d'abord déterminer les éléments qui devront être considérés. Des valeurs extrêmes viennent parfois « fausser » des moyennes. Par contre, plus le nombre d'éléments est grand et moins les éléments extrêmes risquent de brouiller les cartes. De plus, en interprétant une moyenne, il faut savoir avec précision qui y est inclus. Ainsi, la moyenne des notes d'un groupe d'étudiants contient-elle uniquement les notes de ceux qui ont suivi le cours jusqu'au bout? Les notes de ceux qui ont abandonné et de ceux qui ne se sont jamais présentés au cours sont-elles incluses? Si oui, comment sont calculées ces notes? Voilà bien des points à éclaircir avant de se lancer dans de savantes comparaisons.

En bref : Comment calculer une moyenne

1) Quand on connait toutes les valeurs

$$\text{Moyenne} = \frac{\text{Somme des valeurs}}{\text{Nombre de valeurs}}$$

Nombre d'élèves dans chaque groupe
36
36
32
32
32
24
22
20

Exemple : Calcul du nombre moyen d'élèves par groupe dans une école primaire.

$$\frac{36 + 36 + 32 + 32 + 32 + 24 + 22 + 20}{8} = \frac{234}{8} = 29{,}25$$

2) Quand on connait déjà la somme des valeurs et qu'on a le nombre de valeurs

On applique la même formule que dans le cas précédent.

Ici, un total de 234 élèves est distribué en 8 groupes : $\frac{234}{8} = 29{,}25$

3) Quand on connait la fréquence de chacune des valeurs

$$\text{Moyenne} = \frac{(\text{Fréquence 1} \times \text{Valeur 1}) + (\text{Fréquence 2} \times \text{Valeur 2}) + \ldots}{\text{Somme des fréquences}}$$

Valeur (nombre d'élèves)	**Fréquence** (nombre de groupes)
36	2
32	3
24	1
22	1
20	1
	8

$$\frac{(2 \text{ x } 36) + (3 \text{ x } 32) + (1 \text{ x } 24) + (1 \text{ x } 22) + (1 \text{ x } 20)}{8} = \frac{72 + 96 + 24 + 22 + 20}{8}$$

$$= \frac{234}{8} = 29{,}25$$

4) Quand on connait la pondération associée à chaque valeur

$$\text{Moyenne} = (\text{Pondération 1} \times \text{Valeur 1}) + (\text{Pondération 2} \times \text{Valeur 2}) + \ldots$$

Valeur	**Pondération**
36	0,250
32	0,375
24	0,125
22	0,125
20	0,125
	1,000

$$(0{,}25 \text{ x } 36) + (0{,}375 \text{ x } 32) + (0{,}125 \text{ x } 24) + (0{,}125 \text{ x } 22) + (0{,}125 \text{ x } 20)$$

$$= 9 + 12 + 3 + 2{,}75 + 2{,}5 = 29{,}25$$

Notes : Les pondérations peuvent aussi être exprimées en pourcentages (0,25 n'est autre que 25 % ou 25/100). On note que la somme des pondérations est toujours égale à 1 (ou 100 %). Chaque pondération représente la fréquence relative associée à chaque valeur (exemple : 2 classes sur un total de 8 contiennent 36 élèves, soit 2/8, ou 0,25 ou 25 %).

EXERCICES 1

1. 1. Le jour J

Le 6 juin 1944. 745 navires, groupés en 38 convois, approchent des côtes normandes. En tout, 4006 péniches de débarquement sont mises à l'eau, au large des plages, et 185 000 hommes débarquent sur le territoire français. En même temps, 1087 avions de largage, protégés par 13 175 avions de combat, lancent 18 000 parachutistes sur le territoire normand. (Source : Livre Guinness des records, 1995.)

a) Combien y avait-il d'hommes, en moyenne, sur chaque navire?

b) Combien de navires comptaient les convois, en moyenne?

c) Combien y avait-il de parachutistes, en moyenne, dans chaque avion de largage?

d) Chaque avion de largage était protégé par combien d'avions de combat, en moyenne?

2. Si j'avais un char… ou un ordi

a) Dans le cahier automobile de *la Presse de Montréal* du 17 juillet 1995, six concessionnaires accompagnaient leur publicité du prix total (au comptant) des véhicules proposés (tableau 3.3). Quel est le prix moyen des véhicules proposés aux lecteurs de *la Presse*?

TABLEAU 3.3 - Le prix des automobiles annoncées par six concessionnaires

Modèle	Marque	Prix (en $)
Civic	Honda	14 995
Integra RS	Acura	17 995
Protégé S	Mazda	11 995
Precidia	Mazda	14 495
Civic Special Edition	Honda	14 495
940	Volvo	26 900
Precidia RS	Mazda	14 995
Protégé S	Mazda	11 995
626 Cronos	Mazda	16 995

Source : Cahier automobile de La Presse, 17 juillet 1995.

b) En 2011, année qui marque un premier apogée dans le marché des micro-ordinateurs, le nombre d'unités vendues dans le monde s'élevait à 355,2 millions, pour un chiffre d'affaires total de 329 milliards de $US. Aux États-Unis seulement, les données étaient respectivement de 95,4 millions d'unités et de 85,5 milliards de $US (source : Gartner).

Montrez que le prix moyen d'un micro-ordinateur était alors moins élevé aux États-Unis que dans le reste du monde.

3. La moyenne et les fréquences

a) À partir des chiffres du tableau 3.4, calculez le taux d'intérêt moyen des banques canadiennes.

TABLEAU 3.4 - Les taux d'intérêt sur les dépôts à court terme proposés par les 12 principales banques canadiennes

Taux proposé (en %)	Fréquence (nombre de banques)	Fréquence relative
5,750	2	16,67
6,000	4	33,33
6,125	1	8,33
6,250	3	25,00
6,500	1	8,33
6,600	1	8,33
Total	**12**	**100,0**

Source : Les Affaires, 15 avril 1995.

Note : Dépôts de 30 à 59 jours.

b) Calculez la moyenne des notes de chacun des étudiants du tableau 3.5 (chiffres fictifs).

TABLEAU 3.5 - Les résultats des étudiants

Épreuve	Pondération (en %)	Notes de Julie	Notes de Julot
Exercice 1	10	40	80
Examen 1	20	50	80
Exercice 2	10	60	40
Examen 2	20	70	40
Examen final	40	80	60

2. L’ÉCART TYPE : MESURER LA DISPERSION DES DONNÉES

La moyenne est un outil pratique. En un seul chiffre, elle peut résumer plusieurs dizaines, centaines, voire plusieurs millions de données. Cependant, comme tout résumé, la moyenne néglige les détails. Souvent, ces détails présentent peu d’intérêt, mais parfois ils sont essentiels. On constate par exemple que la température moyenne estivale est de 26 degrés à Montréal (moyenne des maximums de juillet) contre 31 degrés à San Juan de Porto Rico. Et pourtant, il arrive souvent que les Montréalais aient à subir des températures plus élevées que les Portoricains. C’est que le thermomètre, très stable dans les Antilles, n’hésite pas à se balader d’un extrême à l’autre au Québec. Certains jours d’été, Montréal bat même le record de chaleur de toute l’Amérique du Nord.

La même moyenne peut recouvrir des données très proches ou encore des données très éparpillées. On peut imaginer, ainsi, deux groupes d’étudiants dont les moyennes sont égales à 67 %. Dans un cas, la plupart des étudiants ont des notes qui varient de 60 à 75 alors que dans l’autre, les résultats se répartissent à peu près uniformément entre 35 et 95. Ici, nos deux moyennes résument des réalités complètement différentes. Dans cette section, nous verrons comment mesurer ce degré de *dispersion* des données.

2.1. Les Beatles : plus que des musiciens

Qui a eu la plus grande influence dans le domaine de la musique populaire depuis la fin de la Deuxième Guerre mondiale? La question étant plutôt subjective, elle pourrait être un bon sujet d’enquête. En attendant, nous avons personnellement couronné les Beatles pour plusieurs raisons. Nés juste un peu avant les baby-boomers et arrivés à maturité dans les années 1960, ils font non seulement des ravages dans les palmarès, mais ils deviennent les idoles de toute une génération, influencent (ou reflètent) les mœurs des jeunes, inspirent même un grand nombre des musiciens populaires qui vont leur succéder. Leur carrière effective ne dure pourtant que 5 ans (de 1964, année de leurs premiers triomphes internationaux à 1969, année de la dissolution officieuse du groupe). Toutefois, pendant ces 5 ans leur musique, et le monde, semblent se transformer à toute vitesse.

Mais revenons à nos moutons. Nous avons fait l’inventaire des 12 albums originaux des Beatles (en excluant les compilations) et du nombre de titres qu’ils contiennent (tableau 3.6). Notre but ici est de mesurer à quel point le nombre de titres varie d’un album à l’autre. Une fois que nous aurons construit notre indicateur de dispersion (ici : l’*écart type*), nous pourrons l’appliquer à d’autres situations.

TABLEAU 3.6 - Le nombre de titres sur les albums originaux des Beatles						
	Album	**Année**	**Nombre de titres**	**Écart**	**Écart au carré**	**Écart absolu**
	[1]	[2]	[3]	[4]	[5]	[6]
1	*Please Please Me*	1963	14	0	0	0
2	*With the Beatles*	1963	14	0	0	0
3	*A Hard Day's Night*	1964	13	-1	1	1
4	*Beatles for Sale*	1964	14	0	0	0
5	*Help*	1965	14	0	0	0
6	*Rubber Soul*	1965	14	0	0	0
7	*Revolver*	1966	14	0	0	0
8	*Sgt. Pepper*	1967	13	-1	1	1
9	Album blanc 1	1968	17	3	9	3
10	Album blanc 2	1968	13	-1	1	1
11	*Abbey Road*	1969	16	2	4	2
12	*Let it Be*	1970	12	-2	4	2
	Total		**168**	**0**	**20**	**10**
	Nombre	12	(nombre total d'albums)			
	Moyenne	14	(nombre moyen de titres par album)			
	Variance	1,67				
	Écart-type	1,29				

Source : collection personnelle.

Les 12 albums contiennent en tout 168 titres. Il y a donc 14 titres en moyenne par album (car 168/12 = 14). Cette moyenne va nous servir de point de repère pour évaluer la dispersion. Y a-t-il beaucoup d'albums dont le nombre de titres s'écarte de cette moyenne et, si oui, s'en écartent-ils beaucoup? À première vue, l'écart est plutôt faible (colonne 4 du tableau 3.6). Certains écarts sont positifs, d'autres sont négatifs. La somme des écarts est toujours nulle, puisque la moyenne représente en quelque sorte le *centre de gravité* de tous les éléments.

L'écart absolu mesure l'écart entre une valeur et la moyenne, sans considérer le fait que cet écart soit positif ou négatif.

Pour additionner les écarts sans qu'ils s'annulent, il faut donc se débarrasser de leur signe. C'est ce que nous avons fait dans la colonne 6. En termes absolus, la somme des écarts est de 10 (en bas de la colonne 6). Comme le calcul porte sur 12 éléments (les 12 albums), on peut dire qu'en moyenne, l'écart absolu est de 0,83 (soit 10/12). Ce simple chiffre suffirait pour nous indiquer que le nombre de titres varie peu d'un album à l'autre.

Pour calculer l'écart absolu moyen

- 1. Calculez la moyenne :
  168/12 = 14
- 2. Mesurez les écarts entre chaque valeur et la moyenne :
  pour les deux premières valeurs : 14 – 14 = 0; pour la troisième valeur : 13 – 14 = –1; etc.

- 3. Mettez ces écarts en valeur absolue et additionnez-les :
  0 + 0 + 1 + 0 + 0 + 0 + 0 + 1 + 3 + 1 + 2 + 2 = 10
- 4. Divisez la somme des écarts absolus par le nombre de valeurs :
  10/12 = 0,83

L'écart type est un indice de dispersion des valeurs autour d'une moyenne. 7

En pratique, l'écart absolu est peu utilisé. On lui préfère l'écart type qui possède des propriétés plus intéressantes et qui n'est pas tellement plus compliqué à calculer. Dans la colonne 5 du tableau 3.6, on s'est débarrassé des signes négatifs en mettant au carré tous les écarts à la moyenne. La somme des écarts mis au carré est égale à 20 (au bas de la colonne 5). Pour terminer, et puisque nous avons mis les écarts au carré au point de départ, nous allons faire la racine carrée du résultat obtenu : l'écart type est égal à la racine carrée de 1,66, soit 1,29. Ouf! Cela semble compliqué? Pas du tout : regardez ce qui suit.

Pour calculer l'écart type

- 1. Calculez la moyenne :
  168/12 = 14
- 2. Mesurez l'écart entre chaque valeur et la moyenne
- 3. Mettez chacun de ces écarts au carré
- 4. Additionnez ces écarts au carré :
  0 + 0 + 1 + 0 + 0 + 0 + 0 + 1 + 9 + 1 + 4 + 4 = 20
- 5. Divisez le résultat par le nombre de données :
  20/12 = 1,66
- 6. Faites la racine carrée du résultat :
  $\sqrt{1{,}66} = 1{,}29$

Les données sur lesquelles nous venons de travailler avaient le mérite d'être simples. Elles nous permettaient d'illustrer facilement la méthode de calcul de l'écart type. Elles n'étaient cependant pas d'une grande utilité au niveau de l'interprétation. Il suffisait en effet de jeter un coup d'œil rapide sur la colonne 3 du tableau 3.6 pour constater que le nombre de titres variait peu d'un album à l'autre. Nous allons donc, sans quitter les Beatles, faire appel à un exemple plus significatif.

La question que nous posons est la suivante : la durée des chansons des Beatles a-t-elle évolué avec le temps? Nous avons choisi, pour y répondre, deux albums typiques des Beatles (tableau 3.7) et nous nous pencherons davantage sur l'analyse des résultats que sur la méthode de calcul. L'album *Help* se situe plus vers le début de la carrière des Beatles (la chanson *Yesterday* se trouve dessus et elle dure 2 minutes et 4 secondes) et *Abbey Road* est le dernier album fait en commun par le groupe (la chanson vedette, *Here Comes the Sun* dure 3 minutes et 4 secondes).

Cette fois, nous avons utilisé notre chiffrier électronique pour calculer les moyennes et les écarts types. Observons les résultats du tableau 3.7, puis essayons de les expliquer. En moyenne, les chansons plus récentes sont plus longues. Mais ce qui est encore plus frappant, c'est que la durée des chansons plus récentes devient de plus en plus variable (l'écart type a beaucoup augmenté). En

1965, les chansons étaient toutes sur le même modèle et duraient entre 2 et 3 minutes. En 1969, c'est beaucoup plus disparate : il y en a de très longues (7:49) et de très courtes (1:06).

Voici notre interprétation de cette évolution, avec tout ce qu'elle peut avoir de subjectif. En 1965, les chansons étaient faites avant tout pour *danser*, les paroles étaient simples, la musique était efficace, énergique et construite selon quelques modèles standards, le public était jeune et insouciant. Il fallait que ça « roule ». En 1969, l'auditoire des Beatles a changé : il est un peu plus vieux, il est plus sérieux, plus intellectuel, plus politisé; la musique est plus sophistiquée, la construction est plus élaborée et plus variée, les paroles sont plus recherchées; les chansons sont plutôt faites pour être *écoutées*.

TABLEAU 3.7 - Durée des chansons des Beatles selon l'album

	Help	**Abbey Road**
(en minutes:secondes)	1965	1969
	2:18	4:16
	1:42	2:59
	2:57	3:24
	2:27	3:28
	2:05	2:49
	2:16	7:49
	3:03	3:04
	2:28	2:45
	1:55	3:57
	2:33	2:31
	2:36	1:06
	2:03	1:13
	2:04	1:58
	2:53	1:31
		1:37
		2:04
Moyenne	**2:22**	**2:54**
Écart-type	**0:23**	**1:33**

Source : collection personnelle.

Proposons, pour conclure, une autre hypothèse plus terre à terre. Jusqu'au début des années 1950, les disques (78 tours) ne pouvaient contenir plus de 3 minutes de musique. Cela explique pourquoi les compositeurs écrivaient toujours des chansons de 2 ou 3 minutes. Et puis l'habitude est restée pendant près de quinze ans après la mort du 78 tours. Comme quoi, l'homme est parfois esclave de lui-même…

EXERCICES 2

1. Les empereurs romains : l'essayer c'est l'adopter

Quelques dizaines d'années après les Césars, Rome a eu la chance d'être gouvernée par des empereurs « sérieux » : les Antonins. Ces derniers avaient pris l'habitude de se choisir un successeur qualifié en l'adoptant. Un seul fit exception à la règle : Marc-Aurèle, surnommé le *meilleur des empereurs*, qui commit l'erreur de désigner son véritable fils comme héritier. Ce sinistre jeune homme, Commode de son prénom, un mélange de Néron et Caligula, sonna le glas de la dynastie des Antonins et périt assassiné.

TABLEAU 3.8 - Les empereurs romains

(Années)	**Naissance**	**Début du règne**	**Fin du règne et décès**
Les Césars			
Auguste	-63	-24	14
Tibère	-42	14	37
Caligula*	12	37	41
Claude*	-10	41	54
Néron*	37	54	68
Les Antonins			
Nerva	26	96	98
Trajan	53	98	117
Hadrien	76	117	138
Antonin	86	138	161
Marc-Aurèle	121	161	180
Commode*	161	180	192

Source : Petit Robert 2.

* Mort assassiné.

a) À partir des chiffres du tableau 3.8 ci-joint, calculez l'âge moyen des empereurs de la dynastie des Césars au moment de leur accession au trône et de leur mort. Faites les mêmes calculs pour la dynastie des Antonins.

b) Calculez la moyenne et l'écart type de la durée du règne pour les trois cas suivants : (1) la dynastie des Césars, (2) la dynastie des Antonins, (3) la dynastie des Antonins à l'exception de son fondateur Nerva.

c) Comparez les deux dynasties.

2. Chansons pour danser, chansons pour écouter

a) Vérifiez les moyennes et les écarts types du tableau 3.7.

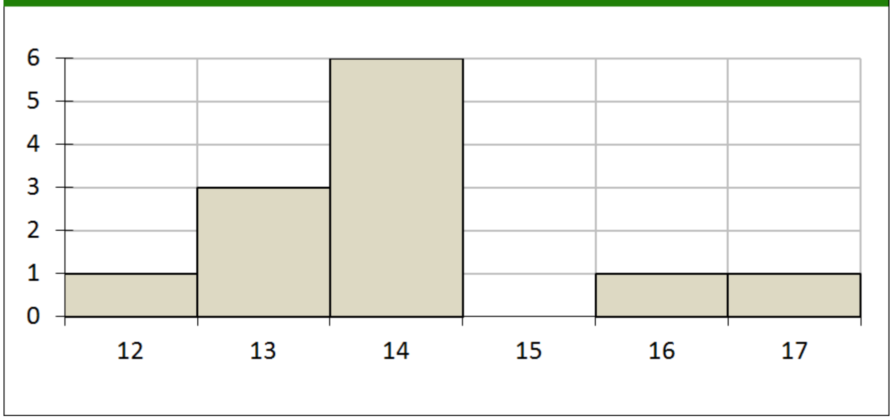
b) Recherche. Choisissez deux disques (avec leur durée) et créez un tableau semblable au tableau 3.7.

Note : vous pouvez remplacer les disques par des livres (avec la longueur de leurs chapitres) des listes de films proposés par des cinémas (avec la durée des films) ou par toute autre situation similaire.

3. LA COURBE NORMALE ET L'ÉCART TYPE

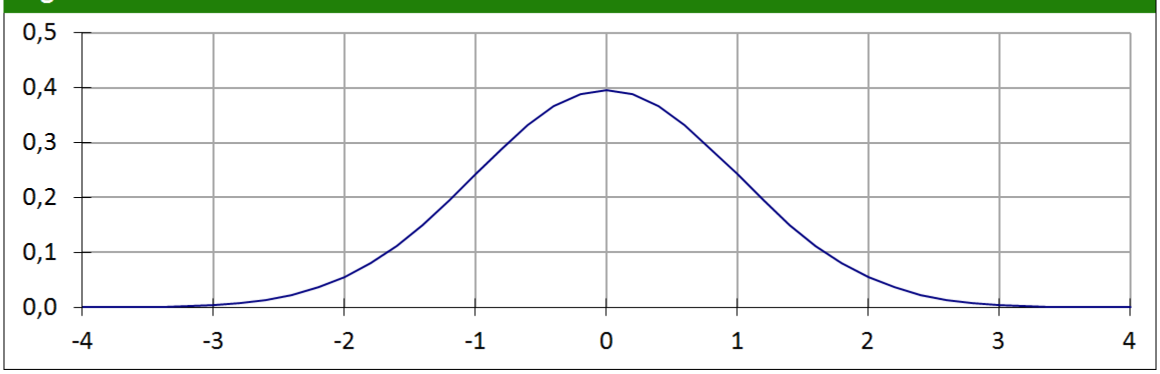
Reprenons les données du tableau 3.6 et représentons-les graphiquement sous forme d'un histogramme de fréquences (figure 3.1). Sur l'axe horizontal, nous indiquons les valeurs que peut prendre la variable *nombre de titres* et sur l'axe vertical nous indiquons le nombre de fois (la fréquence) que la variable prend telle ou telle valeur.

FIGURE 3.1 - Distribution des fréquences de la variable *nombre de titres* sur les albums des Beatles



Dans de nombreux cas, la distribution des valeurs que prend une variable suit une courbe en forme de cloche (ou de chapeau de Napoléon) : la courbe dite *normale* (figure 3.2). Lorsque la distribution est normale, le fait de connaître la moyenne et l'écart type permettent de trouver une foule d'informations.

Figure 3.2 - La distribution normale



3.1. Quand la distribution est-elle normale?

Jusqu'ici, nos exemples étaient relativement simples et les données (les valeurs prises par les variables) étaient peu nombreuses. Notre but était surtout d'illustrer les nouveaux concepts présentés : la *moyenne* et l'*écart type*. Dans la pratique, lorsque les données sont trop peu nombreuses, certaines d'entre elles (celles qui s'éloignent trop de la tendance générale) peuvent jouer les trouble-fête et « fausser » la moyenne et l'écart type. Lorsque les éléments sont très nombreux, les trouble-fête ont par contre moins de chance de se faire remarquer : ils sont noyés dans la masse.

La loi normale apparaît dans de très nombreux contextes.

Lorsque les données sont suffisamment nombreuses et qu'elles dépendent d'une multitude de facteurs dont aucun ne prédomine sur les autres, la distribution de ces données est souvent normale. Le nombre de clients à la banque le jeudi, le temps que nous passons sous la douche, la grosseur des œufs pondus par une poule X, l'épaisseur des anneaux d'un tronc d'arbre ont des bonnes chances de suivre la *distribution normale*. Par contre, la répartition de la population selon l'âge ne suit pas une courbe normale pas plus que la distribution des salaires dans une entreprise.

Même si le nombre de valeurs est peu élevé (il n'y a que 12 albums considérés), on sent déjà poindre la courbe normale dans la distribution du nombre de titres par album 33 tours chez les Beatles. Le nombre de titres est en effet lié à de multiples facteurs d'ordre commercial (l'attente des clients), artistique (l'inspiration des musiciens), matériel (la capacité d'un disque), économique (le coût de production), etc. À plus long terme cependant, il se peut qu'une variable (comme l'apparition du disque compact) vienne changer les règles du jeu de façon radicale et modifier complètement l'allure de la distribution.

La variable *taille des individus* possède les caractéristiques de la distribution normale. Plus on s'éloigne de la taille moyenne, que ce soit en montant ou en descendant, et moins on il y a d'individus. Nous allons utiliser cet exemple pour montrer comment, grâce à la courbe normale et deux simples chiffres (la moyenne et l'écart type), on peut répondre à un tas de questions intéressantes.

3.2. L'écart type et la courbe normale

Au Québec, la taille moyenne des hommes est de 176 cm. Évidemment, il y en a de plus grands et de plus petits. L'écart type, qui mesure cette dispersion, est de 7 cm. Pour les femmes, la moyenne est de 162 cm et l'écart type de 6 cm. Dans un cas comme celui-ci, les hommes diffèrent systématiquement des femmes et il serait maladroit de les mettre dans le même paquet. Notons par ailleurs que la taille moyenne a tendance à augmenter depuis un siècle.

> Pour rendre les explications un peu plus simples, nous avons un peu arrondi les chiffres. Nous nous sommes basés sur les résultats (non publiés) de l'Enquête nationale sur la santé de la population menée en 1994-95 par Statistique Canada, selon laquelle la taille moyenne des Canadiens adultes (18 ans et plus) est estimée à 176,15 cm pour les hommes et à 161,93 cm pour les femmes. En passant, et selon la même enquête, le poids moyen des Canadiens adultes est de 80,3 kg pour les hommes et de 64,6 kg pour les femmes. À titre de comparaison, on pourra consulter le tableau 3.11 qui contient la taille moyenne et le poids moyen des jeunes Australiennes. En ce qui concerne l'écart type de la taille des Canadiens, nous n'avons malheureusement pas été en mesure d'obtenir de données précises (et gratuites), malgré d'ardentes négociations avec les autorités. Nous avons donc dû procéder à nos propres estimations pour l'écart type, en nous basant sur des données américaines et des microdonnées ultérieures.

Les simples chiffres que nous venons de mentionner vont nous permettre de répondre aux questions suivantes :

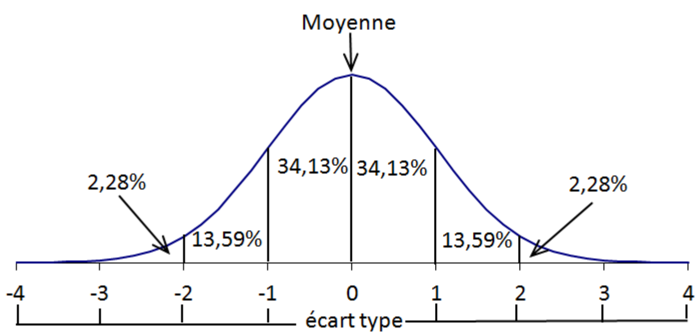
- Julie mesure 168 cm. Est-elle une grande fille?
- Julot mesure 169 cm. Est-il un petit gars?

- Le châssis d'une porte a 2,11 m de haut. Sur les 1000 personnes différentes qui franchiront cette porte au cours de l'année, y en a-t-il beaucoup qui risquent de se cogner la tête?

Il faut d'abord savoir que, dans une distribution normale, la proportion de sujets qui se trouve dans un intervalle donné est toujours la même. Pour être plus concret, disons que 34,13 % des sujets sont situés entre la moyenne (point central ou point 0 sur l'axe horizontal de la figure 3.3) et 1 écart type; 13,59 % des éléments sont situés entre 1 et 2 écarts types de la moyenne; le restant (2,28 %) est situé à plus de 2 écarts types de la moyenne. Il en va de même pour les sujets dont les valeurs sont inférieures à la moyenne (partie gauche de la courbe de la figure 3.3). La surface totale sous la courbe représente 100 % : 50 % à gauche et 50 % à droite de la moyenne.

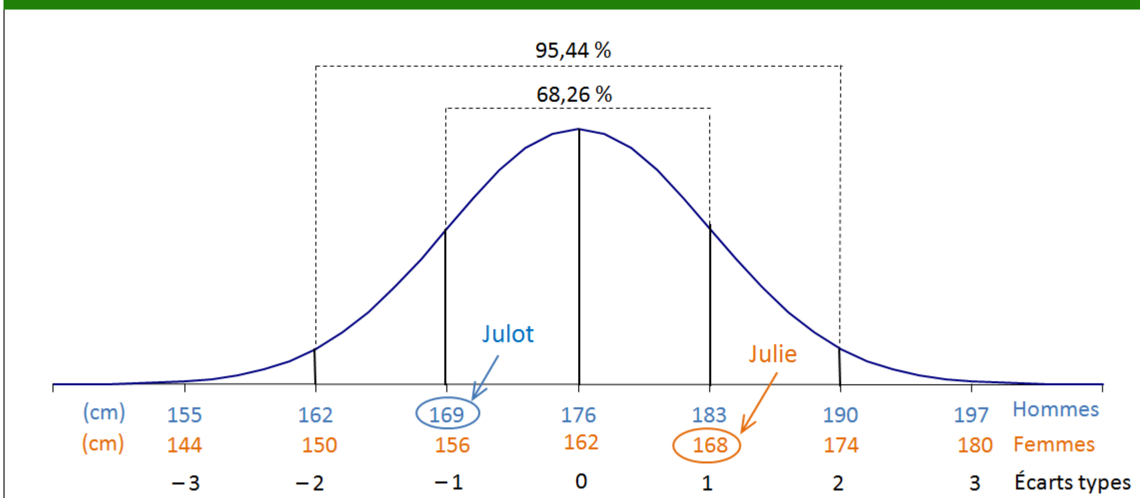
Figure 3.3 - La distribution normale et l'écart type

Nombre d'écart types	Aire correspondante sous la courbe entre 0 et l'écart type	
	(en unités)	(en %)
1	0,3413	34,13%
2	0,4772	47,72%
3	0,4986	49,86%
4	0,4999	49,99%
5	0,4999997	49,99997%



Pour revenir à nos Québécoises, on peut affirmer que Julie est parmi les plus grandes (voir la figure 3.4) puisque 84,13 % des Québécoises sont plus petites qu'elle : les 50 % plus petites que la moyenne et les 34,13 % comprises entre la moyenne et 168 cm. Il reste quand même encore 14,87 % de femmes qui sont plus grandes que Julie. Quant à Julot, il est dans la situation inverse : 84,13 % des Québécois sont plus grands que lui.

Figure 3.4 - La distribution normale et la taille des Québécois



Réglons enfin le cas de la porte de 2,11 cm. Utilisons l'écart type des hommes, qui sont le plus susceptibles de se cogner la tête. Cela nous met à 35 cm de la moyenne (211 cm – 176 cm = 35 cm), soit 5 écarts types (car 5 × 7 cm = 35 cm). Le tableau qui accompagne la figure 3.3 montre que seulement 0,00003 % des sujets (soit 50 % – 49,99997 %) s'éloignent à plus de 5 écarts types de la moyenne. Cette proportion de 0,00003 % équivaut à 3 personnes sur 10 millions. Il y a donc fort peu de chances pour qu'un tel individu, si tant est qu'il se présente, se cogne un jour au cadre de porte. À moins d'utiliser des talons particulièrement épais.

Le nombre d'écarts types qu'il y a entre une valeur particulière et la moyenne de toutes les valeurs est souvent appelé la *cote z*. La cote z de Julie est égale à +1 et celle de Julot est égale à –1, celle du monsieur qui mesure 2,11 mètres et qui se cogne dans les châssis de porte est égale à +5.

Cote z = (Valeur particulière – Moyenne)/Écart type

Cote z de Julie = (168 – 162)/6 = 6/6 = 1
Cote z de Julot = (169 – 176)/7 = –7 /7 = –1
Cote z du grand monsieur = (211 – 176)/7 = 35/7 = 5

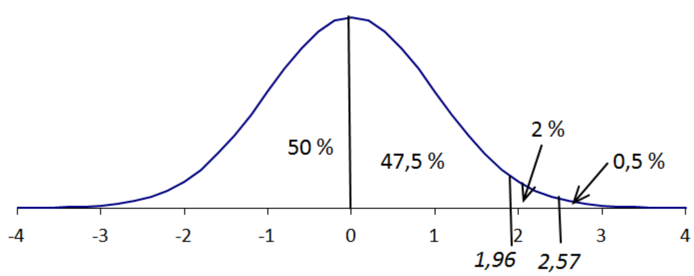
Posons maintenant les questions à l'envers :

Si nous excluons les 0,5 % de Québécois les plus petits et les 0,5 % de Québécois les plus grands, quelle est la taille maximum et quelle est la taille minimum des 99 % de Québécois restants? (La question gagne à être posée séparément pour les hommes et pour les femmes.)

La figure 3.5 reprend la distribution normale sous l'angle que nous venons de lui donner. La proportion de 0,5 % correspond à 2,57 écarts types de part et d'autre de la moyenne. Rappelons que la taille *moyenne* des hommes est de 176 cm et l'*écart type* correspondant, de 7 cm. Pour les hommes, 2,57 écarts types correspondent à 18 cm (2,57 × 7 cm). Il y a donc 99 % des hommes qui mesurent entre 158 cm (soit 176 – 18) et 194 cm (soit 176 + 18).

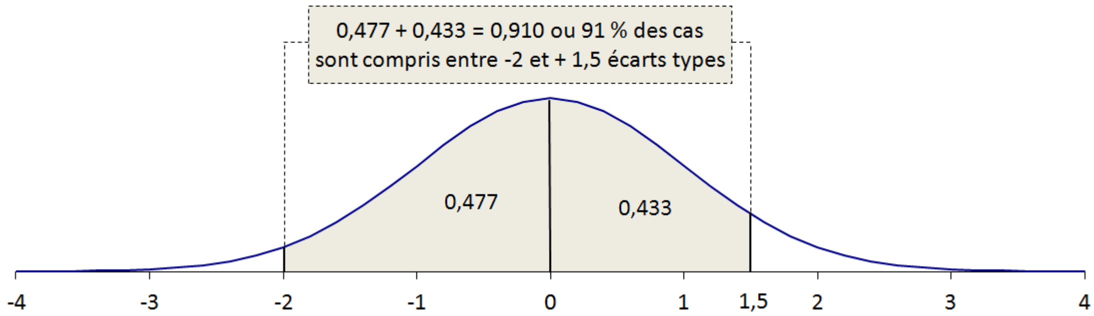
Figure 3.5 - Découpage de la distribution normale en chiffres ronds

Aire correspondante sous la courbe à droite de l'écart type (en unités)	Aire correspondante sous la courbe à droite de l'écart type (en %)	Nombre d'écart types
0,100	10,0	1,28
0,050	5,0	1,64
0,025	2,5	1,96
0,010	1,0	2,33
0,005	0,5	2,57



De la même manière, 99 % des femmes, dont la taille moyenne est de 162 cm avec un écart type de 6 cm, mesurent entre 147 cm (162 – [2,57 x 6]) et 177 cm (162 + [2,57 × 6]).

En bref : Comment déterminer la position sur une courbe normale



A) La valeur du sujet est connue : on cherche sa position dans la distribution

Exemple : Antonine mesure 175 cm. Comment se situe-t-elle par rapport aux autres adultes québécoises, dont la taille moyenne est de 162 cm avec un écart type de 6 cm?

1. Convertissez la distance entre la valeur du sujet et la moyenne en nombre d'écarts types. (Le résultat peut être positif ou négatif.)

Nombre d'écarts types (cote z) = (Sujet – Moyenne) / Écart type

(175 – 166)/6 = 9/6 = 1,5

2. Obtenez la valeur de l'aire correspondante en consultant la table de distribution normale ci-après (sans tenir compte du signe).

1,5 écart type correspond à une aire de 0,433 ou 43 %.

3. Ajoutez 0,5 (ou 50 %) au résultat.

Le chiffre obtenu représente la proportion de sujets inférieurs à notre sujet (si le résultat calculé en 1 était positif) ou supérieurs à notre sujet (dans le cas contraire).

0,433 + 0,5 = 0,933 ou 93,3 %
93,3 % des Québécoises sont plus petites qu'Antonine, et les autres (6,7 %) sont plus grandes qu'elle.

B) La position est connue : on cherche la valeur du sujet correspondant

Exemple : Césarine a une taille inférieure de deux écarts types à la moyenne québécoise.

1. Convertissez les écarts types en unités de mesure de la variable en question, et ajoutez ce résultat à la moyenne.

Valeur du sujet = Moyenne + (Nombre d'écarts types × Valeur de l'écart type)

166 cm – (2 × 6 cm) = 154 cm

C) On cherche la proportion du nombre de sujets compris entre 2 sujets donnés

Exemple : Quelle est la proportion de Québécoises plus petites qu'Antonine et plus grandes que Césarine?

1. Convertissez la valeur des 2 sujets en nombre d'écarts types.

2. Obtenez la valeur de l'aire correspondante en consultant la table de distribution normale.

3. Si les deux écarts types sont de signe opposé : additionnez les aires obtenues à l'étape précédente; s'ils sont de même signe, soustrayez la plus petite aire de la plus grosse.

Antonine : +1,5 écart type; Aire = 0,433
Césarine : –2 écarts types; Aire = 0,477
Résultat : 0,433 + 0,477 = 0,91 (ou 91 %)

Conseil : Faites un dessin de la courbe normale et découpez cette dernière en tranches. Vous pourrez mieux délimiter la section qui vous intéresse.

Table de distribution normale

Les valeurs en bleu représentent l'aire comprise entre la moyenne et l'écart-type correspondant.

Écart type	Aire	Écart type	Aire	Écart type	Aire	Écart type	Aire
0,00	0,000	1,00	0,341	2,00	0,477	3,00	0,499
0,05	0,020	1,05	0,353	2,05	0,480	3,05	0,499
0,10	0,040	1,10	0,364	2,10	0,482	3,10	0,499
0,15	0,060	1,15	0,375	2,15	0,484	3,15	0,499
0,20	0,079	1,20	0,385	2,20	0,486	3,20	0,499
0,25	0,099	1,25	0,394	2,25	0,488	3,25	0,499
0,30	0,118	1,30	0,403	2,30	0,489	3,30	0,500
0,35	0,137	1,35	0,411	2,35	0,491	3,35	0,500
0,40	0,155	1,40	0,419	2,40	0,492	3,40	0,500
0,45	0,174	1,45	0,426	2,45	0,493	3,45	0,500
0,50	0,191	1,50	0,433	2,50	0,494	3,50	0,500
0,55	0,209	1,55	0,439	2,55	0,495	3,55	0,500
0,60	0,226	1,60	0,445	2,60	0,495	3,60	0,500
0,65	0,242	1,65	0,451	2,65	0,496	3,65	0,500
0,70	0,258	1,70	0,455	2,70	0,497	3,70	0,500
0,75	0,273	1,75	0,460	2,75	0,497	3,75	0,500
0,80	0,288	1,80	0,464	2,80	0,497	3,80	0,500
0,85	0,302	1,85	0,468	2,85	0,498	3,85	0,500
0,90	0,316	1,90	0,471	2,90	0,498	3,90	0,500
0,95	0,329	1,95	0,474	2,95	0,498	3,95	0,500

Vous trouverez une table normale détaillée en annexe.

EXERCICES 3

1. La taille des Québécois

La taille des Québécois de sexe masculin est distribuée de façon normale avec une moyenne de 176 cm et un écart type de 7 cm.

a) Quelle est la proportion de Québécois mesurant plus de 183 cm?

b) Quelle est la proportion de Québécois mesurant plus de 185 cm?

c) Quelle est la proportion de Québécois mesurant moins de 162 cm?

d) Quelle est la proportion de Québécois mesurant entre 183 cm et 185 cm?

e) César-Auguste Tremblay mesure 2 écarts types de moins que la moyenne de ses concitoyens. Quelle est sa taille?

2. Cote z

On suppose, dans l'exemple fictif suivant, que les notes sont distribuées de façon normale.

a) Les 150 élèves du professeur X ont obtenu une note moyenne de 64 points à l'examen final (avec un écart type de 12 points). Julie, qui est également une élève du professeur X a obtenu pour sa part la note de 88 points. Calculez la cote z de Julie.

b) Pour les élèves du professeur Y, la moyenne est de 60 points et l'écart type de 10. Julot, élève du professeur Y, a obtenu une note de 85 points. Calculez la cote z de Julot.

c) Julie et Julot ont tous deux posé leur candidature à l'université. Le directeur de l'établissement où ces examens se sont déroulés recommande en premier lieu la candidature de Julot, estimant que ce dernier a obtenu de meilleurs résultats que Julie. Que pensez-vous de l'attitude de ce directeur?

3. Courbe normale et proportions

Les questions suivantes portent sur une variable distribuée de façon normale.

a) Quelle est la proportion de valeurs situées entre 0 et 1 écart type?

b) Quelle est la proportion de valeurs situées entre 0 et 2 écarts types?

c) Quelle est la proportion de valeurs situées au-delà de 3 écarts types?

d) Quelle est la proportion de valeurs situées entre –1 et 1 écart type?

e) Quelle est la proportion de valeurs situées entre –1 et 2 écarts types?

4. D'AUTRES INDICATEURS DE MOYENNE ET DE DISPERSION

En dehors de la moyenne et de l'écart type, il existe quelques mesures pratiques et simples pour évaluer le niveau de dispersion des éléments d'une distribution.

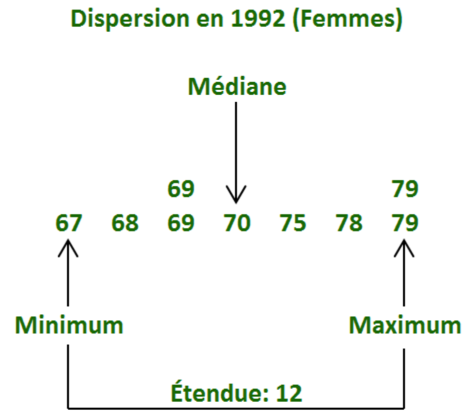
4.1. Le milieu et les bords : autres mesures

La figure 3.6 nous servira à illustrer ces mesures. Nous avons retenu dans ce tableau les pays de langue espagnole de l'Amérique centrale et des Antilles : ce sont des pays d'importance similaire (petits pays en superficie et en population) et homogènes (même région, même culture). Tout au long du commentaire qui va suivre, nous utiliserons les chiffres de la première colonne (les femmes en 1992).

Figure 3.6 - Espérance de vie à la naissance (en années)

Pays hispanophones d'Amérique centrale

	Femmes 1992	**Hommes 1992**	**Femmes 2012**	**Hommes 2012**
Guatemala	67	62	75,3	68,2
Honduras	68	64	76,0	71,2
El Salvador	69	64	76,9	67,5
Nicaragua	69	65	77,6	71,5
R. Dominicaine	70	65	76,5	70,2
Panama	75	71	80,3	74,6
Porto Rico	78	71	82,4	74,9
Cuba	79	73	81,1	77,1
Costa Rica	79	74	82,0	77,5
Moyenne	**72,7**	**67,7**	**78,7**	**72,5**
Écart-type	**4,7**	**4,3**	**2,6**	**3,5**
Coefficient de variation (en %)	**6,5**	**6,3**	**3,3**	**4,8**



Source : Banque mondiale, IDM.

L'étendue mesure l'écart entre les deux valeurs extrêmes d'une distribution.

Les *valeurs extrêmes* sont le *minimum* (ici 67) et le *maximum* (ici 79). L'intervalle dans lequel la variable prend ses valeurs, ou étendue, est la différence entre les deux extrêmes (ici 79 – 67 = 12).

La médiane est la valeur autour de laquelle on retrouve deux groupes représentant chacun la moitié des données d'une distribution.

La médiane représente la valeur de l'élément du milieu de la distribution. Il faut évidemment que les valeurs soient, au préalable, classées par ordre de grandeur. Il y a ici 9 valeurs. La valeur de l'élément du milieu (le cinquième élément) est de 70 : c'est la médiane de cette distribution. Il existe autant de valeurs inférieures à 70 que de valeurs supérieures à 70 dans notre distribution. La médiane permet donc de couper la distribution en 2 parties égales.

Dans une certaine mesure, la médiane nous indique le milieu, mais elle n'est pas nécessairement égale à la moyenne. Dans notre exemple, la médiane est plutôt inférieure à la moyenne (qui est de 72,7 : vérifiez-le si le cœur vous en dit). Dans une distribution normale, par contre, la médiane coïncide avec la moyenne. Enfin, lorsque le nombre de valeurs est pair (ici, il était impair), on coupe la poire en deux : la médiane est alors égale à la moyenne des 2 valeurs centrales.

Au Québec, l'âge médian de la population était de 25,6 ans en 1971. Cela signifie que la moitié des Québécois avaient moins de 25,6 ans. Vingt ans plus tard, cette médiane se situait à 34 ans. En 2012, au moment du « printemps érable », la médiane avait atteint 41,5 ans, et une bonne partie de la population se montrait peu sympathique aux revendications de la jeunesse. Par la suite, la médiane commença à plafonner : par définition, le « vieillissement de la population » ne peut pas être éternel!

Le mode est la (ou les) valeur(s) qui revient (qui reviennent) le plus souvent dans une distribution.

Le mode est la valeur qui revient le plus souvent dans la distribution. Ici, il y a deux modes : la valeur 69 et la valeur 79 reviennent 2 fois chacune. On dira que la distribution est bimodale (ou à deux bosses, comme un chameau). Avouons, pour nuancer, que notre distribution comporte peu d'éléments et qu'aucune valeur ne se détache vraiment nettement. Si on regroupait les éléments, on pourrait cependant observer nettement 2 catégories de pays : ceux qui tournent autour de 68 ou 69 et ceux qui tournent autour de 78 ou 79.

Le coefficient de variation mesure le rapport entre l'écart type et la moyenne.

Le coefficient de variation est une variante un peu plus raffinée de l'écart type. Dans le tableau 3.9, l'écart type est plus grand pour les femmes que pour les hommes (4,7 pour les femmes et 4,3 pour les hommes). Le coefficient de variation représente l'écart type relatif. Ici, le coefficient de variation est à peu près identique pour les femmes (4,7/72,7 = 0,065 = 6,5 %) que pour les hommes (4,3 / 67, 7 = 0,063 = 6,3 %). À vous déterminer si le coefficient de variation convient mieux à la situation étudiée que l'écart type. Car les méthodes quantitatives sont plus une question de jugement que de calcul.

Coefficient de variation = Écart type/Moyenne.

Coefficient de variation de l'espérance de vie des femmes = 4,7/72,7 = 0,065 = 6,5 %

Au fait, pourquoi les femmes vivent-elles généralement plus longtemps que les hommes? Posez la question à votre entourage et vous récolterez sans doute de beaux spécimens de préjugés. La cause des écarts serait tout banalement de nature chimique : le métabolisme des hommes est plus élevé que celui des femmes (d'où, entre autres, des besoins supérieurs en calories ingérées) et la « machine » s'userait plus vite.

4.2. Les quintiles : cinq parts égales?

Par quintile on entend chacune des quatre valeurs qui partagent une distribution (rangée en ordre) en cinq groupes de même effectif et, par extension, ces cinq groupes eux-mêmes.

En plus d'identifier le milieu, la médiane permettait de couper la distribution de valeurs en 2 paquets, ou *tranches*. Eh bien, il arrive aussi que l'on coupe la distribution en 5 tranches (voir le tableau 3.10 sur lequel figurent les revenus d'une population divisée en 5 tranches : les plus pauvres, les assez pauvres, ceux du milieu, les assez riches et les plus riches). On parle alors de quintiles (qui n'est autre que le mot « cinquième » en latin).

Dans le tableau 3.9, les quintiles sont utilisés de façon particulière pour observer comment les revenus se distribuent dans la population. Aux Pays-Bas, le premier quintile le plus pauvre (c'est à dire les 20 % des Néerlandais les plus pauvres) se partage 8,2 % du revenu national (voir le tableau

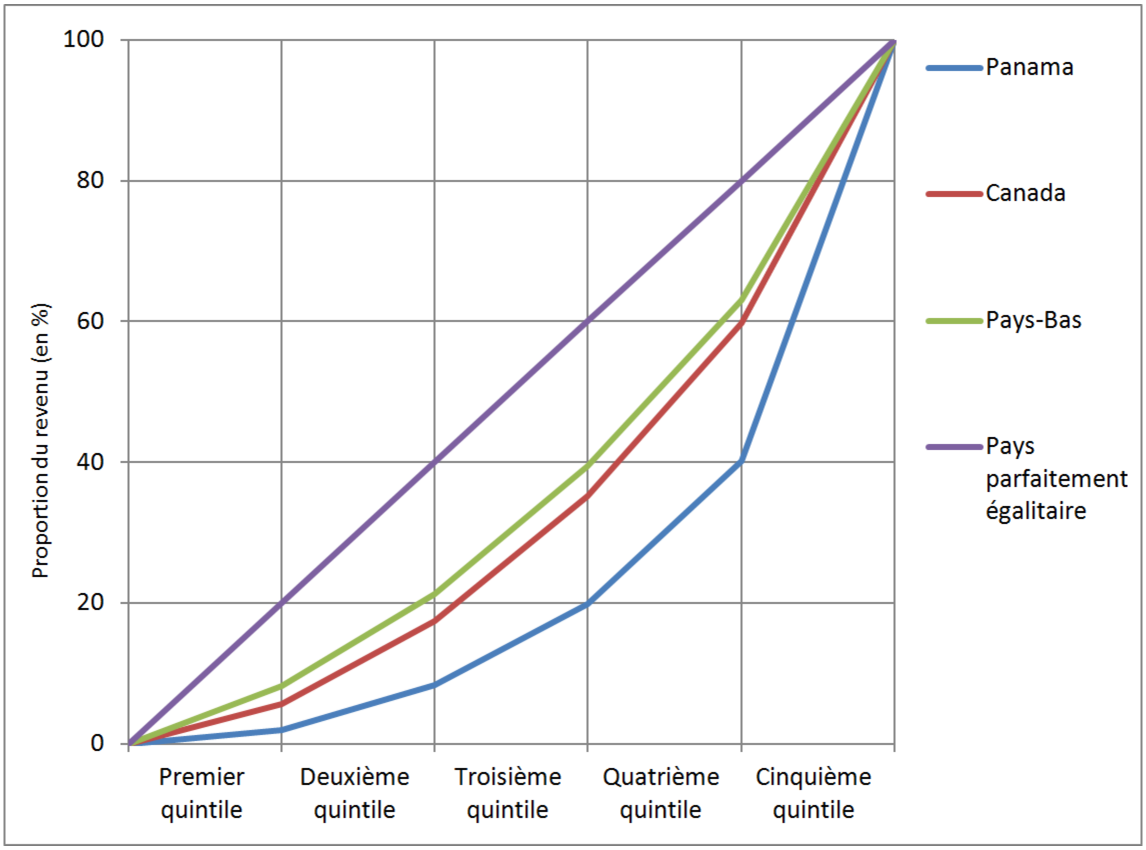
3.9). De l’autre côté, la tranche la plus riche de la population (le cinquième quintile) se partage 36,9 % du revenu. La figure qui accompagne le tableau permet d’évaluer d’un coup d’œil le degré d’inégalité de répartition du revenu dans un pays. Les courbes représentent les proportions *cumulées* du revenu au fur et à mesure qu’on ajoute des tranches. Le Panama est le pays qui s’éloigne le plus d’une répartition égalitaire. Cette représentation graphique s’appelle la courbe de Lorenz.

TABLEAU 3.9 - Répartition du revenu dans divers pays

(en %)		**Quintile le plus pauvre**	**Deuxième quintile**	**Troisième quintile**	**Quatrième quintile**	**Quintile le plus riche**
Panama 1989	Proportion du revenu	2,0	6,3	11,6	20,3	59,8
	Proportion cumulée	*2,0*	*8,3*	*19,9*	*40,2*	*100,0*
Canada 1987	Proportion du revenu	5,7	11,8	17,7	24,6	40,2
	Proportion cumulée	*5,7*	*17,5*	*35,2*	*59,8*	*100,0*
Pays-Bas 1988	Proportion du revenu	8,2	13,1	18,1	23,7	36,9
	Proportion cumulée	*8,2*	*21,3*	*39,4*	*63,1*	*100,0*
Pays parfaitement	Proportion du revenu	20,0	20,0	20,0	20,0	20,0
égalitaire	*Proportion cumulée*	*20,0*	*40,0*	*60,0*	*80,0*	*100,0*

Source : Banque mondiale, Rapport sur le développement dans le monde 1994.

Note : On a ajusté le revenu afin de tenir compte du pouvoir d'achat dans chaque pays.

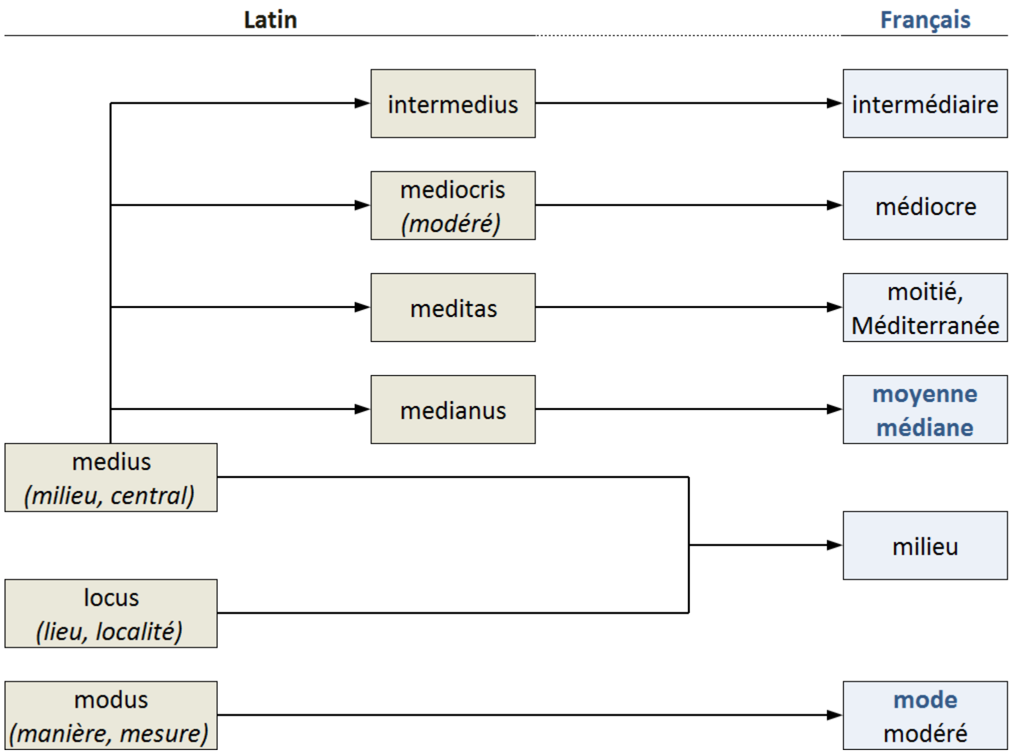


Parfois, on découpe une distribution en 4 tranches (on a alors des quartiles), en 10 tranches (les déciles), voire en 100 tranches (les centiles). Le principe reste le même.

4.3. Pour ne pas y perdre son latin

Nous venons de présenter beaucoup de mots : moyenne, médiane, milieu, moitié. Ce qui compte, c'est de bien comprendre chaque concept et non d'impressionner l'auditoire avec du jargon. D'ailleurs, les quatre mots que nous venons de citer ont tous la même origine : ils proviennent du mot latin *medius* qui veut dire *au milieu*. Observez votre main : votre médius (appelé aussi le majeur) n'est-il pas la médiane de votre collection de doigts? La figure 3.6 montre l'origine de tous ces mots.

Figure 3.7 - L'origine des mots



Parfois, le mot *moyenne* est utilisé à toutes les sauces. Le Québécois « moyen » n'est pas la moyenne des Québécois. Il s'agit plutôt du Québécois *ordinaire* (est-ce Elvis Gratton?) Le mot *moyenne* est utilisé dans le sens d'*intermédiaire* dans les expressions : classe moyenne, Moyen âge, Julot est de taille moyenne, placement à moyen terme. Ces exemples ne font que témoigner de la richesse du langage, reflet de la richesse humaine.

4.4. Pour les curieux : la moyenne harmonique

La moyenne dont nous avons parlé jusqu'ici est la *moyenne arithmétique* (ou *moyenne* tout court). Dans la presque totalité des cas, cette moyenne est amplement suffisante (pour le vérifier, demandez à un de vos professeurs de sciences humaines s'il connaît la moyenne harmonique ou la moyenne géométrique).

La « colle » suivante nous permettra d'expliquer simplement la *moyenne harmonique* (nous verrons la moyenne géométrique dans le prochain chapitre). Notre but est uniquement de montrer, encore une fois, que l'utilisation des chiffres en sciences humaines est plus une question de jugement qu'une question de calcul.

Un automobiliste fait l'aller-retour entre Montréal et Gatineau (400 km) en parcourant les 100 premiers kilomètres à 50 km/h, les 100 kilomètres suivants à 100 km/h, les 100 suivants à 150 km/h et les 100 derniers à 200 km/h. Voilà un ex-chauffeur (son permis lui a été retiré suite à cette expérience) qui, une fois bien réchauffé, n'a pas froid aux yeux. Notre question est la suivante : quelle est la vitesse moyenne de l'automobile sur l'ensemble du parcours?

Étant donné que le parcours est divisé en 4 tronçons égaux, on serait tenté d'additionner les 4 vitesses et de les diviser par 4.

(50 km/h + 100 km/h + 150 km/h + 200 km/h)/4 = 500/4 = 125 km/h.

Mais si les 4 sections du parcours sont de *longueur* égale, le *temps* qu'il faut pour parcourir chacun d'entre eux est différent. Or, la vitesse dépend aussi du temps. Nous connaissons déjà la distance totale, il nous suffit de calculer le temps total du trajet pour obtenir la vitesse moyenne.

Il faut 2 heures pour parcourir les 100 premiers km (à 50 km/h), 1 heure pour le tronçon suivant, 40 minutes pour le troisième tronçon et 30 minutes pour le dernier, soit en tout 4 h 10 min pour parcourir 400 km. La vitesse moyenne est donc de 400 km/4 h 10 min. Notons qu'avant d'effectuer cette division, il faut convertir les minutes en fractions d'heures : 10 min = 10/60 d'heure = 0,166 heure. Exprimé sous forme décimale, le temps du parcours est de 4,166 heures. La vitesse moyenne est de 400 km/4,166 heures = 96 km/h.

Ce dernier résultat (le bon : 96 km/h) est bien inférieur à celui de notre premier calcul (le faux : 125 km/h). N'est-ce pas la preuve que « rien ne sert de courir, il faut partir à point »?

Pour calculer la moyenne harmonique :

- 1. Additionner l'inverse de chaque valeur
  1/50 + 1/100 + 1/150 + 1/200 = 12/600 + 6/600 + 4/600 +3/600 = 25/600 = 1/24
- 2. Diviser le nombre de valeurs par le résultat précédent.
  4/(1/24) = 4 × (24/1) = 96.

Moyenne harmonique = Nombre de valeurs/[(1/Valeur 1) + (1/Valeur 2) + …]

Il y a une leçon à tirer de cet exemple. La formule ou même le nom de la moyenne harmonique ne parviendront peut-être pas à s'implanter dans votre mémoire, mais vous vous souviendrez sans doute de la chose suivante : il est plus important de réfléchir aux données du problème que de se lancer aveuglément dans des calculs.

La moyenne harmonique est employée lorsque les variables considérées sont en réalité des rapports (ici : la distance divisée par le temps) et lorsque c'est le dénominateur qui varie (ici : le temps).

EXERCICES 4

1. Les hommes

À l'aide des données du tableau 3.6, calculez l'étendue, la médiane, le mode et le coefficient de variation de l'espérance de vie des hommes des pays hispanophones d'Amérique centrale en 1992. Justifiez vos réponses avec quelques calculs si nécessaire.

EXERCICES SUPPLÉMENTAIRES

1. Le train? Pas dangereux!

Le tableau suivant indique le nombre de personnes tuées dans des accidents de train dans la patrie du TGV (train à grande vitesse).

TABLEAU 3.10 - Nombre de personnes tuées dans des accidents de train en France entre 1972 et 1992

1972	1973	1974	1975	1978	1982	1983	1984	1985	1987	1988	1991	1992
107	3	10	9	1	3	7	1	78	2	52	14	0

Source : Quid 1995.

Note : Aucune victime en 1976, en 1977, entre 1979 et 1981, en 1986, en 1989, en 1990 et en 1992.

a) Calculez la moyenne du nombre annuel de victimes.

b) Calculez l'écart type du nombre annuel de victimes.

2. Une moyenne délicate

Il y a en Inde 880 millions d'habitants dont 340 millions parlent l'hindi. On dénombre en tout 845 langues différentes dans ce pays.

La Papouasie-Nouvelle-Guinée ne compte que 4,1 millions d'habitants. Le pays est très montagneux et presque entièrement recouvert de forêts. On y parle pas moins de 700 langues*.

Source : *L'état du monde 1994*. Données de 1992.

a) Quel est le nombre moyen de locuteurs par langue en Inde?

b) Quel est le nombre moyen de locuteurs par langue en Nouvelle-Guinée?

Note : un locuteur est une personne qui parle une langue.

3. Seul dans le noir

Selon *La Culture en perspective* (Statistique Canada 87-004-XPB, hiver 1995), il y avait en 1993-94 au Canada 664 cinémas disposant en tout de 1727 écrans. Les 78 812 milliers d'entrées payantes ont alors rapporté 400,5 millions de dollars de recettes.

a) Quel était alors le nombre moyen d'écrans par cinéma?

b) Quel était le nombre moyen de spectateurs (payants) par écran?

c) Quelle était la recette moyenne par entrée?

4. Simon, José, Miguel et les autres

Un texte célèbre de José Martí, poète et patriote cubain du XIX^{e} siècle, raconte comment un voyageur fourbu arrivant à Caracas s'enquiert, avant toute chose, de l'emplacement de la statue du libérateur Simon Bolivar. Puis, l'auteur fait le portrait et l'éloge de Bolivar, de José de San Martín et de Miguel Hidalgo. De ce dernier, Martí dit qu'il était de la race des hommes bons, c'est à dire des hommes qui *veulent savoir*. N'est-ce pas là l'idéal de l'étudiant en sciences humaines? (Référence : *José Martí, Tres héroes*.)

a) Cherchez la date de l'indépendance des pays d'Amérique hispanophones.

b) Calculez la moyenne et l'écart type de ces dates.

c) Commentez.

5. Une moyenne piégée

Il y a 1000 familles dans la ville. La moitié des familles de la ville compte trois enfants. Le reste des familles n'a pas d'enfant (chiffres fictifs).

a) Quel est le nombre moyen d'enfants par famille?

b) Quel est le nombre moyen de frères et sœurs que peut avoir un enfant?

6. Un comportement normal?

En 1993, une enquête* menée auprès de 304 étudiants (118 garçons et 186 filles) dans deux universités de l'est des États-Unis a donné, entre autres, les résultats suivants : l'âge moyen de la première relation sexuelle était de 16,47 pour les garçons (avec un écart type de 2,4) et de 17,4 pour les filles (avec un écart type de 2,3). Nous supposerons, pour les besoins de l'exercice, que la variable *âge de la première relation* était distribuée de façon normale, aussi bien pour les garçons que pour les filles.

Source : *Journal of Health Education*, mai-juin 1996, p. 144 à 152.

Pour faire cet exercice, vous pouvez utiliser la table de distribution normale présentée dans ce chapitre, ou la table plus détaillée fournie en annexe.

a) Quelle est la proportion de filles ayant eu une première relation sexuelle avant l'âge de 17,4 ans? avant l'âge de 19,7 ans? après l'âge de 19,7 ans? avant l'âge de 21 ans? entre l'âge de 17,4 ans et de 19,7 ans? entre l'âge de 16 ans et 20 ans?

b) Quel est l'âge à partir duquel on peut affirmer que 1 % des garçons ont eu une première relation sexuelle? 10 % des garçons? 10 % des filles? 20 % des garçons? 25 % des garçons? 50 % des garçons? 50 % des filles?

7. Un pays sans classes

En 1992, le centile le plus fortuné de la population américaine détenait 42 % du patrimoine (contre 21 % en 1979) et 46 % des actions cotées en Bourse. Le décile le plus fortuné détenait, quant à lui, 89 % des actions cotées en Bourse. Dans le même intervalle de temps (1979-1992), la famille médiane n'avait augmenté son patrimoine que de 10 %. (Sources : Courrier international, 8 février 1996.)

a) Quelle était la proportion des actions détenues par les 1 % des Américains les plus fortunés en 1992? par les 9 % suivants? par les 90 % les moins fortunés?

b) Quelle était la proportion d'Américains qui détenaient un patrimoine plus élevé que la famille médiane?

c) Montrez que l'écart entre riches et pauvres s'est accru aux États-Unis entre 1979 et 1992.

d) Entre 1992 et 2012, le PIB américain par habitant en dollars constants a augmenté de 37 % (source : Banque mondiale, IDM). En combinant ce fait aux informations contenues dans la figure 2.3, pouvez-vous déterminer si les inégalités se sont accrues ou non au cours de ces vingt ans aux États-Unis?

8. Changement de médiane

Au Canada*, l'âge médian des personnes infectées par le VIH est passé de 32 ans en 1982 à 27 ans en 1983-84 et à 23 ans en 1994. Compte tenu de cela, que pensez-vous des affirmations suivantes?

Source de ces données : Tendances sociales, été 1996, SC. 11-008-XPF.

a) La moitié des personnes infectées en 1994 avaient 23 ans ou plus.

b) Les personnes infectées en 1994 avaient en moyenne 23 ans.

c) La moyenne d'âge des personnes infectées a sensiblement diminué en l'espace de 12 ans.

9. À l'autre bout du monde

Répondez aux questions suivantes à l'aide du tableau 3.11.

TABLEAU 3.11 - Les femmes australiennes

			1964	1974	1984	1994
1	Poids moyen (16 à 20 ans)	kg	55,5	56,2	56,9	57,6
2	Taille moyenne (16-20 ans)	cm	162,2	162,9	163,6	164,3
3	Espérance de vie	années	74,2	74,8	78,3	81,0
4	Pourcentage de fumeuses adultes	%	28,0	30,0	31,0	23,3
5	Âge médian du mariage	années	24,3	21,4	24,2	26,6
6	Taux de mariage	‰	7,7	8,1	7,4	6,2
7	Durée médiane du mariage	années	14,3	10,7	7,7	7,7
8	Taux de divorce	%	0,7	1,3	2,8	2,7
9	Âge médian des maternités	années	26,4	25,7	27,1	29,0
10	Utilisation de la pilule contraceptive (20-24 ans)	%	25,0	46,0	50,0	52,2
11	Pourcentage des filles ayant réussi leur 12e année	%	18,7	31,6	48,0	79,9
12	Pourcentage des femmes à l'université	%	25,9	38,9	46,6	53,5
13	Proportion des femmes dans la population active	%	25,0	34,0	40,0	43,0
14	Proportion des femmes mariées	%	18,0	39,8	43,5	52,0
15	Ratio des gains femmes/hommes	%	59,0	74,6	78,0	81,6

Source : The Bulletin, 16 avril 1996, Sydney, d'après une compilation de plusieurs études.

a) Le tableau contient 3 moyennes, 3 médianes, 6 proportions et 3 autres types de rapports. Identifiez-les.

b) Quelle est la proportion des femmes dont le mariage dure plus de 7,7 ans en 1994?

c) Expliquez pourquoi les phrases suivantes ont des bonnes chances de « s'avérer » fausses! (*Note* : tous les chiffres concernent l'année 1994.)

i) La moitié des femmes mesure 164,3 cm.
ii) Les femmes se marient en moyenne autour de 26,6 ans.
iii) 6,2 % des femmes sont mariées.
iv) La majorité des femmes utilise la pilule contraceptive.
v) Il y a environ deux fois plus d'hommes que de femmes à l'université.
vi) Les hommes gagnent 18,4 % de plus que les femmes.

d) Tracez des courbes illustrant l'évolution du poids moyen et de la taille moyenne des Australiennes au cours des décennies.

e) Quelle est l'évolution du ratio taille/poids au cours des décennies?

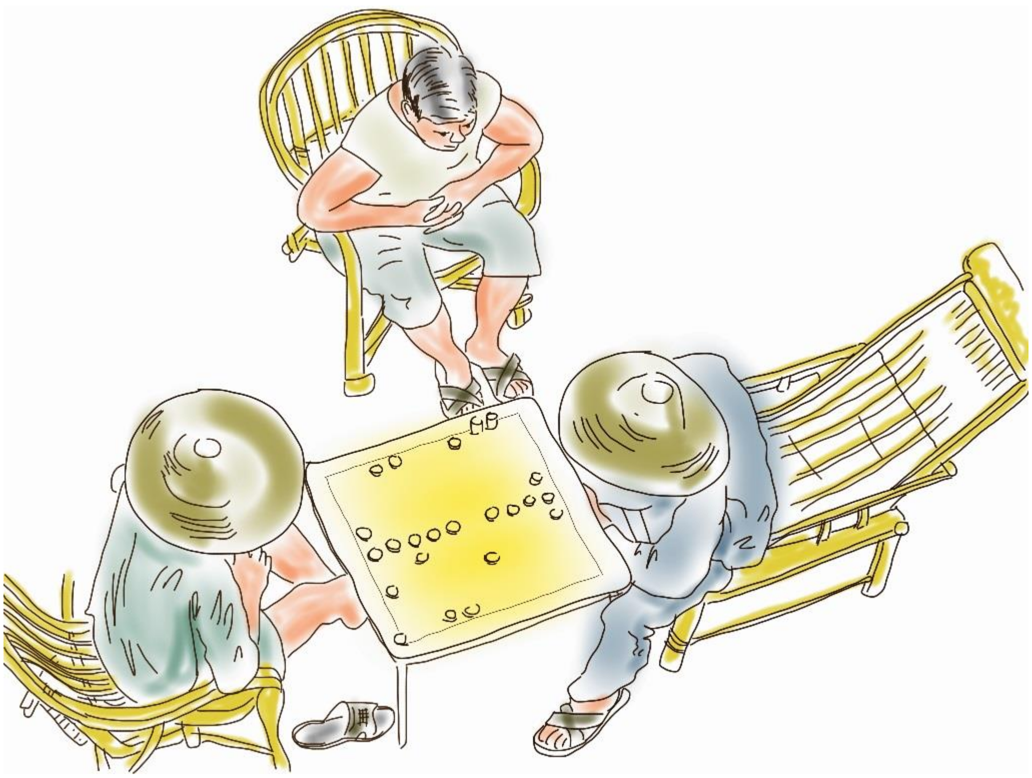
10. Recherche

Essayez d'obtenir l'âge des députés, des conseillers municipaux, des professeurs d'une école ou d'un département ou de tout autre groupe intéressant. Calculez la moyenne d'âge et l'écart type. Comparez les résultats entre eux : différences entre hommes ou femmes, entre périodes et entre lieux différents.

DOSSIER 3 LA RÉFORME DE L'ÉCRITURE CHINOISE

Apprendre à lire et à écrire le chinois n'est pas chose facile, même en Chine.

À cause de sa complexité, l'écriture chinoise est longtemps demeurée l'apanage de la classe dominante. Depuis la proclamation de la république (en 1911) et plus encore depuis la révolution de 1949, la réforme de l'écriture chinoise est à l'ordre du jour. On a songé à transcrire le chinois en alphabet latin (le système *pinyin*), comme cela a déjà été fait, par exemple, pour le vietnamien. Un élève vietnamien peut en effet apprendre à écrire en 100 heures alors qu'il en faut 500 pour un élève chinois.



Toutefois, l'utilisation de l'alphabet latin pose de gros problèmes. Il y a en chinois beaucoup de mots différents qui ont la même prononciation et donc la même transcription en alphabet phonétique pinyin. Écrits en caractères chinois, ces mots perdent leur ambiguïté, un peu comme les mots français *ô*, *oh*, *eau*, *au*, *haut*, *os* qui ne se distinguent que par leur orthographe différente. C'est pourquoi l'alphabet phonétique pinyin n'a eu qu'un succès limité. Il sert surtout à orthographier les noms propres de façon standardisée (Mao Zedong, Sichuan et Beijing) et à enseigner le chinois aux étudiants étrangers débutants.

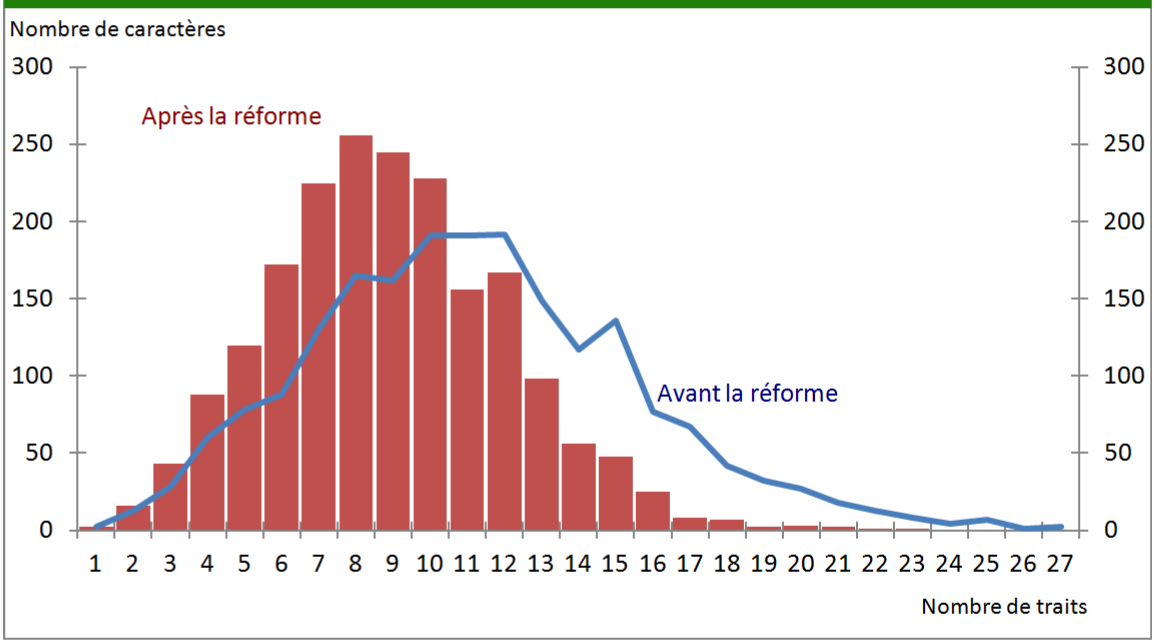
D'ailleurs, les caractères chinois possèdent un avantage évident. Ils permettent à un Pékinois, un Cantonais et un Japonais de communiquer — de façon élémentaire — par écrit, bien qu'ils parlent des langues différentes. Toutes proportions gardées, le Français, l'Anglais et l'Arabe en font autant lorsqu'ils s'échangent des numéros de téléphone, qui s'écrivent de la même façon, mais qui se prononcent différemment.

Au lieu d'une révolution, pourquoi pas une réforme?

Plutôt que d'abandonner les caractères chinois, on a alors cherché à les simplifier, en réduisant le nombre de traits nécessaires pour les tracer. Le comité mis en place dans les années 1950 a établi la liste des 2000 caractères de base que toute personne devrait connaître pour fonctionner convenablement dans la vie courante. Puis il a décidé d'en simplifier 795 et d'en supprimer 31, qui faisaient théoriquement double emploi.

La figure D3.1 indique la distribution des 2000 caractères de base en fonction du nombre de traits, avant et après la réforme. Le nombre moyen de traits par caractère a ainsi été réduit à 8,9, avec un écart type de 3,2. De plus, 71 % des caractères s'écrivent désormais avec 10 traits ou moins alors qu'avant la réforme, cette proportion n'était que de 46 %.

FIGURE D3.1 - Distribution des 2000 caractères chinois de base selon le nombre de traits



Calculer la moyenne et l'écart type de 2000 éléments n'est pas une mince affaire avec une calculatrice, même si ces éléments sont groupés en une vingtaine de catégories. Par contre, la chose est relativement facile avec un chiffrier électronique. C'est ce que nous avons fait pour les 1969 caractères d'après la réforme (voir tableau D3.1). À vous de trouver la moyenne et l'écart type pour les anciens caractères à l'aide d'un chiffrier électronique.

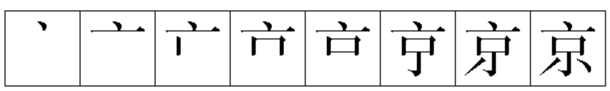
Un petit casse-tête... chinois

Quel rapport existe-t-il entre feu le président Mao Zedong, mort à Beijing et l'illustre metteur en scène Kurosawa Akira, né à Tokyo? Aucun apparemment, du moins si on se fie uniquement à la transcription en alphabet latin. Toutefois, si on utilise les caractères chinois, on s'aperçoit que *ze* et *sawa* s'écrivent de la même façon et veulent donc dire la même chose : le *marais* ou *l'abondance*. Tokyo est simplement la capitale de l'est et Beijing la capitale du nord : *jing* et *kyo* s'écrivent de la même manière et signifient *capitale*. C'est pourquoi l'annonce dans le journal d'une visite de Kurosawa à Mao aurait pu être lue ainsi indifféremment par un Chinois ou un Japonais : M. Marais-noir rencontre M. Poil Marais-de-l'est à sa descente de l'avion qui dessert la ligne capitale-de-l'est/capitale-du-nord (voir figure D3.2).

FIGURE D3.2 - Un petit casse-tête... chinois

Français	Chinois	Caractère simplifié	Caractère traditionnel	Japonais	Nombre de traits
noir	hei	黑	黑	kuro	12
marais	ze	泽	澤	sawa	8
nord	bei	北	北	kita	5
est	dong	东	東	to	5
capitale	jing	京	京	kyo	8
poil	mao	毛	毛	ke	4
quatre	si	四	四	yon	5
rivière	chuan	川	川	kawa	3
cap	qi	崎	崎	saki	11

Exemple :
8 traits



À quand le jumelage entre Trois-Rivières et le Sichuan, et entre la rivière du Cap-rouge (en banlieue de Québec) et Kawasaki (en banlieue de Tokyo)?

QUESTIONS

a) Comparez les deux distributions de la figure D3.1 à la courbe normale.

b) À première vue, quelle est la distribution qui possède la plus grande moyenne? Quelle est la distribution qui a le plus grand écart type?

c) Calculez le nombre moyen de traits par caractère avant la réforme.

d) Calculez l'écart type du nombre de caractères avant la réforme.

e) Calculez la proportion de caractères de plus de 12 traits avant et après la réforme.

CHAPITRE 4 LES DONNÉES CHRONOLOGIQUES

TABLE DES MATIÈRES

1. Le taux de variation
2. L’indice de variation
3. Les stocks et les flux
4. Les variations à long terme

• Exercices supplémentaires

• Dossier

Ce chapitre s’intitule « Les données chronologiques » parce que nous y étudierons l’évolution dans le temps des valeurs que peuvent prendre des variables.

Les données chronologiques sont particulièrement utilisées en économie, en démographie et, naturellement, en histoire. Mais toutes les disciplines des sciences humaines y ont recours, ne serait-ce que parce que le passé éclaire généralement le présent et permet d’entrevoir l’avenir.

Manipuler des données chronologiques peut sembler particulièrement simple, mais là encore, il est essentiel d’acquérir quelques techniques de base et de procéder avec méthode. Une fois muni des outils appropriés, l’étudiant — ou le bon journaliste — pourra non seulement mieux lire dans le passé, mais il sera également en mesure de communiquer les informations trouvées de façon claire et efficace.

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

- Comment peut-on mesurer la vitesse à laquelle les valeurs d’une variable augmentent ou diminuent avec le temps?
- Comment, connaissant cette vitesse et la valeur de départ, peut-on déterminer la valeur d’arrivée de cette variable?
- Comment, connaissant cette vitesse et la valeur d’arrivée, peut-on déterminer la valeur de départ de cette variable?
- Comment notre méconnaissance des indices de variation peut-elle provoquer de grossières erreurs de raisonnement et de mesure?
- Comment peut-on suivre sur une longue période de temps, l’évolution des valeurs prises par une variable en éliminant ses variations périodiques et récurrentes?

1. LE TAUX DE VARIATION

Reportons-nous à la mi-août 1995, en Amérique latine. Alors que la récolte de café est amorcée depuis quelques semaines, les autorités brésiliennes doivent se rendre à l'évidence : la production de café ne dépassera pas les 10 millions de sacs pour la campagne de 1995-96. Dès l'année précédente, alors que les zones productrices étaient touchées par un hiver austral particulièrement rigoureux (gelées de juin et sécheresse de juillet) le Brésil, premier exportateur mondial s'attendait à une réduction dramatique de sa production pour la récolte suivante. On prévoyait alors pour 1995-96 une production de 12,7 à 14,8 millions de sacs. Maintenant que les grains sont arrivés à maturité, on constate que la production sera encore plus faible que prévu. À titre de comparaison, la campagne précédente (récolte 1994-95) avait permis de récolter 25 millions de sacs. (Notons qu'un sac de café contient officiellement 60 kilogrammes de ces précieux grains.)

Comme de nombreux journaux, le *Listin Diario de Santo Domingo* du 14 août 1995 publie la nouvelle dans ses pages économiques. Pour rédiger son article, le journaliste s'inspire largement du texte envoyé par l'agence de presse brésilienne. Après tout, un journal, ça paraît… tous les jours : il faut donc être rapide et efficace. Il ne reste plus qu'à trouver un titre concis et accrocheur, un titre qui fasse ressortir en peu de mots les principaux éléments d'information (quoi? où? quand? et combien?). Le *Listin* titre : « La production de café du Brésil diminuera cette année de 60 %. » Tout est dit, ou presque. Il ne reste plus qu'à savoir pourquoi, et pour cela on est obligé de lire l'article lui-même.

1.1. Tout est dit en un seul chiffre : le taux de variation

Ce qui nous intéresse ici, c'est le chiffre de –60 %, qui résume à peu près toute l'information quantifiable. Le journaliste, parmi tous les chiffres dont il disposait, en a extrait les deux principaux : 25 millions de sacs l'année précédente et 10 millions l'année en cours. Mais, deux chiffres c'était encore un de trop pour lui. Il combine alors ces deux chiffres en un seul : –60 %. Ses connaissances en méthodes quantitatives complètent avantageusement ses compétences journalistiques.

Le taux de variation mesure l'évolution relative de la valeur d'une variable dans le temps.

Le chiffre de –60 % s'appelle un taux de variation. Il est obtenu en comparant deux données : la production de l'an dernier et la production de cette année. Ces données ont une caractéristique commune : elles sont exprimées dans la *même unité* (des millions de sacs de 60 kg). Par contre, ces données viennent de deux *périodes différentes* (l'an dernier et cette année). En somme, le taux de variation mesure ici l'évolution de la valeur d'une variable dans le temps.

Comment calcule-t-on le taux de variation?

La comparaison la plus simple consiste à mesurer l'*écart absolu* entre le point de départ et le point d'arrivée. Ici, l'écart est de (10 – 25) = –15. On peut dire aussi que la production *baisse* de 15 unités. Cependant, le seul chiffre de –15 ne veut pas dire grand-chose si on ne connaît pas le point de départ. En effet, si, par exemple, la production était passée de 3015 à 3000, il n'y aurait pas eu de quoi fouetter un chat. Mais dans le cas présent, la baisse de 15 représente plus de la moitié de la récolte initiale. Pour interpréter la différence entre le point de départ et le point d'arrivée, on ne peut se contenter de l'écart absolu. Il faut calculer l'*écart relatif*. On a perdu 15 unités sur les 25 unités initiales : l'écart relatif ou taux de variation est de –15/25 = –0,6 = –60 %.

Taux de variation = (Valeur finale – Valeur initiale)/Valeur initiale Taux de variation = (F – I)/I

Le taux de variation s'exprime généralement sous forme de pourcentage, mais il ne s'agit là que d'un habillage commode.

Notons que, dans le paragraphe précédent, nous avons fourni deux réponses : –0,6 et –60 %. Il s'agit bien sûr du même chiffre, présenté sous deux habillages différents. Dans le même ordre d'idées, on peut dire que 0,60 $ équivalent à 60 cents (ou centièmes). Lorsque vous devez interpréter un chiffre, ne vous fiez pas à la forme qu'il prend, mais bien sûr la réalité qu'il recouvre. Comparons les deux informations suivantes : le Brésil produit 24 % de la récolte mondiale de café; la récolte brésilienne de café a baissé de 60 % cette année. Dans le premier cas, il s'agit d'une *proportion* (la partie divisée par le tout : voir le chapitre 2). Dans le second cas, il s'agit d'un *taux de variation*. Nous avons affaire à deux *concepts* différents déguisés sous le même *habit*. Étant donné que les pourcentages sont presque exclusivement utilisés pour ces deux concepts, il vaut vraiment la peine de les distinguer clairement.

1.2. Un taux de variation positif : le taux de croissance

Le prix des droits exclusifs de diffusion des Jeux Olympiques a connu une croissance ininterrompue et considérable depuis 1960. La vénérable chaîne CBS obtenait ces droits à Rome (en 1960) pour la bagatelle de 400 000 $. Quatre ans plus tard, à Tokyo, c'était au tour de la non moins vénérable NBC. En 1968, à Mexico, cette dernière se faisait damer le pion par son jeune rejeton, la chaîne ABC. Depuis, la féroce concurrence entre les deux sœurs ennemies a fait monter les enchères : 401 millions en 1992 pour Barcelone, et 456 millions en 1996 pour Atlanta (la patrie de Coca-Cola, est-ce vraiment une coïncidence?)

Si nous appliquons la formule que nous avons énoncée un peu plus haut, nous pouvons affirmer que le prix des droits a augmenté de 13,7 % entre les Jeux de Barcelone et ceux d'Atlanta. En effet, (456 – 401)/401 = 55/401 = 0,137 = 13,7 %. Ce chiffre de 13,7 % représente le *taux de variation* ou *taux de croissance*. Cette dernière appellation est la plus courante lorsque l'on s'attend à voir une valeur augmenter d'une période à l'autre. Et si, par malheur, la variable se met soudain à baisser, on parlera, non sans un certain culot, de « croissance négative ».

> La hausse vertigineuse des droits de diffusion s'est poursuivie de plus belle dans les décennies suivantes. NBC, qui avait versé 4,38 milliards de dollars pour s'assurer l'exclusivité sur le territoire américain des Jeux olympiques de 2014 (Sotchi), 2016 (Rio de Janeiro), 2018 (Pyeongchang) et 2020 (Tokyo), débourse ensuite un montant supplémentaire de 7,65 milliards pour obtenir une rallonge de 12 ans, jusqu'en 2032.

1.3. Trouver le point d'arrivée grâce au taux de variation

Parfois, l'information fournie est incomplète : on possède le point de départ et le taux de variation, mais on voudrait bien connaître le point d'arrivée. Par exemple, sachant que le prix des droits de diffusion a augmenté de 56,8 % entre les Jeux d'Atlanta (où ils coûtaient 456 millions de $) et les Jeux de Sydney, quel est le prix des droits pour les Jeux de Sydney en l'an 2000? Nous utiliserons la formule suivante, qui est adaptée de la formule du taux de variation énoncée un peu plus haut.

$$\text{Valeur finale} = \text{Valeur initiale} + (\text{Valeur initiale} \times \text{Taux de variation})$$
$$F = I + (I \times \text{Taux de variation})$$

Dans notre cas, la valeur finale est de :

$$456 + (456 \times 56{,}8/100) = 715$$

ou encore, présenté de façon moins habituelle, mais plus pratique :

$$456 + (456 \times 0{,}568) = 715$$

N’oublions pas que les trois expressions suivantes : 56,8 %; 56,8/100 et 0,568 ont la même valeur, même si elles sont présentées différemment. Nous vous recommandons d’utiliser la forme décimale (0,568) pour vos calculs et le pourcentage (56,8 %) pour présenter les données au public.

Au fait, que pensez-vous de la variante suivante? Qui sait, elle pourrait nous servir bientôt.

$$456 \times 1{,}568 = 715$$

Quel degré de précision doit-on adopter?

Le plus souvent, on se contente de présenter les taux de variation avec tout au plus un chiffre après la virgule lorsqu’ils sont sous forme de pourcentage. Dans ce cas, il faut donc conserver 3 chiffres après la virgule lorsque le taux est sous forme décimale. On aura ainsi 56,8 % (un chiffre après la virgule) et 0,568 (3 chiffres). On dira par exemple qu’en Chine le taux de croissance annuelle de la population chinoise est de 1,4 % (entre 1985 et 1990) et le taux d’inflation de 7,4 % (en 1980). Par contre, on se contentera d’affirmer que les recettes budgétaires de l’État ont baissé de 20 % (en 1990*). Ce dernier chiffre est suffisamment élevé, et les méthodes comptables suffisamment floues, pour qu’on ne s’embarrasse pas de précision inutile. Remarquez aussi le chiffre de –60 % pour la production de café brésilienne : comme la récolte n’est pas terminée, il serait mal venu de donner des chiffres plus précis qui n’intéresseraient d’ailleurs pas les lecteurs. Enfin, vous pouvez facilement vérifier que toutes les données mentionnées ci-dessus sont des taux de variation.

Source : L’État du monde 1992.

EXERCICES 1

1. 1. Avancer lentement, c'est parfois reculer

a) Le tableau ci-dessous indique la répartition de la population canadienne selon la langue maternelle lors des recensements de 1951, 1991 et 2011. Calculez le taux de variation des effectifs des trois communautés entre 1951 et 1991, puis entre 1991 et 2011.

Tableau 4.1 - La population canadienne selon la langue maternelle

	1951	1991	2011
		(en milliers de gens)	
Anglais	8 281	16 785	19 340
Français	4 069	6 643	7 214
Autre	1 660	3 869	6 568
Total	14 010	27 297	33 122

Sources : Annuaire du Canada 1994; Statistique Canada, Recensement 2011.

Note : Pour l'année 2011, nous avons réparti équitablement les déclarants possédant plusieurs langues maternelles.

b) En 1951, il y avait dans les Provinces maritimes 235 000 francophones. Au cours des 40 années suivantes, le nombre de francophones augmentait de 23 %. Combien y avait-il de francophones en 1991?

c) Les trois groupes linguistiques bénéficient d'un taux de croissance positif. Mais qu'en est-il des proportions de ces trois groupes? Ont-elles diminué ou augmenté?

2. La vie est courte, mais de moins en moins

a) En 1931, l'espérance de vie à la naissance était, au Canada, de 62,1 ans pour les femmes et de 60 ans pour les hommes. En 1991, les chiffres étaient respectivement de 80,6 et 74. Calculez le taux de croissance de l'espérance de vie pour les femmes et les hommes entre 1931 et 1991.

b) En 2007-2009, l'espérance de vie était passée à 83,3 ans pour les femmes et 78,8 ans pour les hommes. Calculez le taux de croissance de l'espérance de vie pour les femmes et les hommes entre 1991 et 2007-2009.

2. L'INDICE DE VARIATION

Si vous avez bien compris la section qui précède, vous avez compris l'essentiel. Mais ce n'est pas une raison suffisante pour nous arrêter là. Car c'est une fois que l'essentiel est atteint que la vie commence à être vraiment intéressante*. Nous vous posons donc un défi sous la forme d'une double colle.

Nous espérons que ce proverbe, que nous venons d'inventer, figurera un jour dans les pages roses du *Petit Larousse*!

Colle nº 1

Sachant que la récolte brésilienne de café a baissé de 60 % pour atteindre 10 millions de sacs, et en admettant que vous ayez oublié la valeur de la récolte initiale, êtes-vous capable de retrouver la valeur de cette même récolte initiale?

Colle nº 2

Sachant que le prix des droits de diffusion pour les Jeux Olympiques a augmenté de 33,7 % entre Séoul (1998) et Barcelone (1992) où ils ont atteint 401 millions de dollars, quel était le prix des droits à Séoul?

Notons que les deux colles sont basées sur un problème similaire : nous savons ajouter un taux de variation à un point de départ pour trouver le point d'arrivée, mais savons-nous faire le chemin inverse? Sommes-nous capables de reconstituer le point de départ à l'aide du point d'arrivée et du taux de variation? Finalement, la seule différence entre les deux colles que nous vous proposons est que vous ne pouvez pas tricher avec la deuxième.

Prenons d'abord la colle nº 1 et essayons de nous enfoncer dans une impasse. C'est d'ailleurs le meilleur moyen d'apprendre. Toute personne soi-disant sensée serait tentée de faire le petit calcul suivant : la récolte actuelle est de 10, j'y ajoute 60 % pour retrouver la récolte initiale, et ça me donne en tout 16 (10 + [60 % × 10]) = 10 + 6 = 16. C'est très simple, malheureusement c'est entièrement faux : nous savons que la récolte initiale était de 25 et non de 16. Notre petit calcul a une faille (sur le plan purement logique et non mathématique). Savez-vous laquelle?

En espérant que vous avez pris le temps de chercher avant de changer de paragraphe, nous allons vous indiquer la source de l'erreur. On ne peut pas ajouter 60 % à la récolte actuelle, car ces 60 % concernent la récolte initiale. Si nous savions que la récolte initiale est de 25, nous pourrions calculer les 60 % qui nous intéressent (60 % × 25 = 15). Nous saurions alors que la production a baissé de 15 (et non pas de 6) et, puisque nous sommes rendus à 10, nous pouvons en déduire que nous sommes partis de 25. Si nous connaissions la récolte initiale, il nous serait facile de résoudre le problème, mais justement, nous ne connaissons pas cette donnée clé : nous la cherchons!

C'est le problème du chien qui court après sa queue. Pour savoir comment le résoudre, vous allez être obligé de lire le reste de cette section. Mais rien ne vous empêche de poser votre manuel pendant quelques minutes (quelques heures?) et de découvrir vous-même la formule magique.

2.1. Pour ajouter un taux de variation : une manière rapide et surtout… réversible

Supposons que l'on veuille ajouter un certain pourcentage (mettons 10 %) à un chiffre (mettons 200 $), quelle que soit la raison. Il y a deux manières de procéder. La manière conventionnelle consiste à multiplier le chiffre par le pourcentage et à ajouter le résultat au chiffre initial :

$$200\ \$ + (200\ \$ \times 10\ \%) = 200\ \$ + 20\ \$ = 220\ \$$$

ou encore :

$$200\ \$ + (200\ \$ \times 0{,}10) = 200\ \$ + 20\ \$ = 220\ \$$$

Nous avons présenté cette formule un peu plus haut, et tous les écoliers la connaissent.

Voici une autre façon d'arriver au même résultat:

$$200\ \$ \times 1{,}10 = 220\ \$$$

L'indice de variation est le rapport entre la valeur finale et la valeur initiale d'une variable.

Lorsqu'on multiplie un nombre par 1, on ne le change pas : son taux de variation est nul. On voit ici qu'en multipliant le nombre par 1,10, on ne fait que lui ajouter 10 % (ou 0,10) de sa valeur : on ajoute 10 cents à chaque dollar. Dans le même ordre d'idée, on ajouterait 12 % en multipliant le nombre par 1,12 et 20 % en le multipliant par 1,20. Ce multiplicateur s'appelle indice de variation (ou facteur de variation).

Quel est l'intérêt de cette deuxième méthode, à part le fait qu'elle est nouvelle (et donc suspecte!)? C'est tout simplement que cette méthode est réversible. Essayons de partir du chiffre final (220 $) pour retrouver le chiffre initial. Il est clair que nous ne pouvons ôter 10 % à 220 $, car cela nous donnerait : 220 $ – 22 $ = 198 $.

La réponse est fausse parce que ce n'est pas à 220 $ que se rapportent les 10 %. Par contre, si nous utilisons notre nouvelle méthode en l'appliquant à l'envers, nous pourrions retrouver le chiffre initial : puisque nous avons *multiplié* 200 par 1,10 pour lui ajouter 10 %, nous allons *diviser* le résultat final par 1,10 pour revenir au point de départ :

$$220\ \$/1{,}10 = 200\ \$$$

Pour ajouter un taux de variation à un nombre, on multiplie le nombre par l'indice de variation correspondant. Pour revenir au point de départ, on divise le résultat final par l'indice de variation. L'indice de variation n'est autre que le taux de variation auquel on ajoute 1.

Valeur finale = Valeur initiale × Indice de variation
Valeur initiale = Valeur finale/Indice de variation
Indice de variation = 1 + Taux de variation
Taux de variation = Indice de variation – 1

Il est temps maintenant de résoudre les deux colles proposées au début de la section. Dans le premier cas, le taux de variation est de –60 % : l'indice de variation est donc égal à 1 + –0,60 = 0,40 (ou 0,4)

Récolte initiale = Récolte finale/Indice de variation
Récolte initiale = 10/0,40 = 25

Dans le deuxième cas, le taux de variation est de 33,7 % : l'indice de variation est donc égal à 1 + 0,337 = 1,337.

Séoul (initial) = Barcelone (final)/1,337
Séoul = 401/1,337 = 300

Les droits de diffusion télévisés pour les Jeux Olympiques étaient donc de 300 millions à Séoul et de 401 millions aux Jeux suivants (à Barcelone). Il est d'ailleurs facile de vérifier que (401 – 300) / 300 = 33,7 %.

Si cette méthode vous paraît nouvelle, sachez que vous l'avez sans doute déjà utilisée bien des fois sans le savoir. Ajouter 100 %, c'est comme multiplier par 2, n'est-ce pas? Le nombre 2 n'est que le facteur de variation correspondant au taux de variation de 100 % (car 1 + 100 % = 2). Et pour revenir en arrière, vous divisez par 2, vous enlevez la moitié, vous ôtez 50 %.

Une erreur à ne pas faire.

Un de nos amis prétend qu'il est plus facile de perdre du poids que d'en gagner et voici sa démonstration : après avoir grossi de 33,3 % (son poids étant passé de 60 à 80 kg, soit une hausse de 20/60), il lui a suffi de maigrir de 25 % (20 kg divisés par 80) pour retrouver son poids initial. Compte tenu de ce que nous venons de voir, il ne faut pas s'étonner que 33,3 % de 60 kg soient égaux à 25 % de 80 kg. Le taux de variation représente un écart relatif entre *deux* moments dans le temps : tout dépend du côté où on se place pour observer les choses.

C'est pour la même raison que lorsque le dollar canadien vaut 0,75 $US, on constate que le dollar américain vaut 1,33 $CAN. L'écart relatif entre 0,75 et 1 est en effet le même qu'entre 1 et 1,33.

1 $CAN vaut 0,75 $US, d'où 1 $US = 1/0,75 = 1,33 $CAN
1 $CAN vaut 1,50 florin, d'où 1 florin = 1/1,50 = 0,67 $CAN
1 $CAN vaut 5 yuans, d'où 1 yuan = 1/5 = 0,20 $CAN
1 $CAN vaut 120 yens, d'où 1 yen = 0,0083 $CAN (ou 100 yens = 0,83 $CAN)

2.2. Un miracle pour New York?

Un maire qui promet de rendre New York plus sécuritaire tout en réduisant les impôts

Lors de sa première campagne électorale à la mairie de New York, le candidat démocrate Giuliani avait promis des miracles aux New Yorkais. Une fois élu, le maire Giuliani semble avoir tenu ses promesses. Il a réussi à convaincre les employés municipaux, les commerçants et les simples citoyens que les efforts demandés profiteraient à tous. Dès la deuxième année de son mandat (1995-96), les dépenses ont diminué de 4,2 %. Seul le budget de la police a augmenté (de 7 %). Les baisses de taxes (un fait plutôt inhabituel) ont surtout profité aux commerçants, aux hôteliers et aux petits entrepreneurs, mais le maire prétend que le dynamisme économique qui en découlera est plus susceptible de sortir la ville du marasme que l'aide sociale. Pour la première fois, on voit le crime battre en retraite dans la métropole de l'Amérique du Nord, et de façon spectaculaire.

Par la suite, le maire Giuliani se fera mondialement connaître en se rendant le 11 septembre 2011 sur les ruines fumantes du *World Trade Center*, pendant que le président G.W. Bush restait prudemment à l'abri au camp militaire de Bellevue (Nebraska).

Un des meilleurs hebdomadaires italiens d'information, *L'Espresso* publie, en 1995, une série de chiffres sur le bilan du maire Giuliani (lui-même d'origine italienne). Quelle que soit l'opinion que l'on peut avoir sur les méthodes du maire (courageuses ou un peu trop musclées?), il faut commencer par examiner les faits, à la lumière de nos connaissances sur les taux de variation. Il devient alors indéniable, grâce aux chiffres du tableau 4.2, que la situation a véritablement changé, quelles qu'en soient les raisons exactes.

TABLEAU 4.2 - Le crime à New York : le bilan du maire Giuliani

	1993	1994	Taux de variation
	(nombre)	(nombre)	(en %)
Délits dans le métro	11 767	9 213	-21,7
Vols d'automobiles	*111 600*	94 525	-15,3
Coups et blessures	*41 127*	39 770	-3,3
Homicides	*1 947*	1 581	-18,8
Total d'arrestations	**285 418**	**344 499**	***20,7***

Source : L'Espresso, 3 marzo 1995.

On nous dit que les crimes dans le métro sont passés de 11 767 (en 1993) à 9213 (en 1994), soit une baisse de 21,7 %. On peut facilement vérifier ce chiffre : (9213 – 11 767)/11 767 = –0,217 = –21,7 %. Pour les autres délits (voir tableau 4.2), la revue publie seulement les dernières données alors disponibles (1994), et le taux de variation par rapport à l'année précédente (1993). Toutefois, rien ne nous empêche d'en déduire les données brutes de 1993 (ces résultats sont inscrits en vert dans le tableau 4.2).

Il y a eu 94 525 vols d'automobiles en 1994, soit une baisse de 15,3 % par rapport à l'année précédente. Combien de vols y a-t-il eu en 1993? Il nous suffit d'utiliser la formule : *Valeur initiale = Valeur finale/Indice de variation*. L'indice de variation est ici 0,847 (soit 1 + [–15,3 %] ou encore 1 – 0,153). La valeur initiale est de 94 525/0,847 = 111 600. Il y a eu 111 600 vols d'automobiles en 1992. On peut d'ailleurs le vérifier ainsi : (94 525 – 111 600)/111 600 = –15,3 %. Nous avons utilisé le même procédé pour obtenir les autres valeurs manquantes du tableau 4.2 (inscrites en vert). Vous pouvez essayer de les vérifier par vous-mêmes.

2.3. La croissance cumulée

Nous savons maintenant ajouter un taux de variation et même revenir au point de départ, mais jusque-là nous nous sommes limités à deux périodes consécutives. Que se passe-t-il, par exemple si une variable connaît une croissance annuelle de 10 % pendant 10 années consécutives? Pour rendre les choses plus intéressantes, supposez que votre vieille tante vous lègue 100 $ que vous placez à 10 % d'intérêt. Que vaudra votre héritage dans 10 ans (en supposant que le taux d'intérêt ne change pas en cours de route et que vous laissiez bien sagement l'argent à la banque)?

On serait tenté, pour se débarrasser rapidement du problème, de dire que le capital va doubler : 10 % pendant 10 ans, cela ne fait-il pas 100 %? Mais soyons plus ambitieux dans notre réflexion, la fortune est à ce prix. La première année, le capital se multiplie par 1,10 (il augmente de 10 %) pour atteindre 110 $. L'année suivante, ce dernier montant se multiplie à nouveau par 1,10 pour atteindre 121 $ (car 110 $ × 1,10 = 121 $). Ce n'est pas encore un montant colossal, mais c'est plus que prévu.

En 2 ans, le capital a été multiplié 2 fois par 1,10 :

$$100\ \$ \times 1{,}10 \times 1{,}10 = 121\ \$$$

En trois ans, le capital serait multiplié 3 fois par 1,10 :

$$100\ \$ \times (1{,}10 \times 1{,}10 \times 1{,}10) = 133{,}10\ \$$$
$$\text{ou encore } 100\ \$ \times (1{,}10)^3 = 133{,}10\ \$$$

En 3 ans, la valeur du placement est passée de 100 \$ à 133,10 \$: cela signifie qu'elle a augmenté de 33,1 % en l'espace de *3 ans*. Pourtant, la progression *annuelle* n'a été que de 10 %, ou, si l'on préfère, le placement a été multiplié par 1,10 chaque année. N'oublions pas que multiplier une quantité par 1,10 équivaut à lui ajouter 10 % (pour ajouter 15 %, on multiplie par 1,15, etc.). En 3 ans, le placement a donc été multiplié par (1,10 × 1,10 × 1,10) ou par $(1{,}10)^3 = 1{,}331$. Trois augmentations successives de 10 % (ou 0,10) se traduisent par une augmentation cumulée de 33,1 % (ou 0,331).

> Essayons, par curiosité, d'aller en sens inverse. Partons du chiffre 1,331 (augmentation totale *cumulée* sur trois ans) pour retrouver le chiffre initial de 1,10 (progression *annuelle*). Il suffit justement d'effectuer l'opération inverse : $(1{,}331)^{1/3} = 1{,}10$. Pour le calcul des exposants, rien de mieux qu'une bonne vieille calculatrice : c'est pour elle un jeu d'enfant.

Revenons maintenant à notre héritage. Examinons sa valeur après 10 ans et, pourquoi pas, après 30 ans.

Après 10 ans, le capital sera de 100 \$ × $(1,\ 10)^{10}$ = 100 \$ x 2,59 = 259 \$.

Après 30 ans, le capital sera de 100 \$ × $(1{,}10)^{30}$ = 100 \$ × 17,45 = 1745 \$. Il aura été multiplié par 17,45 (qui est l'indice de variation sur 30 ans). Il aura augmenté de 1645 % (qui est le taux de variation sur la même période, soit l'indice moins un : 17,45 – 1 = 16,45 = 1645/100 = 1645 %). Un capital de 1745 \$, ça commence à être une somme intéressante, à moins que dans 30 ans ce montant corresponde au prix… d'un paquet de cigarettes.

Une autre erreur à ne pas faire.

Beaucoup de gens seraient prêts à jurer qu'une hausse de 30 % suivie d'une hausse 20 % donnent une hausse globale de 50 %. Quant à nous, nous venons de voir que les taux de variation se cumulent. Dans un premier temps, la valeur de la variable est multipliée par 1,30 (on y ajoute 30 %) et dans un deuxième temps elle est multipliée par 1,20. En tout, elle est donc multipliée par (1,30 × 1,20) = 1,56, ce qui correspond à une hausse combinée de 56 % par rapport à la valeur initiale.

Dans le même ordre d'idées, une hausse de 20 % combinée à une baisse de 10 % équivaut à une hausse nette de 8 %. En effet, la variable est successivement multipliée par 1,20 (soit 1 + 20 %) et par 0,90 (soit 1 – 10 %). Le facteur de croissance combiné est de 1,20 × 0,90 = 1,08. Le taux de croissance combiné est de 8 %.

2.4. La croissance moyenne (ou moyenne géométrique)

Selon la Banque Mondiale, la population du Nigéria devait augmenter de 136 % entre 1970 et 2000. Quelle était la croissance annuelle *moyenne* de la population?

Comment peut-on estimer l'évolution de la valeur d'une variable si on ne connaît pas ses taux de variation successifs?

Nous sommes ici en présence du problème inverse de celui que nous venons de voir (l'héritage de la vieille tante) : nous connaissions la croissance annuelle et nous devions en déduire la croissance *cumulée* après quelques années. Cette fois, nous connaissons la croissance cumulée en 30 ans et nous devons en déduire la croissance annuelle. Bien évidemment, comme cette croissance annuelle a pu varier en cours de route, nous n'obtiendrons qu'un taux de croissance annuelle *moyen*.

Revenons au Nigéria. La population y fait plus que doubler : elle augmente de 136 %. Elle est multipliée par 2,36 (1 + 136 % ou 1 + 1, 36) en l'espace de 30 ans. En *moyenne*, cela signifie que la population est multipliée par 1,029 (soit $2{,}36^{1/30}$) par an. Le taux de croissance est de 0,029 (soit 1,029 – 1) ou 2,9 %.

Indice de croissance moyenne = Indice de croissance cumulée $^{(1/\text{Nombre de périodes})}$

Dans le cas du Nigéria, le nombre de périodes correspond au nombre d'années qui se sont écoulées, puisque nous cherchons à évaluer la croissance annuelle. De la même façon, on pourrait évaluer une croissance mensuelle moyenne sur 6 mois ou une croissance hebdomadaire moyenne sur 13 semaines.

Pour illustrer à nouveau comment calculer un taux de croissance moyen, reprenons un instant les données sur le nombre de francophones au Canada (voir le tableau 4.1). Le nombre de francophones passe de 4 069 000 à 6 643 000 entre 1951 et 1991. Cela équivaut à une croissance de 63,3 % ([6 643 – 4 069]/4069 = 0,633 = 63,3 %) pour cette période de 40 ans. Le nombre de francophones a donc été multiplié par 1,633 (c'est l'indice qui correspond au taux de 63,3 %).

L'indice de croissance annuelle moyenne est donc de $1{,}633^{1/40}$ = 1,012. Le taux correspondant est de 1,2 % (soit 1,012 – 1 = 0,012 = 1,2 %). On peut en déduire que le taux de croissance de la population francophone du Canada a été en moyenne de 1,2 % par an entre 1941 et 1991. C'est pas mal moins que le Nigéria, et pourtant les « Canadiens français » étaient encore très prolifiques dans les années 1940 et 1950. (*Voir le tableau 4.3.*)

TABLEAU 4.3 - L'indice de croissance cumulée en fonction du taux annuel et du nombre d'années			
	Nombre d'années		
Taux annuel	**2**	**10**	**20**
1,0%	1,020	1,105	1,220
5,0%	1,103	1,629	2,653
7,2%	1,149	2,004	4,017
10,0%	1,210	2,594	6,727

Le tableau 4.3 met en relation trois variables : le taux de croissance annuel, le nombre d'années, et l'indice de croissance cumulée. La valeur d'une variable qui subit un taux de croissance annuel de 5 % sera, par exemple, multipliée par 1,629 au bout de 10 ans. Elle aura donc augmenté de 62,9 %. On sait que l'indice (ici 1,629) est égal à 1 + le taux (ici 62,9 %) : on a bien en effet 1+ 62,9 % = 1 + 0,629 = 1,629.

Lorsque deux des trois données en jeu sont connues, on peut en déduire la troisième :
• À partir du taux annuel (5 %) et du nombre de périodes (10 ans) on trouve l'indice cumulé :
$(1 + 5\ \%)^{10} = (1 + 0{,}05)^{10} = (1{,}05)^{10} = 1{,}629$.

• À partir de l'indice cumulé (1,629) et du nombre de périodes (10 ans) on trouve l'indice annuel :
$(1{,}629)^{1/10} = 1{,}05$ (d'où le taux annuel moyen de 1,05 – 1 = 0,05 = 5 %).

• À partir du taux annuel (5 %) et de l'indice cumulé (1,629), on trouve le nombre de périodes nécessaires (pour le moment, le calcul le plus simple consiste à y aller au visé :
8 ans donnent $(1{,}05)^{8} = 1{,}477$: ce n'est pas assez;
11 ans donnent $(1{,}05)^{11} = 1{,}71$: c'est un peu trop;
10 ans donnent $(1{,}05)^{10} = 1{,}629$ en plein dans le mille!

Note : Nous avons utilisé les années à titre d'exemple : la durée de la période peut varier (mois, jours, secondes. etc.).

EXERCICES 2

1. 1. Sur les bords de la rivière Hudson

En 1995-96, les dépenses de la ville de New York s'élevaient à 2,180 milliards de dollars (une hausse de 7 % par rapport à l'année fiscale précédente) pour la police et à 2,840 milliards pour les écoles (en baisse de 7,5 %).

Quelles étaient les dépenses correspondantes pour l'année 1994-95?

2. Télécommunication passive et active

Au début des années 1950, la télévision était cantonnée aux États-Unis (90 % de tous les téléviseurs dans le monde). Elle s'est vite répandue dans les autres pays industrialisés, mais ce n'est qu'à partir de 1975 qu'elle a commencé à conquérir l'ensemble de la planète. De 1975 à 1994, le nombre de foyers équipés de téléviseurs a augmenté de 950 % en Asie.

a) Quel est le taux d'augmentation annuelle moyen du nombre de foyers équipés d'un téléviseur en Asie, entre 1975 et 1994?

b) En vous basant sur les données du tableau 4.4 , comparez le taux de croissance annuel moyen du nombre de foyers équipés de téléviseur dans le monde au cours des années 1980 avec celui de la période 1990-94.

TABLEAU 4.4 - Évolution du parc mondial de téléviseurs et de téléphones

Foyers équipés de téléviseur (en millions)

	1960	1970	1980	1990	1994
Monde	83	244	450	658	888

Abonnements au téléphone (en millions)

	Téléphone fixe			Téléphone mobile	
	2005	2014		2005	2014
Pays développés	570,1	511,1		992,0	1515,4
Pays en développement	673,1	635,6		1213,2	5399,8
Monde	**1243,2**	**1146,7**		**2205,3**	**6915,2**

Sources : « The Futurist » dans Courrier international, 5 octobre 1995 (téléviseurs); UIT, Indicateurs clés 2005-2014 (téléphone).

c) En vous basant sur les données du tableau 4.4 , calculez les taux de croissance des abonnements au téléphone fixe et au téléphone mobile entre 2005 et 2014 pour les pays développés et les pays en développement.

d) Même question que la précédente, mais, cette fois-ci, vous devez calculer les taux de croissance annuels moyens.

3. Apogée du microordinateur

Comme on l'a observé dans un chapitre précédent, les ventes totales de microordinateurs (portables inclus) sont passées de 48 millions d'unités en 1994 à 355,2 millions d'unités en 2011, avant de reculer à 341,3 millions en 2012, et à 305,2 millions en 2013.

a) Calculez le taux de croissance annuel moyen du nombre de microordinateurs vendus entre 1994 et 2011.

b) Calculez le taux de décroissance annuel moyen du nombre de microordinateurs vendus entre 2011 et 2013.

3. LES STOCKS ET LES FLUX

Dans tous les exemples énoncés jusqu'ici dans ce chapitre, il était question de l'évolution de la valeur d'une variable dans le temps. C'est d'ailleurs pourquoi nous avons intitulé le chapitre « Les données chronologiques ». Mais avez-vous remarqué que l'on comparait tantôt des *périodes* (récolte annuelle de café, budget de la ville de New York) et tantôt des *moments* précis (nombre de téléviseurs en 1994, population du Canada au 1er juillet 1951)?

3.1. Choisir ses mots

Dans le même ordre d'idées, on parlera de l'*année* (moment) où les Beatles ont sorti leur premier album et des 8 *ans* (période) qu'a duré leur carrière officielle. Lisez la phrase précédente à voix haute en inversant les mots *année* et *an* : vous pourrez facilement convaincre vos auditeurs que le français n'est pas votre langue maternelle. Il faut admettre cependant que la langue manque parfois de précision. Ainsi, le mot *heure* indique à la fois la période (je t'ai attendu pendant une heure) et le moment (nous avions rendez-vous à une heure*).

Dans ce cas, par contre, certaines langues comme l'anglais ou le chinois sont plus explicites : le mot *heure* sera traduit par deux termes différents.

Le stock mesure la valeur d'une variable à un *moment* précis.

Le mot stock est une image qui rappelle le décompte des inventaires que font les commerces à la fin de l'année. Le berger compte ses moutons, l'avare compte son or, le châtelain compte ses bouteilles de Château Margaux, le sergent compte ses soldats, et le Koweït compte ses réserves de pétrole. Dans tous les cas, il s'agit d'un stock mesuré à un moment précis.

Le flux mesure la valeur d'une variable pendant une *période* de temps.

Le mot flux rappelle le débit d'un cours d'eau ou d'un robinet (le stock serait alors le réservoir alimenté par ce débit). Le nombre d'agneaux mis au monde pendant la saison, les dividendes trimestriels, les emplettes de la journée, le nombre de recrues qui grossissent les rangs de l'armée chaque année, la production quotidienne de barils de pétrole sont tous des exemples de flux, mesurés sur une période de temps.

3.2. Qu'est-ce qui alimente la population du Québec?

La survie d'un peuple est en bonne partie une question de nombre. Le Québec ne serait pas si « distinct » aujourd'hui si ses enfants n'avaient pas été aussi prolifiques pendant les 200 ans qui ont suivi la Conquête (officialisée en 1763). Pour parler moins poétiquement, le *stock* de Québécois ne se serait pas multiplié par 100 sans être alimenté par un *flux* considérable de naissances pendant deux siècles.

Depuis les années 1960, la situation s'est cependant modifiée de façon radicale. Le tableau 4.5 donne des chiffres sur les flux qui alimentent la population québécoise à l'ère des microfamilles. On y retrouve 2 flux d'entrée : les naissances et l'immigration. Ces flux, qui totalisent 145 000 personnes en 2012-2013, correspondent au robinet d'alimentation. Ils font augmenter le niveau du réservoir d'habitants. Les 2 flux de sortie y sont les décès et l'émigration : 74 000 personnes en tout, qui constituent en quelque sorte le conduit d'évacuation du réservoir. On nous pardonnera ces comparaisons de mauvais goût entre des êtres humains et des litres d'eau, en sachant que notre but est avant tout de bien mettre en évidence les notions de flux et de stock. En fin de compte, le flux net de l'année 2012-2013 est de +71 000 individus (145 000 entrées moins 74 000 sorties).

TABLEAU 4.5 - Évolution de la population du Québec

	1991-1992	1992-1993	2011-2012	2012-2013
		(en milliers)		
Naissances	98	94	88	89
Décès	49	51	59	62
Immigration	52	48	54	56
Émigration	58	45	7	12
Total : Flux nets	**43**	**47**	**76**	**71**
Population en fin de période	**7 110**	**7 157**	**8 084**	**8 155**

Sources : Statistique Canada, Cansim 051-0001 et Cansim 051-0004.

Note : Les périodes annuelles vont du 1er juillet au 30 juin.

Étant donné que les flux nets de population sont de +71 000 individus au Québec en 2012-2013, il n'est pas étonnant de voir la population (stock) passer de 8 084 000 habitants au 1er juillet 2012 à 8 155 000 un an plus tard : l'augmentation du stock est égale aux flux nets.

Stock final = Stock initial + Flux d'entrée – Flux de sortie

Variation des stocks = Stock final – Stock initial

Flux nets = Flux d'entrée – Flux de sortie

Variation des stocks = Flux nets

En pratique, il est long et coûteux de mesurer les stocks. C'est pourquoi ils ne sont recensés qu'à des moments éloignés : la caisse n'est vérifiée qu'à la fin de la journée, l'inventaire du magasin n'est effectué qu'à la fin de l'année, et la population n'est recensée que tous les 5 ou 10 ans. Entretemps, les données sur les flux, beaucoup plus faciles à obtenir, permettent de se faire une idée des stocks. Et, puisque la mesure des flux peut être faussée par une série de facteurs (erreurs, oublis et… tricherie), le recensement périodique des stocks permet de remettre les pendules à l'heure.

3.3. Dur comme du bois

Dans certains États forestiers américains où les droits de coupe sont élevés, le bois de construction canadien est vu comme un concurrent déloyal. Les nombreuses feuilles d'érable dont les producteurs canadiens décorent leurs lots, par fierté ou par chauvinisme, sont perçues comme autant de symboles patents de l'invasion étrangère. Mais revenons à une époque où le bois représentait une des principales exportations canadiennes et où le drapeau unifolié se montrait encore très discret.

D'après le tableau 4.6, on constate que le Canada produit plus de bois qu'il n'en consomme : un surplus (ou flux net) de 6017 unités en 1966. Le commerce extérieur génère un flux net à peu près inverse : 5843 unités sortent du pays. La différence (6017 – 5843 = 174 unités) fera augmenter le stock de bois de 174 unités au cours de l'année 1966. Ce stock passe en effet de 1308 à 1482 entre le 31 décembre 1965 et le 31 décembre 1966, soit une augmentation de 174.

TABLEAU 4.6 - Production de bois de construction au Canada			
	1965	**1966**	**1967**
	(en milliers de pieds planche)		
Production	..	10 008	9 962
Consommation	..	3 991	3 994
Exportations	..	6 133	6 487
Importations	..	290	299
Stocks en fin d'année	**1 308**	**1 482**	**1 262**

Résumé des flux en 1966			
	Intérieur	**Extérieur**	**Total**
Entrées	10 008 (production)	290 (importations)	**10 298**
Sorties	3 991 (consommation)	6 133 (exportations)	**10 124**
Flux net	**6 017**	**-5 843**	**174***

Source : Statistiques historiques du Canada 1983.

* Au cours de l'année, les stocks de bois ont augmenté de 174 unités.

Ces mêmes résultats peuvent être obtenus en appliquant la formule :

$$\text{Stock final} = \text{Stock initial} + \text{Flux d'entrée} - \text{Flux de sortie.}$$
$$1\,482 = 1\,308 + (10\,008 + 290) - (3991 + 6133)$$

Ces données nous renseignent également sur le taux de rotation des stocks. On constate que les flux annuels (qui tournent autour de 10 000) sont environ 7 fois supérieurs aux stocks (qui tournent autour de 1400). On peut en déduire que le bois entreposé se renouvelle 7 fois au cours de l'année et qu'il séjourne en moyenne un peu moins de deux mois (1/7 d'année) dans les hangars.

SITUATION MONDIALE DE L'OFFRE ET DE LA DEMANDE DE CÉRÉALES EN 2012/13

Source : FAO, Perspectives de récolte et situation alimentaire, mars 2013.
(Extraits)

« [1] Les dernières prévisions de la FAO concernant la production mondiale de céréales de 2012 ont été révisées en hausse de 4 millions de tonnes par rapport au chiffre de février, pour s'établir à 2 306 millions de tonnes (y compris le riz usiné), ce qui reste environ 2 pour cent de moins que le volume record de l'année précédente.

[2] […] Alors que l'utilisation fourragère de blé devrait progresser de 2,4 pour cent pour s'établir à 649 millions de tonnes, niveau record, la chute de 10 pour cent attendue en ce qui concerne l'utilisation de maïs pour la production de carburant à base d'éthanol aux États-Unis, qui selon les prévisions passerait de 127 millions de tonnes en 2011/12 à 114 millions de tonnes en 2012/13, sous-tend la contraction globale de 3,2 pour cent de l'utilisation mondiale de céréales secondaires dans des secteurs autres que les secteurs alimentaire et fourrager.

[3] […] Les prévisions concernant les stocks de report de céréales secondaires restent inchangées par rapport à février, à savoir 165 millions de tonnes. Ainsi, les réserves mondiales seraient en recul de 6 pour cent (10 millions de tonnes) par rapport à leur niveau d'ouverture, des prélèvements sur les réserves étant attendus aux États-Unis et dans l'UE, de près de 9 millions et 4,3 millions de tonnes, respectivement. En revanche, les réserves mondiales de riz devraient gagner 7,3 pour cent (11,7 millions de tonnes), pour passer à 172 millions de tonnes, grâce aux grandes quantités accumulées en Chine, mais aussi en Thaïlande, où le programme d'achat continue de détourner du marché les disponibilités, qui viennent gonfler les réserves publiques.

[4] […] Le rapport entre les stocks de clôture des grands exportateurs de céréales et l'utilisation totale (définie comme la somme de l'utilisation intérieure et des exportations) devrait, selon les estimations, passer de 17,9 pour cent pour la campagne précédente à 16,4 pour cent en 2012/13. »

Questions

a) Parmi les éléments quantitatifs que nous avons mis en gras dans le texte, lesquels représentent des *taux de variation*, lesquels représentent des *rapports*, et lesquels représentent des *données brutes* (et parmi ces dernières, distinguez les *flux* des *stocks*)?

b) On indique, dans l'article, que « l'utilisation fourragère de blé devrait progresser de 2,4 pour cent pour s'établir à 649 millions de tonnes ». Quel était le montant de cette utilisation l'année précédente?

b) On affirme, dans le texte, que « les réserves mondiales de riz devraient gagner 7,3 pour cent (11,7 millions de tonnes), pour passer à 172 millions de tonnes ». Vérifiez que les trois chiffres cités dans cette phrase sont bien cohérents.

EXERCICES 3

1. Des mots bien choisis

Parmi les expressions suivantes, identifiez les *flux* et les *stocks*.

Naissances, solde migratoire, population, nombre de chômeurs, nombre d'emplois créés, nombre de prisonniers, nombre de personnes condamnées, nombre d'États membres de l'ONU, nombre de nouveaux États admis à l'ONU.

2. Le bois unifolié

À l'aide des chiffres du tableau 4.6, vérifiez que la variation des stocks de bois en 1967 correspond bien aux flux nets.

3. Crise et prospérité

a) Entre 1931 et 1941, la population du Canada est passée de 10,377 millions à 11,507 millions d'habitants. Au cours de cette décennie, les naissances se sont élevées à 2,294 millions, les décès à 1,072 million, tandis que l'immigration atteignait 149 000 et l'émigration 241 000 (*source :* Statistique Canada, *Recensements de la population*). Calculez le flux naturel et le flux migratoire pour la décennie. Vérifiez si les flux nets correspondent à l'accroissement de la population pendant cette même période.

b) Entre 1901 et 1911, le portrait était tout différent. La population du Canada passait de 5,371 millions à 7,207 millions d'habitants. Au cours de cette décennie, les naissances, les décès, l'immigration et l'émigration étaient respectivement de 1,925 million, 900 000, 1,550 million et 740 000. Comme pour la question précédente, calculez le flux naturel et le flux migratoire pour la décennie, et vérifiez si les flux nets correspondent à l'accroissement de la population pendant cette même période.

4. LES VARIATIONS À LONG TERME

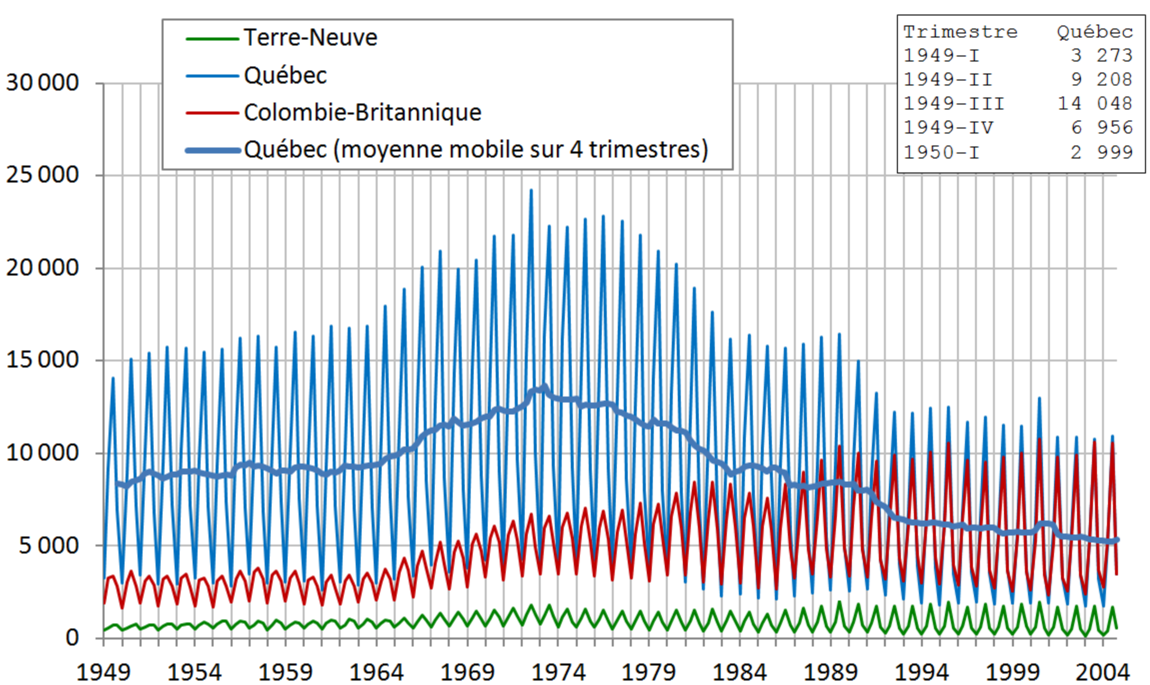
Pour conclure ce chapitre, nous verrons comment une représentation graphique des données chronologiques permet parfois de mieux saisir leur évolution. Nous chercherons également à bien séparer la tendance à long terme des variations à court terme. Enfin, nous verrons comment tricher (et comment ne pas se faire avoir) avec des courbes de données chronologiques.

4.1. Le retour mythique aux bonnes vieilles traditions

Cela fait longtemps que l'on parle d'un retour en force du mariage au Québec. On prétend, avec toutes sortes de pseudo-chiffres à l'appui, qu'après quelques années d'égarement, les jeunes sont enfin revenus aux « vraies valeurs ». Cependant, dans le domaine des sciences humaines, il vaut toujours mieux prendre les excès de nostalgie avec une certaine méfiance. La même réserve s'applique d'ailleurs face aux ennemis du passé qui croient clouer le bec à toute critique en commençant chaque affirmation par « de nos jours… ».

Pour notre part, nous avons décidé d'aller y voir de plus près. Alors des chiffres, s'il vous plait, afin que nous puissions les interpréter avec toutes les précautions nécessaires. Statistique Canada a longtemps publié le nombre trimestriel de mariages. C'est exactement ce qu'il nous fallait pour nous faire une idée de l'évolution *à long terme* du phénomène. Comme il s'est écoulé beaucoup de trimestres pendant la période considérée, nous avons choisi de représenter les données sous forme de graphique : les trimestres figurent sur l'axe horizontal et le nombre de mariages sur l'axe vertical. Pour faire bonne mesure, nous avons ajouté, à titre de comparaison, deux provinces canadiennes, une grosse et une petite, une moderne et une traditionnelle.

FIGURE 4.1 - Les mariages au Québec… et ailleurs (données trimestrielles)



Source : Statistique Canada, Cansim 53-0001

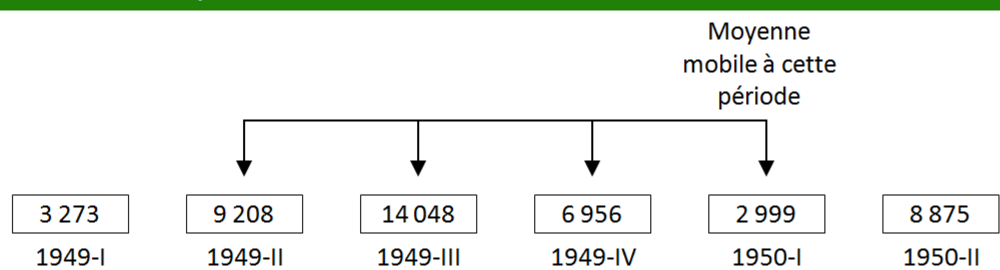
Pour analyser les données à long terme il est souvent utile (et facile, avec un chiffrier électronique) de les représenter sous forme de graphique.

La représentation graphique offre l'avantage de faire ressortir de façon frappante certaines tendances à propos du mariage au Québec. D'une part, les variations saisonnières sont fortes et très systématiques. On se marie surtout en été (3^{e} trimestre), un peu moins au 2^{e} et au 4^{e} trimestre et très rarement en hiver (1^{er} trimestre). Sur ce plan, rien n'a changé depuis l'après-guerre. D'autre part, le nombre de mariages a eu tendance à monter jusqu'au milieu des années 1970 pour diminuer sensiblement par la suite. D'ailleurs, n'est-ce pas depuis cette inversion de tendance qu'on entend le plus parler d'un retour du mariage? Il n'y a là rien d'étonnant : on ne peut en effet regretter que ce qui a disparu.

Pour mieux observer la tendance qui se cache derrière les variations saisonnières, nous avons calculé une *moyenne mobile* sur 4 trimestres. Cela veut dire que nous faisons, pour chaque trimestre, la moyenne du trimestre en cours avec les trois trimestres précédents. Ainsi, pour le 4^{e} trimestre 1949, nous faisons la moyenne des 4 trimestres de 1949. La moyenne mobile est alors de (3 273 + 9 208 + 14 048 + 6 956)/4 = 8371). Pour obtenir la moyenne suivante, nous incorporons le 1^{er} trimestre 1950 et nous laissons tomber le 1^{er} trimestre 1949 (9 208 + 14 048 + 6 956 + 2 999)/4 = 8302. Chaque moyenne comporte 4 trimestres (soit une année complète) et les influences saisonnières devraient alors disparaître.

Sur la figure 4.2, on peut voir la moyenne mobile dont il vient d'être question, pour le 1^{er} trimestre 1950. Pour calculer la moyenne mobile du trimestre suivant, on avance le « râteau » d'un cran vers la droite. Soulignons également que la moyenne mobile peut être utilisée sur des périodes diverses : sur 7 jours, sur 3 mois, sur 5 ans, etc. Tout dépend du phénomène étudié.

Figure 4.2 - La moyenne mobile



La tendance à long terme.

Avant de tirer des conclusions trop hâtives de la baisse des mariages depuis le début des années 1970, examinons ce qui peut influencer la tendance à long terme du mariage. Il y a deux éléments principaux : d'une part l'évolution démographique (taille de la population, proportion de jeunes en âge de convoler) et d'autre part l'évolution des mœurs. Notre intention n'est pas ici d'étudier le sujet à fond, mais simplement de montrer qu'avec un peu de méthode on est déjà bien armé pour traiter d'un sujet. La hausse des mariages n'a rien d'étonnant entre 1949 et 1965, compte tenu de la croissance soutenue de la population du Québec. La poussée de 1965-75 s'explique sûrement en bonne partie par l'arrivée à maturité des enfants d'après-guerre (les *baby-boomers*). Depuis 1975, bien que la population ait continué d'augmenter, il se peut que la proportion de jeunes se soit réduite. Toutefois, la descente est trop prononcée et trop systématique pour qu'on puisse l'imputer à de simples considérations démographiques. Il faut plutôt y voir un changement dans les comportements : on se marie plus vieux et, même si certaines personnes se marient maintenant plusieurs fois, on a plus souvent recours à l'union libre qu'autrefois (voir le tableau 4.7 en ce qui concerne ces hypothèses et, pour le reste, dites-vous que cela ferait un bon sujet de recherche).

Tableau 4.7 - L'évolution des couples au Québec et au Canada

Proportion de conjoints vivant en union libre

	1991	2001	2011	2013
		(en %)		
Terre-Neuve	7,7	11,3	15,6	15,2
Québec	19,4	30,1	37,1	36,7
Ontario	7,9	11,1	13,1	13,0
Colombie-Britannique	11,2	12,6	15,2	15,0
Canada	11,6	16,4	19,7	19,5

Source des données brutes : Statistique Canada, Cansim 051-0042.

Âge médian et âge moyen du mariage au Québec

	Hommes		Femmes	
	2000	2004	2000	2004
Âge médian	32,0	33,0	30,0	30,0
Âge moyen	35,3	35,9	32,5	33,2

Source : Statistique Canada, Cansim 101-1002.

L'évolution des autres.

La Colombie-Britannique a continué à connaître une forte croissance de sa population après les années 1960. Ce dynamisme, à la fois en quantité (population plus nombreuse) et en qualité (population plus jeune), explique en partie la hausse des mariages après le seuil fatidique du milieu des années 1970. Il est probable aussi que les changements de mœurs aient été moins prononcés qu'au Québec, puisque la Colombie-Britannique finit par se rapprocher du Québec (au chapitre du nombre de mariages) bien que sa population soit environ deux fois moindre.

On remarque enfin que les variations saisonnières étaient moins prononcées, autrefois, en Colombie-Britannique et à Terre-Neuve. Cela s'expliquait sans doute par un climat moins inconstant que celui du Québec. Mais on constate, au fil des ans, des amplitudes de plus en plus grandes. Se pourrait-il que pour certains le mariage soit devenu un gros investissement et un phénomène social? Il faut alors bien choisir sa saison, d'autant plus que les fiancés qui vivent déjà en concubinage sont moins pressés qu'autrefois.

4.2. Derrière les mouvements immédiats, la tendance profonde

Les données chronologiques portent naturellement sur des variables susceptibles d'évoluer à long terme. Mais elles subissent parfois des influences à court terme. Les mariages, comme on l'a vu, fluctuent d'une saison à l'autre; l'affluence dans les hôpitaux peut varier au cours de la semaine; et les hauts et les bas de l'économie, qui durent parfois plusieurs années influencent par exemple le chômage ou le commerce extérieur à court terme.

Les corrections saisonnières

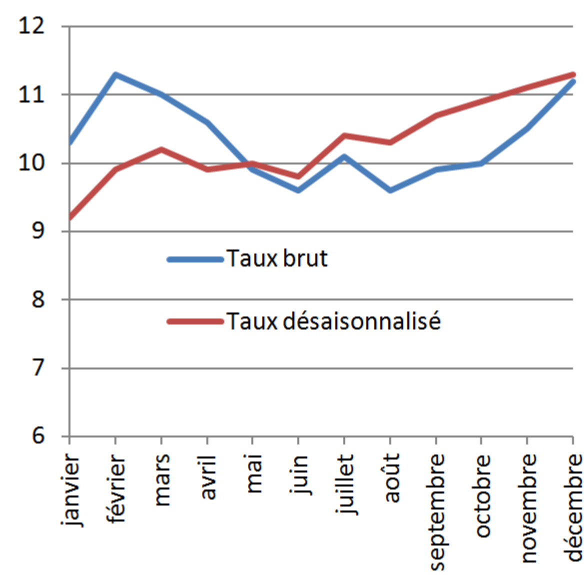
Lorsque les variations à court terme sont bien cernées et qu'elles reviennent de façon régulière et systématique, on a recours aux corrections saisonnières. En pratique, elles se font souvent sur une échelle annuelle, lorsque les variations à court terme sont semblables d'une année à l'autre.

Certains hôtels réalisent la moitié de leur chiffre d'affaires lors des vacances de juillet. Septembre marque depuis très longtemps la rentrée des classes et avril le paiement des impôts. Quant à la saison de ski, il y a fort à parier qu'elle aura toujours une prédilection pour l'hiver. Dans tous ces cas, et en supposant que les données sont mensuelles, on pourrait calculer, d'après l'observation des années précédentes, une série de douze écarts (un pour chaque mois) équivalents aux variations purement saisonnières.

La figure 4.3 illustre le principe des variations saisonnières en ce qui concerne le chômage. On voit sur le graphique que le taux brut est fortement influencé par les fluctuations saisonnières. De janvier à avril, le chômage est toujours relativement élevé, avec un sommet en février (le mois le plus triste de l'hiver?). Entre août et novembre, le chômage est toujours relativement bas. Le taux de chômage désaisonnalisé nous permet de voir ce qui se cache derrière le taux brut. Malgré les apparences (le taux brut descend pendant une bonne partie de l'année), la situation de l'emploi se détériore sérieusement au cours de l'année 1977, marquée par une récession. On y observe par exemple qu'en mars 1977, le taux de chômage brut a très peu diminué bien que la saison de création d'emplois soit déjà amorcée. C'est donc signe que le chômage a, en réalité, tendance à augmenter, ce que confirme le taux de chômage désaisonnalisé.

Figure 4.2 - La désaisonnalisation du chômage au Québec (année 1977)

	brut	désaisonnalisé	Écart
	(en %)		
janvier	10,3	9,2	-1,1
février	11,3	9,9	-1,4
mars	11,0	10,2	-0,8
avril	10,6	9,9	-0,7
mai	9,9	10,0	0,1
juin	9,6	9,8	0,2
juillet	10,1	10,4	0,3
août	9,6	10,3	0,7
septembre	9,9	10,7	0,8
octobre	10,0	10,9	0,9
novembre	10,5	11,1	0,6
décembre	11,2	11,3	0,1



***Source : Revue statistique du Canada*, juin 1978.**

La moyenne mobile.

Lorsque les variations à court terme surviennent de façon irrégulière et imprévisible, on peut avoir recours à la moyenne mobile pour déceler la tendance à long terme derrière les fluctuations à court terme. Nous avons vu plus haut la moyenne mobile sur 4 trimestres pour les mariages. Il existe également des moyennes mobiles sur 12 mois (on fait la moyenne des 12 derniers mois : le mois en cours et les 11 mois précédents).

Le choix du nombre de périodes utilisées pour calculer la moyenne mobile dépend de la variable étudiée et de l'usage que l'on veut en faire.

Les changements de périodicité.

Pour étudier l'évolution à long terme du nombre de mariages, nous aurions pu nous contenter de données annuelles. Comme les données trimestrielles dont nous disposons sont des données brutes, il nous suffit alors d'additionner les 4 trimestres de chaque année.

Lorsque les données sont des rapports, l'opération est un peu plus compliquée. Supposons que le taux de chômage (chômeurs/personnes actives) est de 11 % pendant les 4 premiers mois de l'année et de 10 % pendant les 8 mois suivants. On ne peut évidemment pas additionner les 12 taux de chômage (cela ferait [4 × 11] + [8 × 10] = 44 + 80 = 124 % de chômeurs!). Dans ce cas, on devrait faire une moyenne annuelle. On obtiendrait : [(4 × 11) + (8 × 10)]/12 = 124/12 = 10,3 %.

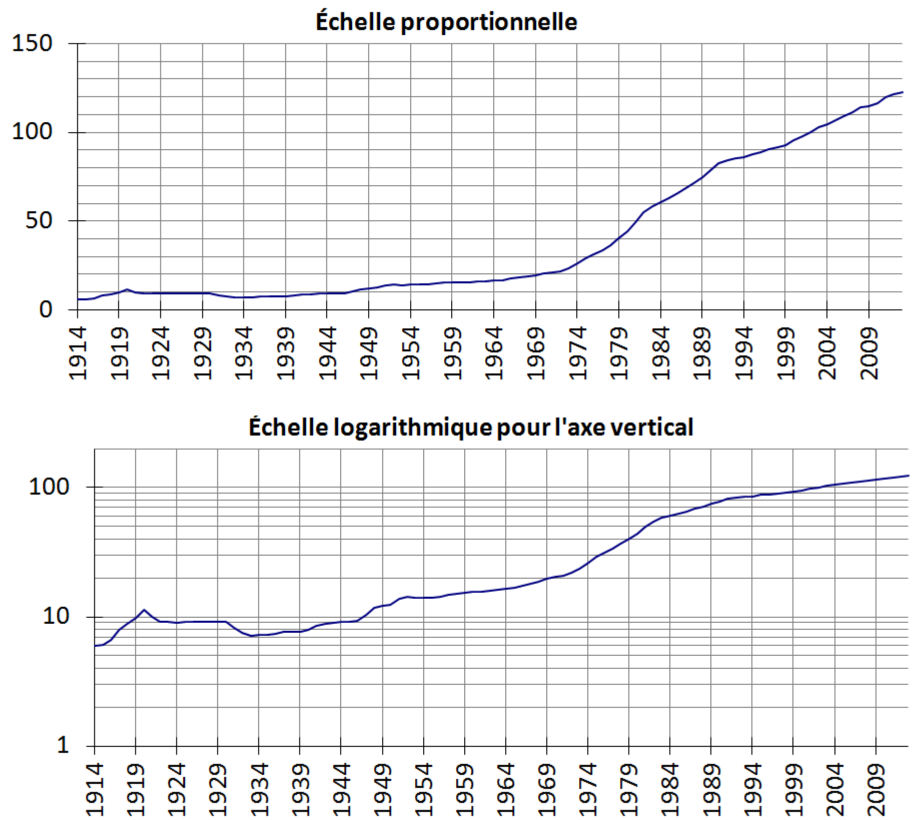
Comme on le voit, ces petits calculs ne présentent pas de grandes difficultés sur le plan mathématique. Ce qui compte, c'est avant tout de bien les choisir. Souvent, la même variable est publiée avec différentes périodicités : il est par exemple courant de trouver le nombre de naissances par mois, par trimestre et par année. Lorsque cela est possible, choisissez la périodicité qui convient le mieux à votre étude.

4.3. L'échelle semi-logarithmique

Comme nous l'avons vu, les courbes sont un des moyens privilégiés pour représenter l'évolution des valeurs d'une variable à travers le temps. Cependant, lorsqu'une série chronologique porte sur une longue période et que les données tendent à augmenter constamment, une telle courbe tend à se déformer.

La première courbe de la figure 4.4, qui représente l'évolution de l'indice des prix au cours du XX^{e} siècle, laisse croire, par exemple, que l'inflation était beaucoup plus forte dans les années 1980 que dans les années 1960 (la pente de la courbe est beaucoup plus abrupte). Mais cette impression est trompeuse, car les prix ont augmenté à peu près au même rythme dans les décennies 1980 et 1960.

Figure 4.4 - Évolution à long terme de l'indice des prix au Canada



Source: Statistique Canada, Cansim. Base : 2002 = 100.

L'échelle logarithmique permet de représenter fidèlement le rythme de progression d'une variable.

Observez, maintenant la deuxième courbe de la figure 4.4. Cette fois, l'échelle de l'axe vertical a été ajustée. Avez-vous remarqué que les marques de graduations 1, 10, 100 sont situées à égale distance les unes des autres? C'est tout simplement qu'entre chacune de ces marques, les données progressent au même rythme : elles sont multipliées chaque fois par 10 (la marque suivante serait 1000). Ce genre d'échelle, appelée *semi*-logarithmique parce qu'elle ne concerne qu'un seul des deux axes, permet de représenter fidèlement l'évolution d'une série chronologique. Grâce à la nouvelle d'échelle, on peut constater clairement que l'indice des prix a beaucoup fluctué entre les deux guerres et que la poussée qu'il a subie dans les années 1970 s'est nettement atténuée par la suite.

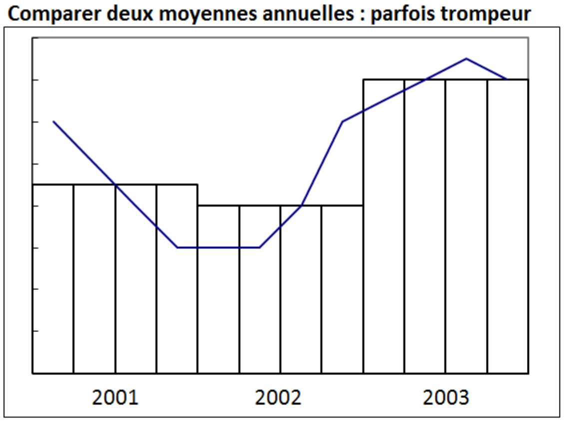
Il n’est pas très difficile de tracer un graphe à échelle logarithmique, surtout si on se sert d’un chiffrier électronique. Mais parfois, cela n’en vaut pas vraiment la peine. Lorsque la période de temps est relativement courte ou que les données ont tendance à baisser autant qu’à monter, un bon vieux graphe traditionnel fait très bien l’affaire.

4.4. Quelques précautions à prendre

Pour conclure, nous vous proposons quelques façons de tricher sur la présentation des données chronologiques. Mais, rassurez-vous, notre but n’est pas de vous corrompre, mais bien de vous aider à démasquer les vrais tricheurs (voir la figure 4.5).

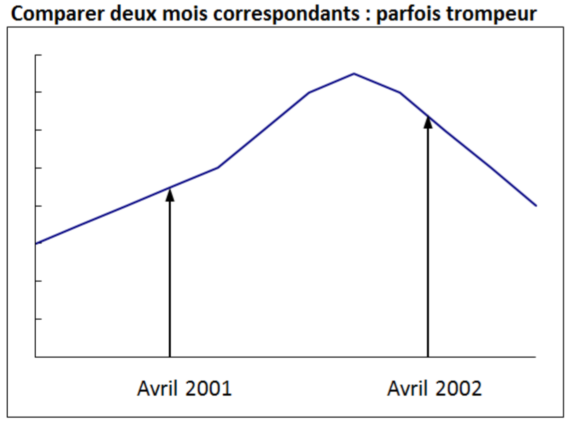
Figure 4.5 - Comment tricher avec les données chronologiques

Comparer deux moyennes annuelles : parfois trompeur



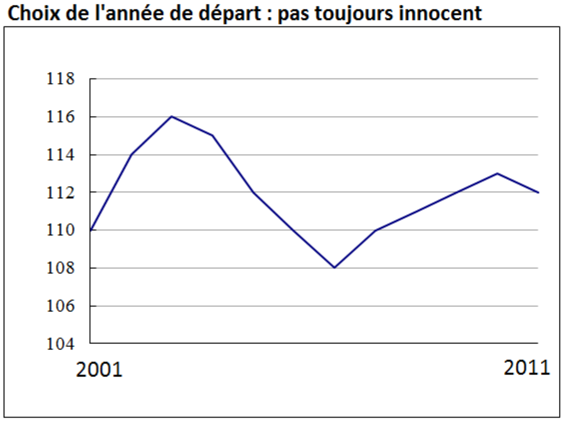
En 2002, les données trimestrielles montrent qu'on est en plein remontée par rapport à 2001. Pourtant, la moyenne annuelle chuté.

Comparer deux mois correspondants : parfois trompeur



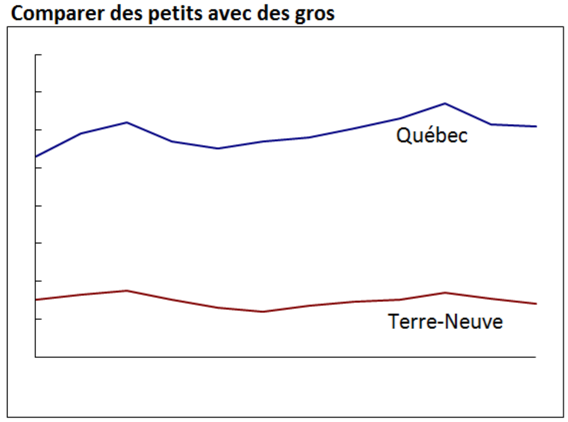
En avril 2002, on semble être en meilleure position que 12 mois plus tôt, alors qu'on est en pleine dégringolade.

Choix de l'année de départ : pas toujours innocent



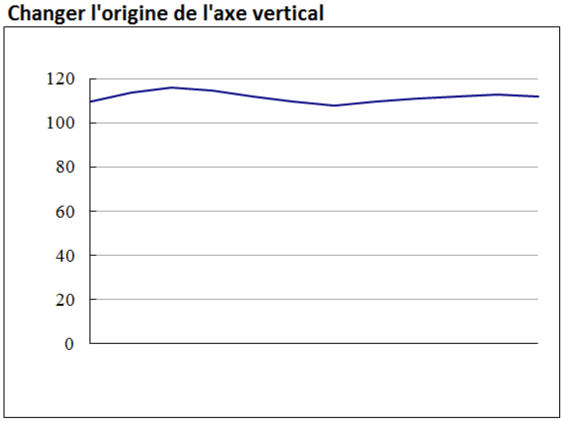
En 2011, la situation semble meilleure qu'au point de départ (2001). Mais 2001 était une mauvaise année alors que 2011 en est une bonne.

Comparer des petits avec des gros



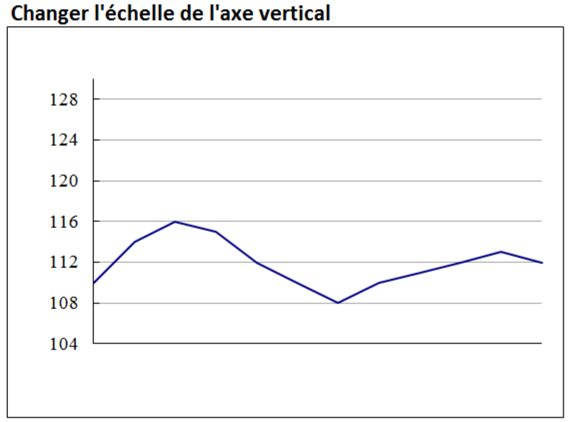
Le Québec fait mauvaise figure : il compte beaucoup plus de chômeurs que Terre-Neuve.

Changer l'origine de l'axe vertical



Ce graphique représente les mêmes données que celui qui est au-dessus, mais le point de départ de l'axe vertical est différent : les valeurs semblent plus grandes et les variations plus faibles.

Changer l'échelle de l'axe vertical



Ce graphique représente les mêmes données que celui du dessus à gauche, mais l'échelle de l'axe vertical est plus serrée : les valeurs et les variations semblent plus faibles.

EXERCICES 4

1. L'inflation en graphiques

Utilisez les tables historiques sur l'inflation pour construire les trois graphiques suivants :

a) Évolution à long terme des indices des prix à la consommation depuis 1913 (échelle semi-logarithmique)

b) Évolution à long terme des indices des prix à la consommation depuis 1913 (échelle non logarithmique)

c) Évolution des taux d'inflation annuels depuis 1950.

EXERCICES SUPPLÉMENTAIRES

1. 1. Le bilan de Société de l'assurance automobile du Québec

Complétez le tableau 4.8.
(*Note :* Les données de 2013 sont fournies à titre de comparaison seulement.)

Tableau 4.8 - Les accidents de la route au Québec

	Nombre de victimes		Taux de variation	
	1993	1994		2013
			(en %)	
Mortelles	982	824		399
Graves	6 517		-0,1	1 727
Légères		42 657		35 972
Total	**50 797**			**38 098**

Source : Société de l'Assurance automobile du Québec.

2. Honnête et… intelligent

a) La coalition au pouvoir, constituée du parti progressiste et du parti conservateur, propose d'accroître les impôts de 20 % et de les diminuer ensuite de 25 %. Vous êtes éditorialiste dans un grand journal et vous êtes tenté d'approuver publiquement la double mesure gouvernementale… mais vous ne le ferez pas. Pourquoi?

b) Vous venez d'obtenir deux hausses de salaire consécutives. La première, de 10 %, souligne votre changement d'ancienneté et la seconde, de 20 %, est due à la promotion qui a suivi votre dernier éditorial sur la coalition au pouvoir. En recevant votre chèque de paie, vous constatez que votre salaire a augmenté de 32 %. Vos collègues, à la moralité douteuse, vous conseillent de ne pas signaler l'erreur au service de la comptabilité. Puisque vous êtes honnête et intelligent, quelle devrait être votre attitude? Pourquoi?

3. Le dollar flotte

a) Sachant que le dollar canadien vaut 0,50 livre, 1,25 florin ou 1000 lires, dites combien valent chacune des devises mentionnées en dollar canadien (chiffres fictifs)

b) Recherche. Trouvez dans le journal le taux de change (en dollars canadiens) du dollar américain, du franc français, du yen japonais et d'une autre devise de votre choix. Calculez le taux de change inverse : la valeur du dollar canadien en dollars américains, etc.

4. La croissance moyenne

a) Le « Grand bond en avant » coïncide avec le deuxième plan quinquennal chinois (1958-62). Pendant cette période, la croissance annuelle moyenne de la production a été de 0,65 %. Quel est le taux de croissance sur l'ensemble de la période du plan? (Source : La Chine au présent, juin 1996.)

b) En Chine, la population masculine passe de 301,8 millions en 1953 à 581,8 millions en 1990. Durant la même période, la population féminine passe de 280,8 millions à 548,7 millions. Quel est le taux de croissance annuel moyen pour les hommes, les femmes et pour l'ensemble de la population? (Source : Zhongguo tongji nianjian 1994.)

c) Selon les prévisions de la Banque Mondiale, la population du Canada devait augmenter de 36 % entre 1970 et 2000. Quel était le taux de croissance annuel moyen prévu entre 1970 et 2000?

5. La revanche des berceaux

À l'aide des chiffres du tableau 4.5 , p. xxx [Évolution de la population du Québec], calculez pour l'année 1995 les flux nets de population et la variation du stock. Comparez les deux résultats.

6. Attention au ratio Q

Selon un article du New York Times repris par Courrier international (13 juin 1996), le ratio Q avoisinerait 1,8. Attention, il ne s'agit ni du ratio tour de hanche/tour de taille, ni du rapport entre le nombre de mâles et le nombre de femelles, ni même d'un cousin éloigné du mystérieux point G, mais plus prosaïquement du rapport entre la valeur des actifs cotés en Bourse et le coût de leur remplacement. Ce ratio de 1,8 signifie que les actifs sont cotés à 80 % au-dessus de leur prix de remplacement. Le niveau normal du ratio Q est de 0,7, ce qui signifie que le marché des actions devrait reculer un jour ou l'autre d'environ 60 %. L'indice Dow Jones pourrait ainsi dégringoler de 5500 à 2500 points. Malgré tout cela, James Tobin, inventeur du ratio Q, et prix Nobel d'économie, ne semble pas pressé de vendre ses actions.

a) Parmi tous les chiffres cités, lequel est un taux de variation?

b) Vérifiez le chiffre de 60 % proposé dans le texte en comparant les deux ratios Q.

c) Vérifiez le même chiffre en comparant les deux indices Dow Jones.

d) Montrez que la prévision de 2500 points pour l'indice Dow Jones a été calculée à partir des trois données suivantes : le ratio Q actuel, le ratio Q normal et l'indice Dow Jones actuel.

e) Comment concilier le fait que le cours des actions représente plus du double de leur valeur normale (ratio de 1,8 contre 0,7 ou indice de 5500 contre 2500) et qu'en même temps on affirme que le marché devrait reculer de seulement 60 %?

7. Le compte des terres du Québec

Le texte suivant est extrait des *Comptes des terres : région de Chaudière-Appalaches* (2014), document publié par l'Institut de la statistique du Québec.

« Dans l'ensemble de la région, les surfaces artificielles ont augmenté d'environ 78 km² entre ~2001 et ~2006. Cette variation représente 15 % de la superficie de ~2001. En contrepartie, les terres agricoles ont perdu à peu près 8 % de leur superficie au cours de cette période, ou environ 256 km², et les milieux humides boisés, environ 6 % ou 55 km². Les autres changements notables concernent les forêts, qui ont eu tendance à « s'enrésiner » : les forêts de conifères et mixtes à couvert fermé ont gagné, respectivement, environ 19 % (410 km²) et 3 % (133 km²), tandis que les forêts de feuillus ont perdu à peu près 12 % (312 km²). [La région de Chaudière-Appalaches possède une superficie totale de 16 130 km².] »

a) Parmi toutes les données que nous avons mises en caractères gras, lesquels représentent des flux, lesquels représentent des stocks et lesquels représentent des taux de variation?

b) À combien s'élève le stock final des forêts de conifères?

c) À combien s'élève le stock final des forêts de feuillus?

d) Quelle est la proportion de forêts de feuillus par rapport à l'ensemble de la région (en 2006)?

8. Le décollage de l'agriculture chinoise

Le tableau 4.9 ci-dessous indique l'évolution de la production agricole chinoise pendant le demi-siècle qui a suivi la révolution de 1949.

Tableau 4.9 - Évolution de la production agricole en Chine

	Céréales	Porc, boeuf, mouton	Fruits
		Production annuelle par habitant en kg	
1952	288	6	4,3
1957	306	6	5,1
1962	240	3	4,1
1965	272	8	4,5
1970	293	7	4,6
1975	310	9	5,9
1978	319	9	6,9
1980	327	12	6,9
1985	361	17	11,1
1986	367	18	12,6
1987	372	18	15,4
1988	356	20	15,1
1989	364	21	16,4
1990	393	22	16,5
1991	378	24	18,9
1992	380	25	21,0
1993	387	27	25,6

Source : Zhongguo tongji nianjian 1994 (中国统计年鉴).

a) Tracez les 3 courbes sur un graphique (à la main ou à l'aide d'un chiffrier électronique).

b) Calculez la croissance annuelle moyenne pour les périodes 1952-1978 et 1978-1993, pour chacune des variables.

c) Recherche : Quels sont les évènements politiques qui ont marqué la fin des années 1950 et la fin des années 1970 en Chine? Comment cela peut-il s'observer sur les courbes?

d) Les variations à court terme et la tendance de fond : Laquelle des trois variables est la plus instable à court terme? Quelles sont les conséquences à long terme de l'implantation de la politique de réforme en 1978?

Sous-questions supplémentaires

e) Tracez une courbe de tendance à long terme (à l'aide de votre chiffrier électronique).

f) Comparez les données à celles d'un autre pays.

9. Laboratoire : le tourisme progresse toujours

Le tableau 4.10 ci-dessous montre que le nombre de visiteurs internationaux est en constante progression. Utilisez un chiffrier électronique pour répondre aux questions suivantes.

Tableau 4.10 - Arrivées et recettes du tourisme international

	Arrivées (millions)	**Recettes** (milliards $US)		**Arrivées** (millions)	**Recettes** (milliards $US)
1960	69	7	1978	267	69
1961	75	7	1979	283	83
1962	81	8	1980	286	105
1963	90	9	1981	289	107
1964	105	10	1982	289	101
1965	113	12	1983	292	102
1966	120	13	1984	319	112
1967	130	14	1985	330	118
1968	131	15	1986	341	142
1969	144	17	1987	367	174
1970	166	18	1988	402	202
1971	179	21	1989	431	218
1972	189	25	1990	459	265
1973	199	31	1991	466	272
1974	206	34	1992	504	309
1975	222	41	1993	518	314
1976	229	44	1994	546	347
1977	249	56	1995	567	373

Source : Organisation mondiale du tourisme (OMT).

a) Calculez le taux de croissance annuel du nombre d'arrivées.

b) Calculez le taux de croissance annuel du nombre d'arrivées en utilisant une moyenne mobile de 4 ans.

c) Calculez le taux de croissance annuel du nombre d'arrivées en utilisant une moyenne mobile de 8 ans.

d) Tracez un graphe représentant l'évolution du nombre d'arrivées au cours des années.

e) Calculez le taux annuel de croissance des recettes.

f) Avancez des hypothèses qui pourraient expliquer pourquoi les recettes fluctuent beaucoup plus que les arrivées.

g) Rendez-vous sur le site de L'OMT afin d'actualiser les données du tableau.

DOSSIER 4 CINQ FEMMES POUR UN HOMME

Avez-vous déjà entendu dire que, à tel ou tel endroit, on compte cinq femmes pour un homme? Ce lieu mythique se trouve tantôt dans une région éloignée (l'Abitibi), tantôt dans une grande métropole (Montréal), ou encore dans une ville pas comme les autres (une capitale administrative comme Québec ou Ottawa). Dans ce dossier, nous vous proposons de tirer les choses au clair et de déterminer si ce paradis — ou cet enfer, selon le point de vue où l'on se place — existe vraiment. Nous en profiterons pour jeter un coup d'œil plus général sur la famille et le mariage.

Il existe bien d'autres idées reçues à propos de la population. L'une d'entre elles porte sur le nombre d'enfants nécessaires pour assurer la pérennité d'une population. Un autre préjugé concerne la proportion de femmes au Québec.

En fin de compte, nous vous proposons la liste de lieux communs suivante, sur laquelle nous vous demanderons de vous prononcer :

- 1. On compte 5 femmes pour un homme dans certaines villes ou régions.
- 2. Il faut 2,1 enfants par couple (en non 2 tout juste) pour assurer le renouvellement des générations.
- 3. Le Québec compte 52 % de femmes.

Deux de ces trois lieux communs relèvent tout simplement du mythe. Nous mettons le lecteur au défi de trouver lesquels. Attention, il ne s'agit pas ici de se disputer sur les chiffres précis, mais de juger du bien-fondé général de ces lieux communs.

Plus de femmes mariées que d'hommes mariés?

Le tableau D4.1 (ci-après) donne une vue d'ensemble de l'état civil des hommes et des femmes âgées de 15 à 34 ans au Canada. Nous nous sommes basés sur deux recensements espacés d'une vingtaine d'années, afin d'observer l'évolution d'une génération à l'autre.

Examinons d'abord les données de 1991 (nous vous laisserons le soin de refaire le même exercice pour 2011). On remarque d'emblée, dans le tableau, que le nombre d'hommes est sensiblement égal au nombre de femmes, avec un léger avantage pour les hommes dans le groupe des 15 à 24 ans : il s'agit là d'une situation tout à fait typique pour l'espèce *Homo sapiens*.

Tableau D4.1 - État civil de certains groupes d'âge au Canada

1991 (en milliers)

	15-24 ans		25-34 ans		Total 15-34 ans	
	Hommes	Femmes	Hommes	Femmes	Hommes	Femmes
Célibataires	1750	1500	840	559	2590	2059
Mariés	185	367	1478	1717	1663	2084
Autres	9	20	102	170	111	190
Total	**1944**	**1887**	**2420**	**2446**	**4364**	**4333**

2011 (en milliers)

	15-24 ans		25-34 ans		Total 15-34 ans	
	Hommes	Femmes	Hommes	Femmes	Hommes	Femmes
Célibataires	2292	2157	1600	1320	3892	3477
Mariés	42	86	663	886	705	972
Autres	6	10	86	142	93	152
Total	**2340**	**2253**	**2350**	**2348**	**4690**	**4601**

Source : Annuaire du Canada, Ottawa, 1994. Recensement du Canada 2011.

Notes : Les célibataires sont ceux et celles qui n'ont jamais été mariés. Les autres sont les personnes séparées, divorcées ou veuves.

Si on y regarde de plus près, on note que les célibataires de ce groupe d'âge sont surtout des hommes et que les personnes mariées sont surtout des femmes. Étant donné que les mariages se font, en 1991, exclusivement entre deux personnes du sexe opposé, le fait peut sembler curieux à première vue.

Les femmes se marient-elles plus jeunes que les hommes? Les femmes se marient-elles avec des hommes plus vieux qu'elles?

Réglons d'abord le cas des hommes de 15 à 24 ans (pour l'année 1991) : si les hommes célibataires cherchaient à se marier exclusivement avec des femmes de leur groupe d'âge, 250 000 d'entre eux ne pourraient trouver de partenaire (soit 1 750 000 – 1 500 000). On est loin de la croyance populaire voulant qu'il existe 5 femmes pour un homme. Apparemment, les hommes sont moins pressés de se marier que les femmes, du moins à cet âge-là. Le même phénomène peut être constaté chez les hommes de 25 à 34 ans, même s'il commence à être temps pour eux de penser au mariage : le « déficit » s'élève alors à 271 000 (soit 840 000 – 559 000).

Si on considère l'ensemble des hommes célibataires de 15 à 34 ans, il leur manque plus d'un demi-million de partenaires. Certes, il y a plus d'hommes que de femmes dans ce groupe d'âge, soit 31 000 hommes de plus que de femmes entre 15 et 34 ans (vous pouvez vérifier dans le tableau D4.1), mais cet écart est somme toute très minime. Il ne reste plus qu'une explication : en admettant que peu d'hommes soient mariés avec des femmes plus vieilles qu'eux, on peut déduire qu'environ un quart des femmes de 15 à 34 ans sont mariées avec des hommes de 35 ans et plus. Qu'en pensez-vous?

Cela dit, est-il plus facile de se trouver un partenaire pour un homme que pour une femme, comme le veut la croyance populaire? Rien, après examen de ces chiffres, ne nous permet de l'affirmer, du moins à l'échelle nationale. On peut en effet déduire du tableau D4.1 qu'il y a, en 2011, au Canada, 98,1 femmes pour 100 hommes dans le groupe des 14 à 35 ans (soit 4601/4690).

Nous avons pu constater, en épluchant les données détaillées du recensement, que la tendance est sensiblement la même à l'échelle régionale pour ce groupe d'âge. On retrouve, par exemple, 96,3 femmes pour 100 hommes à Québec, 99,1 à Montréal et 101,1 à Ottawa. On est loin du ratio de cinq femmes pour un homme!

Les autres types de famille

Pour se marier, il faut être deux, mais pour vivre à deux il n'est pas indispensable de se marier. Les tableaux D4.2 et D4.3 ci-après montrent que la proportion d'unions libres est en augmentation constante dans l'ensemble du Canada, et plus particulièrement au Québec. Par contre, la proportion de familles monoparentales s'est mise à redescendre après 1991.

Tableau D4.2 - Familles monoparentales (en % des familles)

	1981	1986	1991	2011
Terre-Neuve	12,7	14,2	15,9	16,1
Québec	17,6	20,8	21,7	14,2
Ontario	16,3	17,8	19,3	15,1
Colombie-Britannique	17,3	20,1	20,3	13,4
Canada	**16,6**	**18,8**	**20,0**	**14,8**

Source : Annuaire du Canada, Ottawa, 1994. Statistique Canada, Cansim 111-0009.

Tableau D4.3 - Taux de prévalence des unions libres (en % des unions)

	1981	1986	1991	2011
Terre-Neuve	2,2	3,5	9,3	15,6
Québec	8,1	13,7	21,7	37,1
Ontario	5,6	7,2	7,4	13,1
Colombie-Britannique	8,1	9,9	13,5	15,2
Canada	**6,4**	**9,2**	**13,5**	**19,7**

Source : Annuaire du Canada, Ottawa, 1994. Recensement du Canada 2011.

Le mariage entre conjoints de même sexe a été inauguré en 2005 au Canada. Lors du recensement de 2011, une proportion de 0,35 % des couples mariés était constituée de conjoints de même sexe (0,37 % pour les hommes et 0,33 % pour les femmes). La proportion correspondante est un peu plus élevée en ce qui concerne les unions libres, où elle atteint 2,8 % (3,0 % pour les hommes et 2,6 % pour les femmes). (*Source :* Recensement du Canada 2011. Notez bien que ces pourcentages constituent des *stocks* (nombre de couples à telle date), et non des *flux* (nombre d'unions célébrées pendant l'année).

Les familles canadiennes tendent à compter de moins en moins d'enfants.

On dit parfois que les écoles seraient plus faciles à administrer s'il n'y avait pas d'élèves : les gestionnaires pourraient vaquer à leurs occupations sans être constamment dérangés. Dans le même ordre d'idée, les parents n'auraient-ils pas la tâche plus facile s'ils cessaient d'avoir des enfants? Le tableau D4.4 montre qu'on se dirige peut-être dans la bonne direction, puisque le nombre moyen d'enfants par femme est inférieur à 2 au Québec et au Canada.

Tableau D4.4 - Nombre moyen d'enfants par femme

	1990	2001	2011
Terre-Neuve	1,55	1,24	1,45
Québec	1,72	1,47	1,69
Ontario	1,82	1,51	1,52
Colombie-Britannique	1,81	1,38	1,42
Canada	**1,82**	**1,51**	**1,61**

Source : Statistique Canada, Cansim 102-4505.

Vous êtes-vous déjà demandé pourquoi les femmes devraient avoir « un peu plus » de 2 enfants pour que la population se maintienne à un niveau stationnaire?

La chose est véridique, pour assurer le renouvellement de la population, un couple devrait avoir *plus* de 2 enfants (un garçon, une fille et des poussières?). On dit souvent que cela est dû au fait que certaines personnes sont stériles ou passent leur vie dans un monastère, quand on ne donne pas des explications encore plus farfelues. Examinons plutôt le problème de façon logique : pour perpétuer l'espèce, il suffirait que chaque femme donne naissance en moyenne à une fille, ni plus ni moins. Étant donné que la proportion de bébés garçons est légèrement supérieure à celle de bébés filles et que certaines femmes meurent avant d'atteindre leur maturité, un nombre moyen de 2 enfants par femme en âge de procréer serait insuffisant pour atteindre cet objectif.

Je t'épouse… un peu, beaucoup, souvent

La plupart des jeunes mariés se recrutent parmi les célibataires. Cependant, au fil des ans, la proportion de mariages dans lequel l'un ou l'autre des partenaires a déjà convolé auparavant tend à augmenter. C'est ce qu'indique le tableau D4.5. On y constate que si le nombre total de mariages a diminué en 20 ans, le nombre de remariages a augmenté. Seriez-vous capable, en observant le tableau, de déterminer le nombre de mariages impliquant un homme jamais marié et une femme ayant déjà été mariée au moins une fois?

Tableau D4.5 - Nombre de mariages au Canada selon l'état civil des fiancés

	1971	1991
	(en milliers)	
Tous les mariages	191	172
L'homme n'a jamais été marié	169	131
La femme n'a jamais été mariée	169	133
Un des deux partenaires a déjà été marié	32	56
Les deux partenaires ont déjà été mariés	13	24

Source : Annuaire du Canada, Ottawa, 1994.

Après avoir nous-mêmes longtemps réfléchi à la question, nous nous sommes rendu compte que les cinq lignes du tableau 5 cachaient en réalité quatre catégories *exclusives* : [homme jamais marié + femme jamais mariée], [homme déjà marié + femme jamais mariée], [homme jamais marié + femme déjà mariée], [homme déjà marié + femme déjà mariée]. En partant de cette constatation, nous avons construit le tableau D4.6 qui contient (en dehors des totaux) 2 colonnes et 2 lignes. Nous avons inscrit dans ce tableau les données que nous connaissions déjà grâce au tableau D4.5 (en caractères gras). Il a suffi ensuite de boucher les trous à l'aide de simples soustractions.

Tableau D4.6 - Répartition des mariages au Canada selon l'état civil des fiancés

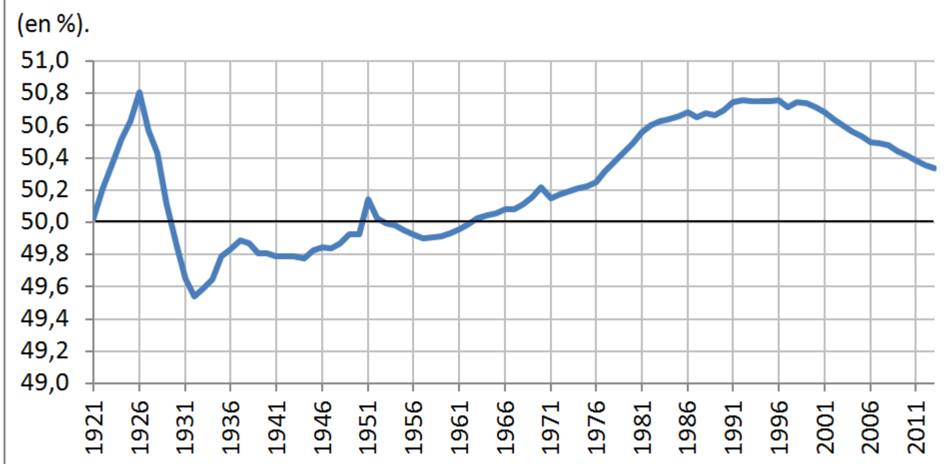
(en 1991)		Hommes		
		Jamais marié	Déjà été marié	**Total**
Femmes	Jamais mariée	116	17	**133**
	Déjà été mariée	15	**24**	39
	Total	**131**	41	**172**

Note : données dérivées du tableau D4.5.

La statistique quasi officielle voulant que le Québec compte 52 % de femmes relève-t-elle du mythe ou de la réalité?

Pour les gens « bien informés », il va de soi qu'au Québec, et probablement dans la plupart des pays normaux, les femmes constituent la majorité de la population. Il existe même un chiffre quasi officiel, que la presque totalité des personnes théoriquement éclairées sur la question vous confirmera d'emblée : le Québec compte 52 % de femmes. Cette donnée a, par exemple, été mise de l'avant lors d'une émission à Radio-Canada (23 mars 2007) par Gabriel Chèvrefils, représentant de Québec solidaire, « le seul parti à présenter 52 % de candidates aux élections ». C'est déjà plus qu'une statistique, c'est un véritable programme politique. Or, il se trouve que cette proportion n'a jamais dépassé 50,8 % depuis un siècle, comme en peut le constater en examinant la figure D4.1.

Figure D4.1 - Évolution de la proportion de femmes au Québec



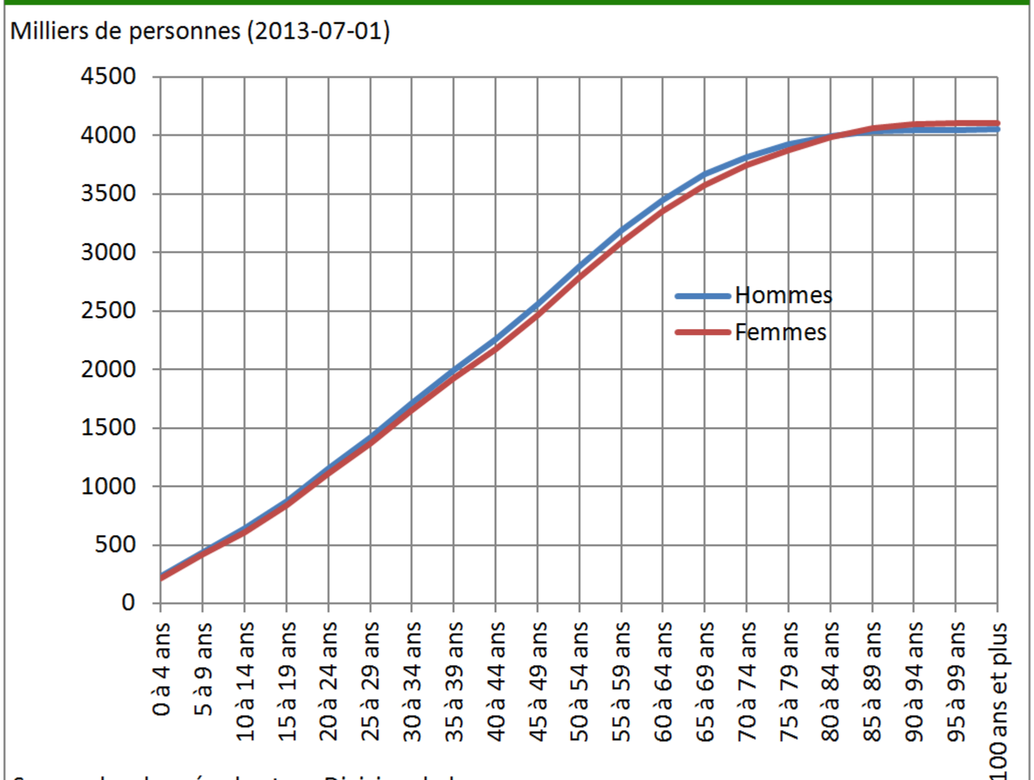
Source des données brutes : Statistique Canada, Cansim 051-0026.

Les femmes ont même déjà été « minoritaires » au Québec, du moins jusqu'en 1963. De fait, il naît plus d'hommes que de femmes chez l'*Homo sapiens* (environ 105 contre 100), et les femelles de l'espèce vivent plus longtemps que les mâles. La proportion de femmes s'est accrue au Québec jusqu'en 1997, où elle a atteint 50,76 %. Par la suite, la réduction de l'écart entre l'espérance de vie des hommes et des femmes a entraîné un retournement de tendance (la proportion de femmes était tombée à 50,33 % en 2013).

Notre propos n'est pas ici de minimiser l'importance des femmes dans la population. De toute façon, les différences sont si minces que l'on peut considérer, au bout du compte, que les deux sexes sont à égalité. C'est plutôt le chiffre magique de 52 % qui nous intéresse en tant que phénomène psychologique, social et politique. D'où sort ce fameux chiffre? Quel rôle joue-t-il? Comment se fait-il qu'il fasse l'objet d'une telle unanimité tout en ne reposant sur aucune réalité?

Pour examiner la situation sous un autre angle, nous vous proposons, dans la figure D4.2, une façon originale de représenter la pyramide des âges. On y retrouve les fréquences *cumulées* de chaque groupe d'âge. Aussi étonnant que cela puisse paraître, il y a, au Québec, autant d'hommes que de femmes dans le groupe d'âge constitué des 84 ans et moins (3,998 millions d'hommes et 3,988 millions de femmes en 2013, pour être plus précis). Ce n'est qu'en comptant les personnes âgées de 85 ans et plus que la proportion de femmes dépasse celle des hommes.

Figure D4.2 - Population du Québec par classe d'âge cumulée et par sexe



Source des données brutes : Division de la démographie, Statistique Canada.

Pour schématiser la chose, on pourrait diviser la population du Québec en trois groupes : les enfants (0 à 19 ans), les personnes d'âge moyen (20 à 64 ans) et les personnes âgées (65 ans et plus). En 2013 (année correspondant aux données de la figure D4.2), les hommes sont légèrement majoritaires dans les deux premiers groupes (respectivement 51,0 % et 50,6 %), et sensiblement minoritaires dans le troisième (44,3 %).

QUESTIONS

1. Les jeunes célibataires (tableau D4.1, année 1991)

a) Quelle est la proportion de femmes mariées parmi les 15 à 24 ans?

b) Quelle est la proportion d'hommes mariés parmi les 15 à 24 ans?

c) Quelle est la proportion de femmes parmi les personnes de 15 à 24 ans qui sont mariées?

d) Quelle est la proportion d'hommes parmi les personnes de 15 à 24 ans qui sont mariées?

e) Quelle est la proportion d'hommes parmi les 15 à 24 ans?

f) Quel est le ratio de masculinité (rapport homme/femme) parmi les 15 à 24 ans?

2. Les concubins (tableau D4.2)

Tracez un graphe illustrant l'évolution du taux de prévalence des unions libres entre 1981 et 2011 (tracez sur le même graphe quatre courbes représentant les quatre provinces).

3. Deux enfants et quelques (tableau D4.4)

Comment expliquez-vous que la moyenne canadienne soit de 1,82 enfant par femme en 1990 alors que toutes les données citées sont inférieures ou égales à ce chiffre?

4. Mariés et remariés (tableaux D4.5 et D4.6)

a) Calculez le taux de croissance du nombre de mariages entre 1971 et 1991.

b) Quel est, en 1971, le nombre de mariages impliquant deux partenaires jamais mariés auparavant (si nécessaire, construisez un tableau similaire au tableau D4.6)?

CHAPITRE 5 LES INDICES

TABLE DES MATIÈRES

1. Qu’est-ce qu’un indice?

2. Les indices synthétiques

3. L’indice du développement humain

4. L’indice des prix à la consommation

• Exercices supplémentaires

• Dossier

Vous avez peut-être déjà entendu le gouvernement d’Ottawa claironner que le Canada est le meilleur pays au monde en matière de « développement humain ». Pour le prouver, on cite le score attribué par un organisme des Nations-Unies, le PNUD (Programme des Nations Unies pour le développement). Le Canada s’est classé en tête à plusieurs reprises, devant les États-Unis et le Japon avec une note dépassant 0,93, alors que le Mexique se retrouvait en milieu de peloton avec une note tournant autour de 0,80, et que le Niger occupait la dernière place avec une note variant de 0,2 à 0,3 (théoriquement, la note maximum est de 1 et la note minimum de 0). Ces scores sont ce qu’on appelle des *indices*, c’est-à-dire une manière de représenter de façon chiffrée une combinaison d’indicateurs représentatifs d’une réalité humaine.

L’usage des indices en sciences humaines est assez répandu, ne serait-ce que parce qu’ils sont pratiques. Cependant, la manière de construire certains indices comporte une dose d’arbitraire, car l’être humain ne se traduit pas facilement en équations, et c’est bien normal! Voilà quelques bonnes raisons d’étudier ce que sont les indices, et comment on les construit.

Nous reviendrons plusieurs fois, au cours de ce chapitre, sur l’indice du développement humain (IDH) des Nations Unies, mais soulignons dès à présent les points suivants. D’une part, si on calculait l’IDH du Québec, on obtiendrait le même score que pour le Canada. D’autre part, on constate que c’est à cause d’un plus faible niveau de scolarisation que le Japon se fait dépasser par le Canada : voilà qui est surprenant… et qui vaudra la peine d’être tiré au clair.

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

- Comment peut-on utiliser des indices simples pour faire des comparaisons dans le temps et dans l’espace?
- Comment peut-on combiner un ensemble d’indices simples pour en faire des indices synthétiques?
- Les indicateurs utilisés pour construire un indice donné ont-ils été choisis de façon objective ou de façon arbitraire?
- Quelles sont les limites des indices?

- Comment les indices d'inflation permettent-ils de mesurer adéquatement l'évolution de la valeur des choses?

1. QU'EST-CE QU'UN INDICE?

Le développement humain, pour reprendre notre exemple, est un *concept* très vaste et multidimensionnel. L'indice du développement humain (IDH) n'est qu'une tentative de représenter ce concept sous forme chiffrée. Pour passer du concept à l'indice, il faut d'abord bien cerner le concept lui-même ainsi que ses dimensions. Selon les Nations-Unies, le développement humain consiste à « élargir le champ des possibilités de l'individu ». Pour qu'un individu ait la possibilité de se développer, il doit donc posséder trois grands atouts : une bonne santé, une instruction de base et un niveau de vie décent. Il reste à trouver des indicateurs chiffrés qui reflètent de façon plus concrète ces atouts. Dans le cas de l'IDH, ces indicateurs pourraient, par exemple, être l'espérance de vie (santé), le taux d'alphabétisation (instruction) et le revenu par habitant (niveau de vie). Il ne resterait plus qu'à évaluer l'importance relative de chacun de 3 indicateurs et de les combiner en un seul chiffre : l'*indice* de développement humain.

L'indice est une combinaison d'indicateurs chiffrés qui représentent les diverses dimensions d'une réalité humaine (ou *concept*).

Le mot *indice* a la même origine que le mot *index*. L'index est en effet le doigt qui *indique*, le doigt qui *dit* ce qui est important ou significatif. En sciences humaines, l'indice est la représentation d'un ou plusieurs des éléments les plus significatifs (les *indicateurs*) qui reflètent le concept étudié. Les indicateurs doivent être choisis avec soin : il faut qu'ils soient à la fois représentatifs du concept et faciles à obtenir.

1.1. Le point de référence

Les indices servent avant tout à faire des *comparaisons* dans le temps et dans l'espace.

L'indice permet donc de représenter un concept humain en un seul chiffre. L'IDH, par exemple, sert à comparer les pays entre eux, l'indice boursier sert à comparer le cours des actions dans le temps. Même s'il existe plusieurs sortes d'indices (et la plupart sont d'ailleurs plus simples que l'IDH), tous les indices ont un point commun : ils servent à faire des *comparaisons*.

Contrairement aux données brutes, les indices ne sont pas mesurés avec des unités absolues (comme le dollar, la tonne, le nombre d'individus, etc.). Puisque les indices servent à faire des comparaisons, les valeurs qu'ils prennent doivent s'interpréter par rapport à un *point de référence* choisi.

(200 x 130,7)/100 = 261,4

Voici un premier exemple du point de référence d'un indice. En 2013, l'indice des prix à la consommation (base 2002 = 100) est de 122,8 au Canada. Cet indice n'est pas exprimé en dollars. Il indique seulement le chemin parcouru depuis l'année de référence, qui est ici 2002 et pour laquelle on a choisi une valeur arbitraire et commode (le chiffre 100). L'indice nous dit simplement que, toutes proportions gardées, un panier de provisions typique qui coûtait 100 $ en 2002 coûterait 122,8 $ en 2013. On pourrait également dire qu'un panier qui coûtait 200 $ en 2002 coûterait 245,6 $ en 2013 (car 122,8/100 = 245,6/200).

La base est la valeur attribuée au point de référence de l'indice (par exemple : 2002 = 100).

Dans l'exemple que nous venons de donner, le point de référence de l'indice est l'année 2002, qui se voit attribuer une valeur de 100. Cette valeur est appelée la *base* de l'indice. Toute valeur supérieure à 100 signifie que les prix ont augmenté et toute valeur inférieure à 100 indiquerait que les prix ont baissé.

Revenons à l'exemple de l'IDH : quel est le point de référence? En réalité, il y a ici deux points de référence : le minimum (qui vaut 0) et le maximum (qui vaut 1). Les valeurs que prend l'indice seront interprétées en fonction de ces deux points. On pourra dire notamment que le Canada (avec 0,950, par exemple) est très près du maximum et ne devance le Japon (0,937) que de très peu. Cet indice un peu particulier fera l'objet d'une étude plus détaillée dans la section 3 de ce chapitre.

On remarque que les points de référence sont toujours des chiffres ronds, ce qui facilite les comparaisons. Le point de référence le plus courant est 100. Dans le cas de l'IDH, on aurait pu utiliser une échelle de 0 à 100 (au lieu d'une échelle de 0 à 1). Le Canada aurait alors obtenu 95 et le Japon 93,7. Ce n'est finalement qu'une question de présentation.

1.2. Une comparaison dans l'espace

Un indice élémentaire est basé sur un seul indicateur.

De nombreux indices sont basés sur un indicateur unique. On les appelle les indices élémentaires. Leur but principal est de permettre des comparaisons commodes entre époques ou endroits différents. Voici deux exemples d'indices élémentaires : le premier permet de comparer les niveaux de vie entre les régions de l'Union européenne; grâce au second, on peut examiner l'évolution de la population de certaines provinces canadiennes à travers le temps.

Au cours de la décennie 1980, la Grèce, l'Espagne et le Portugal se joignent à l'Europe des Neuf (future Union européenne). Une dizaine d'années plus tard, les disparités entre ces pays et les autres pays membres sont toujours aussi marquées. Le quotidien londonien *Financial Times* reprend alors les résultats d'une étude de l'organisme statistique de l'Union européenne (Eurostat) sur le niveau de vie moyen dans les régions européennes. Curieusement, les données ne sont pas exprimées en livres, ni en dollar, ni dans une quelconque monnaie européenne, mais en *indices*.

Tableau 4.7 - Les régions riches et pauvres d'Europe

	Indices du PIB par habitant, États-Unis = 100
Les sept régions les plus riches	
Hambourg	196
Bruxelles	174
Darmstadt	174
Paris	169
Vienne	166
Bavière	157
Brême	155
Les régions pauvres	
Aucune région de la Grèce et du Portugal n'atteint le niveau des États-Unis (100) et en Espagne, seule la région des Baléares dépasse le score de 100.	

Source : Financial Times, 22 février 1995. Données de 1992.

Pourquoi des indices? Prenons le problème dans son ordre logique. Le concept à mesurer est le niveau de vie : le PIB (produit intérieur brut) par habitant semble un indicateur approprié et suffisant. Toutefois, on ne peut fournir aux lecteurs la simple la liste du PIB par habitant dans la monnaie de chaque région concernée. Il faut rendre cet indicateur plus *présentable*. Tout d'abord, le PIB par habitant est converti dans une devise commune (le dollar américain). C'est mieux, mais ce n'est pas assez, car d'une année à l'autre le pouvoir d'achat de toute monnaie varie au gré des prix. En fait, la monnaie, contrairement au mètre ou au baril de pétrole, est un étalon de mesure variable. Si, au lieu d'exprimer le PIB par habitant en dollars, on le comparait à une valeur de référence bien concrète (le PIB par habitant des États-Unis, par exemple), les données seraient bien plus faciles à comprendre. On pourrait même alors comparer des situations d'époques très différentes en matière de pouvoir d'achat et de niveau de richesse.

Les indices du tableau 5.1 peuvent alors se lire ainsi : le PIB par habitant est presque deux fois plus élevé à Hambourg qu'aux États-Unis (196 par rapport à 100) et les 7 régions d'Europe les plus riches dépassent largement la moyenne américaine (pour l'année 1992).

Comment se calcule cet indice élémentaire?

Pour calculer ces indices, les spécialistes d'Eurostat avaient besoin des données brutes sur chaque région. Ils avaient besoin, par exemple, de savoir qu'en 1992 le PIB moyen par habitant était de 23 830 $ aux États-Unis (lieu de référence), de 46 707 $ dans la région de Hambourg et de 40 273 $ dans la région parisienne. Ils ont ensuite choisi comme base de l'indice la valeur 100. On peut constater que l'indice est beaucoup plus explicite que les valeurs en dollars.

Indice du lieu X = (Valeur du lieu X/Valeur du lieu de référence) × Base

Indice de Hambourg = (46 707/23 830) × 100 = 1,96 × 100 = 196

Dans cet exemple, on remarque que l'indice est composé d'un seul indicateur (c'est un indice élémentaire) et qu'on a choisi un point de référence dans l'espace (États-Unis = 100).

1.3. Une comparaison dans le temps

Nous venons de voir un indice élémentaire (composé d'un seul indicateur) basé sur un point de référence dans l'espace. Voyons maintenant un autre indice élémentaire dans lequel le point de référence est situé dans le temps (généralement l'année).

La moitié gauche du tableau 5.2 indique la population de certaines provinces canadiennes depuis 1951. Il s'agit de données brutes : toute l'information s'y trouve, mais elle ne ressort peut-être pas de manière frappante. La moitié droite du tableau reprend les mêmes données sous forme d'indice*. L'année de base est 1951 et la base choisie est égale à 100. Cette fois, la réalité saute aux yeux. Si le Québec suit relativement bien l'Ontario jusqu'au milieu des années 1960, il se met par la suite à perdre rapidement du terrain.

Tableau 5.2 - Répartition de la population par province au Canada

	(En milliers)				(Indice 1951=100)			
1er juillet	Québec	Ontario	Colombie-Britannique	Terre-Neuve	Québec	Ontario	Colombie-Britannique	Terre-Neuve
1951	4 056	4 598	1 165	361	100	100		
1956	4 628	5 405	1 399	415	114	118		
1961	5 259	6 236	1 629	458	130	136		
1966	5 781	6 961	1 874	493	143	151		
1971	6 156	7 768	2 250	535	152	169		
1976	6 410	8 431	2 545	564	158	183		
1981	6 568	8 838	2 836	577	162	192		
1986	6 734	9 477	3 020	578	166	206		
1991	7 081	10 471	3 380	580	175	228		
1996	7 247	11 083	3 874	560	179	241		
2001	7 396	11 897	4 077	522	182	259		
2006	7 632	12 662	4 242	511	188	275		
2011	8 008	13 264	4 499	525	197	288		
2013	8 155	13 538	4 582	527	201	294		

Source des données brutes : Statistique Canada, Cansim 051-0001.

Nous vous laisserons le soin de calculer les indices de Terre-Neuve et de la Colombie-Britannique en exercice.

Indice à l'année X = (Valeur l'année X/Valeur à l'année de référence) × Base

Indice en 1980 (au Québec) = (6 568/4 056) × 100 = 1,62 × 100 = 162

Dans cet exemple, on remarque que l'indice est composé d'un seul indicateur (c'est un indice élémentaire) et qu'on a choisi un point de référence dans le temps (1951 = 100).

EXERCICES 1

1. L'unité dans la disparité

Le tableau 5.3 représente un indice élémentaire.

Tableau 5.3 - Le revenu des particuliers par province au Canada

Revenu personnel par habitant par rapport à la moyenne canadienne

(Indice Canada = 100)

	Québec	Ontario	Colombie-Britannique	Terre-Neuve
1950	85,9	121,2	122,2	51,1
1955	86,5	119,5	123,1	53,2
1960	87,1	117,9	115,1	55,6
1965	90,1	117,0	113,2	59,0
1970	89,6	119,0	107,6	62,8
1975	92,2	110,6	108,6	68,5
1980	96,0	106,4	112,5	67,1
1985	94,2	109,1	101,9	68,9
1990	92,7	113,4	101,1	71,4

Source des données brutes : Statistique Canada, Cansim 380-0050.

a) Quel est le concept étudié? Quel est l'indicateur choisi?

b) Le point de référence de l'indice est-il situé dans le temps ou dans l'espace? Quelle est la base de l'indice?

c) Question plus avancée : commentez l'évolution du revenu des Québécois au fil des années. Votre commentaire doit montrer que vous savez comment lire les indices fournis.

2. Le Québec perd du terrain?

Les questions portent sur le tableau 5.2.

a) Complétez le tableau 5.2 en calculant les indices de Terre-Neuve et de la Colombie-Britannique.

b) Le point de référence de l'indice est-il situé dans le temps ou dans l'espace? Quelle est la base de l'indice?

2. LES INDICES SYNTHÉTIQUES

Un indice synthétique est basé sur plusieurs indicateurs. Il combine plusieurs indices élémentaires.

Certains phénomènes humains ne peuvent être représentés par un indicateur unique. Le coût de la vie dépend du prix d'un grand nombre de produits différents et ces prix peuvent varier de façon indépendante : un indice des prix doit tenir compte d'un ensemble représentatif de produits. À la bourse, certaines actions montent tandis que d'autre baissent, certaines compagnies pèsent lourd et d'autres non : l'indice boursier doit tenter de refléter le phénomène dans son ensemble. La fréquence des accouchements dépend en bonne partie de l'âge de la mère : l'indice de fécondité doit tenir compte des femmes de tout âge. Tous ces exemples concernent des indices synthétiques, c'est à dire des indices combinant plusieurs indices élémentaires.

Au fond, l'indice synthétique n'est que la moyenne *pondérée* d'un ensemble d'indices élémentaires. Avant de construire notre premier indice synthétique, nous devons d'abord régler deux questions : comment évaluer ces pondérations et comment tenir compte du fait que ces pondérations peuvent évoluer avec le temps?

2.1. La pondération des indicateurs

Les indicateurs doivent refléter la réalité tout en étant commodes à obtenir.

Lors de la construction de l'indice synthétique, il faut d'abord savoir bien choisir les indicateurs. Ces derniers doivent être peu nombreux (pour des questions de coût et de simplicité) et faciles à obtenir pour la période, la région ou le domaine étudiés. Par ailleurs, il est nécessaire d'évaluer l'importance relative de chaque indicateur : nous l'avons dit, certains pèsent plus lourd que d'autres.

On remarque que chacun des indices synthétiques que nous avons énumérés au début de cette section regroupe des indices élémentaires de même nature. L'indice des prix, par exemple, est constitué d'une combinaison de prix de divers produits. Les prix sont tous exprimés dans la même unité (le dollar) et chacun fait *partie* d'un *tout* (le coût total des dépenses). Dans un tel cas, la pondération de l'élément correspond à la proportion qu'il représente dans le total : on divise la partie par le tout. Cette pondération représente l'importance relative de l'élément dans l'indice synthétique.

L'évaluation des pondérations qu'il faut allouer à chaque élément de l'indice synthétique dépend des circonstances.

Dans une situation complexe, on est souvent obligé de se contenter d'un nombre limité d'indices élémentaires pour construire un indice synthétique. Cela est correct dans la mesure où les indices élémentaires choisis sont suffisamment représentatifs. Dans un tel cas, les pondérations ne peuvent être trouvées par un simple calcul de proportion. Il faudra avoir recours à d'autres méthodes. Une enquête montrera, par exemple, que le prix du bœuf « pèse » près de 3 fois plus lourd que celui du porc dans le budget d'un ménage canadien typique (selon l'enquête de 2011).

Enfin, lorsque l'indice synthétique regroupe des indices élémentaires de nature différente, les pondérations ne peuvent plus résulter d'un simple calcul. Leur évaluation doit faire l'objet d'un choix éclairé. C'est le cas pour l'indice du développement humain que nous étudierons dans la prochaine section. Pour le moment, nous nous en tiendrons à des cas où les éléments qui composent l'indice synthétique sont de nature semblable.

2.2. L'effet de structure

Beaucoup d'indices sont basés sur des données chronologiques. Or, avec le temps, tout peut évoluer, notamment l'importance relative de chaque élément. L'exemple fictif ci-dessous illustre à la fois la simplicité de l'indice synthétique et les quelques précautions à prendre lors de sa construction et de son interprétation. Il s'agit d'une entreprise qui compte 200 employés : des chercheurs et des techniciens. Notre objectif final est de construire un indice de salaire moyen basé sur l'année 1 (colonne 9 du tableau 5.4). Cet indice est *synthétique*, car il doit tenir compte de deux indicateurs, qui représentent les catégories de salariés.

Tableau 5.4 - L'effet de structure

	Chercheurs			**Techniciens**			**Ensemble des employés**		
	Salaire unitaire (en $)	Nombre d'employés	Masse salariale (en $)	Salaire unitaire (en $)	Nombre d'employés	Masse salariale (en $)	Nombre d'employés	Masse salariale (en $)	Salaire moyen (en $)
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Année 1	800	100	80 000	400	100	40 000	200	120 000	600
Année 2	900	60	54 000	440	140	61 600	200	115 600	578
Taux de variation (%)	12,5	-40,0	-32,5	10,0	40,0	54,0	0,0	-3,7	-3,7
Indice (année 1 = 100)	112,5	60,0	67,5	110,0	140,0	154,0	100,0	96,3	96,3

(Données fictives)

Le salaire moyen des employés est obtenu en faisant la moyenne *pondérée* des salaires de chaque catégorie d'emploi à l'année 1. Les chercheurs gagnent 800 $ (par semaine) et les techniciens 400 $. Les pondérations reflètent la proportion de chaque catégorie de travailleurs au sien de l'entreprise.

Pondération d'une catégorie = Nombre de travailleurs de la catégorie / Nombre total de travailleurs.
Pondération des chercheurs (année 1) = 100/200 = 0,5
Pondération des techniciens (année 2) = 100/200 = 0,5
Salaire moyen (année 1) = (Pondération des chercheurs × Salaire des chercheurs) + (Pondération des techniciens × Salaire des techniciens)
Salaire moyen (année 1) = (0,5 × 800 $) + (0,5 × 400 $) = 400 $ + 200 $ = 600 $

À l'année 1, il y a autant de chercheurs que de techniciens : le salaire moyen aurait donc pu être obtenu en faisant une moyenne *simple* des deux niveaux de salaire.

Salaire moyen (année 1) = (Salaire des chercheurs + Salaire des techniciens) / 2
Salaire moyen (année 1) = (800 $ + 400 $)/2 = 600 $

À l'année 2, la proportion de chercheurs et de techniciens n'est plus la même : pour obtenir le salaire moyen, il faut recalculer les pondérations.

Pondération des chercheurs = 60/200 = 0,3
Pondération des techniciens = 140/200 = 0,7
Salaire moyen (année 2) = (0,3 × 900 $) + (0,7 × 440 $) = 270 $ + 308 $ = 578 $

Il y a une autre manière de calculer le salaire moyen : il suffit de diviser la masse salariale totale de l'entreprise par le nombre total d'employés. La masse salariale est le montant global de salaire versé à un groupe de travailleurs.

Masse salariale des chercheurs = Salaire des chercheurs × Nombre de chercheurs
Masse salariale des techniciens = Salaire des techniciens × Nombre de techniciens
Masse salariale totale = Masse salariale des chercheurs + Masse salariale des techniciens
Salaire moyen = Masse salariale totale/Nombre total d'employés
Salaire moyen (année 1) = [(800 $ × 100) + (400 $ × 100)]/200 = (80 000 $ + 40 000 $)/200 = 120 000 $/200 = 600 $

Étrange paradoxe : le salaire moyen diminue alors que chaque catégorie d'employé voit son salaire augmenter!

Entre l'année 1 et l'année 2, le salaire de chaque catégorie d'emploi augmente. En effet, celui des chercheurs passe de 800 $ à 900 $ (soit une augmentation de [900 - 800]/800 = 0,125 = 12,5 %) et celui des techniciens passe de 400 $ à 440 $ (soit une augmentation de 40/400 = 10 %). Normalement, on devrait s'attendre à ce que le salaire moyen ait lui aussi augmenté. Pourtant, le salaire moyen (colonne 9 du tableau) passe de 600 $ à 578 $: il baisse de 3,66 % ([578 - 600]/600 = -22/600 = -3,66 %). Un observateur de mauvaise foi aurait beau jeu d'accuser l'entreprise d'exploiter sa main-d'œuvre. Quant à nous, il nous faut tirer les choses au clair.

Entre l'année 1 et l'année 2, il n'y a pas que les salaires qui changent. La *structure* d'emploi se modifie également. L'entreprise compte relativement plus de techniciens (les moins bien payés) et moins de chercheurs (les mieux payés) à la fin qu'au début. Cet exemple illustre la difficulté de comparer des moyennes à deux moments différents. Le poids de chaque élément (les pondérations) peut alors se modifier.

2.3. Des indices élémentaires à l'indice synthétique

Pour contourner le problème du changement de structure, nous construirons notre indice synthétique à partir des indices élémentaires. Continuons avec notre exemple des salaires. L'indice synthétique représente le salaire moyen dans l'entreprise et les indices élémentaires représentent le salaire de chaque catégorie d'emploi. (Nous utiliserons ici des indices à base 100.)

Indice élémentaire à l'année N = (Valeur à l'année N/Valeur à l'année de base) × Base

Indice du salaire des chercheurs à l'année 2 = (900/800) × 100 = 112,5
Indice du salaire des techniciens à l'année 2 = (440/400) × 100 = 110

Pour obtenir l'indice synthétique, nous faisons la moyenne pondérée des indices élémentaires. Toutefois, nous disposons de deux ensembles de pondérations : celles de l'année 1 et celles de l'année 2. Il va falloir faire un *choix*. Il paraît plus logique d'utiliser les pondérations de l'année de base et de considérer que la structure n'a pas changé que de faire l'inverse. Il est en effet difficile d'imaginer que les pondérations actuelles s'appliquent rétroactivement à la situation initiale. Mais tout cela n'est qu'une question de convention. Le tricheur choisira l'une ou l'autre variante de l'indice synthétique pour embellir la réalité. La personne avisée s'assurera de comparer des choses comparables et de démasquer les tricheurs.

Indice synthétique = (Pondération 1 × Indice de l'élément 1) + (Pondération 2 × Indice 2) + etc.

Indice synthétique = (Indice élémentaire des chercheurs × Pondération des chercheurs) + (Indice élémentaire des techniciens × Pondération des techniciens)

Indice de Laspeyres

Indice synthétique (pondérations de l'année 1) = (112,5 × 0,5) + (110 x 0,5) = 111,25

Indice de Paasche

Indice synthétique (pondérations de l'année 2) = (112,5 × 0,3) + (110 x 0,7 = 33,75 + 77 = 110,75)

Ces deux variantes de l'indice synthétique portent le nom de leur glorieux inventeur. Dans la grande majorité, des cas, on utilise la première variante et on se contente de parler d'indice synthétique. Mais si, en fouillant un jour dans des statistiques, vous rencontrez ces noms, vous saurez qu'ils ne cachent rien d'effrayant.

Si on constate un léger écart entre les deux indices, on s'aperçoit néanmoins que les résultats sont réalistes. Sachant que les augmentations de salaire par catégorie d'emploi varient entre 12,5 % (pour les chercheurs) et 10 % (pour les techniciens), on s'attend normalement à ce que l'augmentation du salaire moyen soit comprise entre ces deux extrêmes. Et en effet, dans le premier cas (indice synthétique basé sur les pondérations de l'année 1), les salaires augmentent de 11,25 % entre les deux périodes (l'indice passe de 100 à 111,25) et dans le second cas, ils augmentent de 10,75 %. De toute façon, ces indices synthétiques nous montrent bien que l'entreprise en question paie mieux ses employés qu'au point de départ, même si la masse salariale distribuée a diminué.

2.4. Les sorties au cinéma ne sont plus ce qu'elles étaient

Pour vous permettre de vérifier si vous avez bien compris ce qu'est un indice synthétique, nous vous proposons un exemple simple et typique. Le tableau 5.5 illustre l'évolution des prix rattachés à une sortie au cinéma pour une personne. Comme on peut le voir, il en coûte de plus en plus cher d'aller voir un film : le billet d'entrée, le ticket d'autobus et même le cornet de *popcorn* augmentent, quoique dans des proportions variables. Notre but ici est de construire un indice des prix basé sur l'année 1 (à laquelle on attribuera une base de 100).

Tableau 5.5 - Indice synthétique : une sortie au ciné

Prix de la sortie (en $)	1 billet de cinéma	2 billets d'autobus	1 cornet de popcorn	Dépense totale
Année 1	6	2	2	10
Année 2	8	4	3	15
Taux de variation (en %)	*33,3*	*100,0*	*50,0*	*50,0*
Indices élémentaires (année 1 = 100)				
	1 billet de cinéma	2 billets d'autobus	1 cornet de popcorn	
Année 1	100,0	100,0	100,0	
Année 2	133,3	200,0	150,0	
Taux de variation (en %)	*33,3*	*100,0*	*50,0*	
Pondération des éléments (coefficients budgétaires)				
	1 billet de cinéma	2 billets d'autobus	1 cornet de popcorn	Total
Année 1	0,600	0,200	0,200	1,000
Année 2	0,533	0,267	0,200	1,000

(Données fictives)

Les indices élémentaires sont inscrits dans la deuxième partie du tableau 5.5. À l'année 2, l'indice du prix du billet de cinéma est de 133,3 (soit [8/6] × 100). Les pondérations figurent dans la troisième partie du tableau. À l'année 2, le billet de cinéma représente 0,533 du budget total de la sortie (soit 8 $/15 $ = 0,533).

L'indice synthétique pour l'année 2 est calculé de la manière suivante :

Indice synthétique (pondérations de l'année 1) = (133,3 × 0,6) + (200 x 0,2) + (150 × 0,2) = 80 + 40 + 30 = 150

(En vérifiant ces calculs, vous constaterez peut-être de légers écarts à cause de la manière dont les chiffres ont été arrondis.)

Selon l'indice basé sur les pondérations de l'année 1, le prix moyen d'une sortie au cinéma a augmenté de 50 % (soit [150 - 100]/100 = 50/100 = 0,5 = 50 %). Cela correspond d'ailleurs au taux de variation de la dépense totale qui passe de 10 $ à 15 $.

Étant donné que nos chiffres ne portent que sur deux années, l'indice basé sur les pondérations de l'année 2 présente peu d'intérêt ici. Si on devait le calculer, cet indice serait égal à (133,3 × 0,533) + (200 × 0,267) + (150 × 0,2) = 71,11 + 53,33 + 30 = 154,44). Par contre, les pondérations de l'année 2 pourraient servir à construire l'indice d'une éventuelle année 3.

FORMULES : L'INDICE SYNTHÉTIQUE

Indice synthétique = Moyenne pondérée des indices élémentaires

Moyenne pondérée = (Pondération de l'élément 1 × Élément 1) + (Pondération 2 × Élément 2) + etc.

Si chaque élément fait partie d'un même tout, on calcule une pondération ainsi :

Pondération de l'élément 1 = Valeur de l'élément 1/Somme des éléments

Sinon, les pondérations font l'objet d'un choix « éclairé ».

Indice élémentaire à l'année N = (Valeur à l'année N/Valeur à l'année de base) × Base

Exemple : calcul de l'indice synthétique du prix de la sortie au cinéma (tableau 5.5)

Indices élémentaires à l'année 2 :
Indice de l'élément 1 = (8/6) × 100 = 133,3
Indice de l'élément 2 = (4/2) × 100 = 200
Indice de l'élément 3 = (3 /2) × 100 = 150

Pondérations (basées sur l'année de départ) :
Pondération de l'élément 1 = 6/10 = 0,6
Pondération de l'élément 2 = 2/10 = 0,2
Pondération de l'élément 3 = 2/10 = 0,2

Indice synthétique :
(0,6 × 133,3) + (0,2 × 200) + (0,2 × 150) = 80 + 40 + 30 = 150

EXERCICES 2

1. Grandeur et décadence du 33 tours

À l'aide des chiffres du tableau 5.6, calculez l'indice synthétique des prix des disques et des cassettes, avec comme base l'année 1984 = 100. Utilisez les prix moyens pour construire les deux indices élémentaires, et la valeur des ventes pour calculer les pondérations. Les autres données du tableau sont purement décoratives.

Figure 5.6 - Vente de disques et cassettes au Canada: le déclin du disque en vinyle

	Livraisons nettes (milliers d'unités)				
	Disques 30 cm	Disques compacts	Autres disques	**Total disques**	**Totals cassettes**
1984	32 759	209	14 811	**47 779**	**35 550**
1985	25 550	1 256	12 328	**39 134**	**41 075**
1986	22 477	3 489	10 464	**36 430**	**42 555**
1987	17 882	*9 388*	7 238	**34 508**	***44 406***
1988	12 395	*14 708*	5 786	**32 889**	***51 356***

	Valeur des ventes (milliers de $)		Prix moyen ($)		Vente par habitant (unités)	
	Disques	Cassettes	Disque	Cassette	Disques	Cassettes
1984	133 885	119 803	2,80	3,37	1,86	1,38
1985	118 057	137 922	3,02	3,36	1,51	1,58
1986	136 200	145 242	3,74	3,41	1,39	1,62
1987	176 351	172 626	5,11	3,89	1,30	1,67
1988	206 136	*244 286*	6,27	4,76	1,22	1,91

Source des données brutes : Statistique Canada, Cansim 2904.

Note : les données en italiques sont des estimations; les données en vert constituent des données calculées.

3. UN CAS TYPIQUE : L'INDICE DU DÉVELOPPEMENT HUMAIN

Nous avons affirmé à plusieurs reprises, dans ce manuel, que les chiffres ne mentent jamais. Ce sont plutôt ceux qui en font un usage abusif, devant un auditoire crédule, qui mentent parfois. L'indice du développement humain (IDH) mis au point par la Programme des Nations unies pour le développement (PNUD) est un bon exemple de donnée chiffrée détournée à des fins politiques, du moins dans les premières années de son existence. Le temps est venu d'examiner comment cet indice est construit, étape indispensable avant une interprétation correcte des chiffres. Aux fins d'analyse, nous travaillerons sur les données de 1992, époque où certains pays (que nous ne nommerons pas) arrangeaient quelque peu les données qu'ils fournissaient au PNUD, afin de mieux briller dans ce classement international. Le lecteur trouvera les derniers IDH sur le site du PNUD.

3.1. L'état de la question

Pour commencer, jetons un coup d'œil sur les résultats du palmarès. Dans le tableau 5.7, on retrouve les 10 pays qui obtiennent le meilleur score ainsi qu'une sélection d'autres pays.

Tableau 5.7 - Indice du développement humain et autres indices

Pays	Rang dans l'IDH	Indice du développement humain (IDH)	Indice sexospécifique du développement humain (ISDH)	PIB par habitant (États-Unis = 100)	Rang dans l'IDH
			1992		2012
Canada	1	0,950	0,891	86	11
États-Unis	2	0,937	0,901	100	3
Japon	3	0,937	0,896	86	10
Pays-Bas	4	0,936	0,851	75	4
Finlande	5	0,934	0,918	68	21
Islande	6	0,933		75	13
Norvège	7	0,932	0,911	78	1
France	8	0,930	0,898	82	20
Espagne	9	0,930	0,795	56	23
Suède	10	0,929	0,919	77	7
Royaume-Uni	18	0,915	0,862	72	26
Mexique	53	0,842	0,741	31	81
Niger	174	0,207	0,196	3	186

Source : Rapport mondial sur le développement humain 1995, PNUD. Rapport 2014.

Note : le PIB par habitant est calculé selon la méthode de la parité des pouvoirs d'achat.

Nous avons ajouté au tableau 5.7 quelques autres indices, ainsi que le classement 2012. L'ISDH (indicateur sexospécifique du développement humain) est une variante de l'IDH qui tient compte des inégalités entre les sexes. Un pays où la discrimination sexuelle est plus grande est considéré comme moins développé sur le plan humain et voit son score diminuer. Nous y reviendrons plus loin. Le PIB (produit intérieur brut) par habitant est la méthode de classement des pays la plus courante. Mais le développement humain ne se limite pas à une simple dimension économique.

Étant donné que chaque PIB est comptabilisé en monnaie nationale, les données ont été converties en dollars américains en tenant compte du pouvoir d'achat de chaque devise, et non du taux de change officiel. Le tout a été ramené en indice, pour lequel les États-Unis servent de point de référence. On voit que les classements diffèrent selon la variable choisie : les États-Unis, premiers au chapitre du PIB par habitant, se voient détrôner par le Canada si on se fie à l'IDH, et le Canada perd lui-même sa première place si on tient compte de la discrimination sexuelle.

3.2. La composition de l'IDH

Pour permettre à l'être humain de bien se développer à travers son travail, ses loisirs, ses activités sociales, culturelles et politiques, il faut lui donner un minimum d'atouts. Une personne qui n'a pas accès aux soins médicaux, qui n'a pas la chance de s'instruire et qui a de la difficulté à satisfaire ses besoins matériels les plus essentiels a peu de possibilités de se développer. C'est du moins le point de vue qui a amené le PNUD à identifier trois dimensions du développement humain : la santé, le niveau d'éducation et le niveau de vie. Nous avons déjà parlé de ces trois dimensions au début du chapitre. Voyons maintenant comment traduire ces principes par des données mesurables. Le tableau 5.8 donne les résultats obtenus par quelques pays.

Tableau 5.8 - Les composantes de l'indice du développement humain (IDH) en 1992

Rang dans l'IDH	**Pays**	**Espérance de vie à la naissance** (années) [1]	**Taux d'alphabétisation des adultes** (en %) [2]	**Taux de scolarisation tous niveaux confondus** (en %) [3]	**PIB par habitant (en parité des pouvoirs d'achat)** (en $US) [4]	**PIB corrigé par habitant** (en $US) [5]
1	Canada	77,4	99	100	20 520	5 359
2	États-Unis	76,0	99	95	23 760	5 374
3	Japon	79,5	99	77	20 520	5 359
8	France	76,9	99	86	19 510	5 347
18	Royaume-Uni	76,2	99	77	17 160	5 341
53	Mexique	70,8	89	65	7 300	5 213
174	Niger	46,5	12	14	820	820

		Indicateur d'espérance de vie [6]	**Indicateur de niveau d'éducation** [7]	**Indicateur de PIB** [8]	**Indicateur du développement humain** **[9]**
1	Canada	0,87	0,99	0,98	**0,950**
2	États-Unis	0,85	0,98	0,99	**0,937**
3	Japon	0,91	0,92	0,98	**0,937**
8	France	0,87	0,95	0,98	**0,930**
18	Royaume-Uni	0,85	0,92	0,98	**0,916**
53	Mexique	0,76	0,81	0,96	**0,842**
174	Niger	0,36	0,13	0,13	**0,207**

Source : Rapport mondial sur le développement humain 1995, PNUD.

Indicateur 1 : l'espérance de vie mesure la longévité.

L'espérance de vie à la naissance est un indicateur tout indiqué pour représenter sous forme de chiffre la possibilité de vivre en bonne santé et longtemps. Cet indicateur est facile à obtenir pour tous les pays et reflète de façon éloquente les chances qu'a l'individu de protéger sa santé.

Le chiffre ainsi obtenu est maintenant transformé en indicateur, en le mettant sur une échelle de 0 à 1. Le PNUD fixe la valeur maximum de la longévité humaine à 85 ans (le Japon obtient 79,5) et la valeur minimum à 25 ans (n'importe quelle société, même primitive, peut atteindre ce résultat). Un pays où l'espérance de vie serait de 25 ans se verrait donc attribuer la note de 0; un pays où l'espérance de vie serait de 85 ans se verrait attribuer la note de 1.

Indicateur pour un pays = (Valeur du pays – Valeur minimale)/(Valeur maximale – Valeur minimale)

Indicateur d'espérance de vie pour le Mexique = (70,8 – 25)/(85 – 25) = 45,8/60 = 0,76

Indicateur 2 : le taux d'alphabétisation et le taux de scolarisation indiquent le niveau d'éducation.

Pour l'accès à l'éducation, le PNUD combine deux indicateurs : le taux d'alphabétisation des adultes (qui reflète l'état de la situation) et le taux de scolarisation (qui nous dit où on s'en va). La moyenne entre les deux indicateurs est une moyenne pondérée : le taux d'alphabétisation compte pour les 2/3 et le taux de scolarisation pour 1/3.

Le taux de scolarisation est également ramené sur une échelle de 0 à 1 (ou 0 à 100 %), le pays ayant obtenu le meilleur score (le Canada) servant de base. Ainsi, le score de 65 (%) pour le Mexique signifie que dans ce pays, le taux de scolarisation équivaut à 65 % du taux canadien. En d'autres mots, le taux de scolarisation est *indexé* sur la valeur d'un pays de référence.

Indicateur pour un pays = (2/3 × Taux d'alphabétisation) + (1/3 × Taux de scolarisation)

Indicateur du niveau d'éducation pour le Mexique = (2/3 × 88,6 %) + (1/3 × 65 %) = 81 % = 0,81

Indicateur du niveau d'éducation pour le Mexique = (2 × 88,6 % + 1 × 65 %)/3 = 81 % = 0,81

Indicateur 3 : le PIB par habitant mesure le niveau de vie.

C'est le PIB par habitant qui a été choisi comme indicateur du niveau de vie. Le PIB représente la valeur de la production annuelle sur un territoire donné. Par ailleurs, la méthode de calcul choisie pour évaluer le PIB tient compte de la disparité du coût de la vie d'un pays à l'autre. Même si le PIB par habitant peut prendre n'importe quelle valeur positive, le PNUD estime qu'il varie en pratique entre 100 $ (le minimum vital) et 40 000 $ (niveau au-delà duquel la richesse matérielle n'a plus aucune utilité en matière de développement humain).

Mais ce n'est pas tout. Passé un certain seuil, l'augmentation du revenu d'un individu apporte de moins en moins de satisfaction. Le PNUD fixe ce seuil à 5 120 $ US, soit la moyenne mondiale du PIB par habitant. Jusqu'à ce seuil, chaque dollar est compté à sa pleine valeur, au-delà du seuil, chaque dollar compte de moins en moins. On voit ainsi, dans le tableau 5.8, le Niger conserve son score de 820 $, alors que le score du Canada est ramené de 20 520 $ à 5 359 $. Un pays qui aurait obtenu le maximum (40 000 $) verrait son score ramené à 5 448 $*.

Pour les curieux (les autres s'abstenir), voici un exemple du calcul de conversion : le PIB par habitant du Canada est de 20 520 $; cette valeur est découpée en tranches de 5120 $ (la moyenne mondiale), soit 20 520 = 5120 + 5120 + 5120 + 40; on traite ensuite les tranches de la façon suivante : 5120 + (2 × $5120^{1/2}$) + (3 × $5120^{1/3}$) + (4 × $5120^{1/4}$) + (5 × $40^{1/5}$) = 5359. Ouf, n'est-ce-pas?

Comme les autres indicateurs, le PIB par habitant est ramené sur une échelle de 0 à 1. Un pays qui aurait un PIB par habitant de 100 $ (le minimum) se verrait attribuer la note de 0; un pays qui aurait un PIB par habitant de 5 448 $ (le maximum) se verrait attribuer la note de 1.

Indicateur pour un pays = (Valeur du pays – Valeur minimale)/(Valeur maximale – Valeur minimale)

Indicateur du niveau de vie pour le Canada = (5 359 – 100)/(5 448 – 100) = 0,98.
Indicateur du niveau de vie pour le Mexique = (5 213 – 100)/(5 448 – 100) = 0,96.

L'IDH représente la moyenne des trois indicateurs

Le PNUD a estimé que chacun des indicateurs avait la même importance dans l'évaluation du développement humain. Il leur attribue donc à chacun une pondération de 1/3. Cela revient à faire une moyenne simple dans laquelle chaque élément est traité en parts égales.

IDH d'un pays = (Indicateur 1 + Indicateur 2 + Indicateur 3)/3

IDH du Mexique = (0,76 + 0,81 + 0,96)/3 = 2,53/3 = 0,843
IDH du Mexique = (0,333 × 0,76) + (0,333 × 0,81) + (0,333 × 0,96) = 0,843

3.3. Un regard critique

Il est temps d'examiner de plus près l'attitude triomphaliste du gouvernement canadien de l'époque. Ce dernier prétend en gros que le Canada est le pays où on vit le mieux au monde, et que cette affirmation est d'autant plus crédible qu'elle vient d'une source étrangère et sérieuse (ce qui constitue un double sophisme). Pourtant, le PNUD prend la peine de dire que l'IDH n'est pas une mesure du bien-être ou du degré de bonheur d'une société. Par ailleurs, il ne faut pas oublier que nous mesurons ici des phénomènes humains et non la trajectoire des planètes : le degré de précision n'est pas le même. Ainsi, il est intéressant de comparer le Japon (IDH de 0,937) à Cuba (0,769), à la Chine (0,594), au Bangladesh (0,364) et au Niger (0,207). D'ailleurs, le PNUD divise lui-même les pays en 3 catégories : les pays à développement humain élevé (IDH de 0,8 et plus), moyen (IDH entre 0,5 et 0,799) et faible (IDH inférieur à 0,5). Par contre, la différence entre le Canada et le Japon est si minime qu'elle n'est pas pertinente.

Si on observe de plus près les indicateurs fournis par chaque pays, on constate des phénomènes pour le moins étonnants. Le Japon battrait facilement le Canada si son taux de scolarisation dans l'enseignement supérieur (18 à 24 ans) n'était pas si faible (19,5 % contre 73,4 % pour le Canada). Il serait surprenant que les Japonais aient un tel retard sur le Canada dans ce domaine. Cela ne signifierait-il pas plutôt que les Canadiens mettent plus de temps à apprendre que les studieux confrères Japonais, ou, tout simplement, que les chiffres du Canada sont carrément truqués? D'ailleurs, les taux de scolarisation dans l'enseignement supérieur des États-Unis (72,9 %), de la France (51,2 %) et du Royaume-Uni (31,9 %) semblent montrer que les systèmes nord-américains, européens et japonais ne sont tout simplement pas comparables.

Enfin, le niveau du PIB par habitant est très minimisé dans la formule de l'IDH : comparez, dans le tableau 5.8, le PIB par habitant et l'indicateur de PIB qui en est déduit. De plus, si on utilise le taux de change (et non le pouvoir d'achat) pour faire les comparaisons entre pays, on constate que le PIB par habitant du Japon dépasse de moitié celui du Canada. Le coût de la vie étant plus élevé au Japon (à peu près 50 %, également), les deux pays se voient attribuer le même PIB par habitant. Il n'en demeure pas moins qu'un Japonais qui voyage au Canada doit trouver la vie bon marché, alors que le Canadien en visite au Japon, tout champion de l'IDH qu'il est, ne doit pas toujours se sentir très riche.

3.4. L'ISDH, indice *sexospécifique* du développement humain

L'ISDH est une sorte d'IDH amélioré. Derrière les moyennes sur lesquelles est construit l'IDH se cachent parfois de profondes inégalités : entre races, entre sexes, entre classes sociales, etc. Or, 50 personnes très riches, en bonne santé et diplômées universitaires accompagnées de 50 personnes très pauvres, malades et sans instruction ne font pas 100 êtres humains développés. Dans le calcul de l'ISDH, on tient justement compte d'une des formes d'inégalités la mieux partagée à travers le monde : celle qui existe entre les sexes. Plus les inégalités entre sexes seront grandes dans un pays, plus son ISDH sera faible par rapport à son IDH (à moins que cette inégalité se manifeste au détriment des hommes). À l'opposé, un pays où il n'y aurait pas d'inégalités obtiendrait le même ISDH que son IDH.

Les indicateurs retenus pour l'ISDH sont sensiblement les mêmes que pour l'IDH : espérance de vie, taux d'alphabétisation et de scolarisation, PIB par habitant. Pour mesurer l'inégalité, on tient compte, pour chaque indicateur, de l'écart entre les hommes et les femmes (voir tableau 5.9).

Tableau 5.9 - Les composantes de l'indice sexospécifique du développement humain (ISDH)

Rang	Pays	ISDH	IDH	Espérance de vie (années)		Taux d'alphabétisation (%)		Taux de scolarisation (%)		Part des revenus du travail (%)	
				Femmes	Hommes	Femmes	Hommes	Femmes	Hommes	Femmes	Hommes
1	Suède	0,919	0,929	81	75	99,0	99,0	79,3	76,7	41,6	58,4
9	Canada	0,891	0,950	81	74	99,0	99,0	100,0	100,0	29,3	70,7
47	Cuba	0,726	0,769	77	74	94,2	95,5	67,7	67,9	27,2	72,8
127	Niger	0,196	0,207	48	45	5,8	19,3	10,2	18,0	39,6	60,4

Source : Rapport mondial sur le développement humain 1995, PNUD.

La longévité.

En principe, l'espérance de vie des femmes dépasse celle des hommes de 5 ans. En Suède, l'écart est de 5,7 ans, au Canada de 6,5 et au Niger de 3,2. La Suède et le Canada améliorent leur score.

Le niveau d'éducation.

Là encore, l'écart est en faveur des femmes pour la Suède alors qu'il est en défaveur des femmes au Niger. La Suède gagne des points et le Niger en perd. Le Canada ne bouge pas.

Le niveau de vie.

Étant donné qu'il n'existe pas de PIB par homme ou par femme (alors qu'il existe une espérance de vie pour chaque sexe), on a choisi d'utiliser la répartition des revenus du travail entre les sexes pour mesurer l'inégalité du niveau de vie. Les femmes reçoivent 41,6 % des revenus du travail en Suède, 39,6 % au Niger et seulement 29,3 % au Canada. On voit que la Suède est proche de l'égalité à ce chapitre et que le Canada en est loin. Deux remarques s'imposent ici. D'une part, la plus faible proportion obtenue par les femmes peut être due à des salaires plus bas, mais aussi à une plus faible participation au marché du travail. D'autre part, pour évaluer l'inégalité, il faut comparer la part du revenu reçu par les femmes à la proportion de ces mêmes femmes dans la population (qui n’est pas nécessairement de 50 %). Si au Canada, les femmes représentent près de 51 % de la population, cette proportion descend jusqu’à 47 % dans certains pays (Corée du Sud, Inde).

Dans l’ensemble, la Suède et Cuba font bonne figure au chapitre de l’ISDH, et pourtant le niveau de vie est 5 fois plus élevé en Suède qu’à Cuba. Cela prouve que l’égalité est avant tout le résultat d’un choix politique et non une affaire de gros sous. Les autres pays qui se classent mieux dans l’ISDH que dans l’IDH (les pays à inégalités faibles) sont, entre autres, le Danemark, la Finlande, la Norvège, la Pologne, la Hongrie, la Tchéquie, la Malaisie, Sri Lanka et la Jamaïque. À l’opposé, dans les pays pour lesquels le classement selon l’ISDH est nettement inférieur à celui de l’IDH (les pays à inégalités fortes), on retrouve l’Argentine, le Chili, l’Arabie Saoudite, l’Égypte, l’Algérie, les Pays-Bas, l’Espagne et le… Canada.

EXERCICES 3

1. Sommes-nous les meilleurs?

Cet exercice se rapporte au tableau 5.8 (les composantes de l’IDH).

a) Comment se fait-il que le Japon obtienne le même IDH que les États-Unis malgré le fait que les Japonais vivent plus vieux que les Américains?

b) À partir des données brutes (partie gauche du tableau 5.8), calculez les indicateurs d’espérance de vie, de niveau d’éducation et de PIB et l’IDH pour la France et le Royaume-Uni. Comparez vos résultats avec les chiffres figurant dans la partie droite du tableau.

2. Derrière l’indice, des principes

a) Quelles sont les trois dimensions du développement humain retenues dans la construction de l’IDH? Quels sont les indicateurs choisis pour évaluer ces dimensions?

b) L’IDH est-il un bon indicateur du développement humain? Donnez le pour et le contre.

4. L'INDICE DES PRIX À LA CONSOMMATION

Nous avons déjà fait connaissance avec l'indice des prix à la consommation au chapitre précédent. On y voyait (à la figure 4.4) que les prix avaient baissé au Canada entre les deux guerres et qu'ils n'avaient cessé d'augmenter depuis. Les changements de prix affectent la valeur (ou pouvoir d'achat) de la monnaie : nous verrons comment en tenir compte grâce à l'indice des prix à la consommation (IPC). L'étude de l'indice des prix à la consommation nous permettra en même temps de revoir une bonne partie des notions abordées dans ce chapitre. (Les données historiques de l'IPC figurent en annexe de ce manuel.)

4.1. L'IPC est un indice synthétique

Il existe de nombreux prix et tous n'ont pas le même impact sur le niveau général d'inflation. Une hausse du prix du bœuf a plus de conséquences qu'une hausse du prix du poisson, car le bœuf pèse 3 fois plus lourd que le poisson dans le budget d'un consommateur canadien moyen. Statistique Canada, qui est chargée de la construction de l'IPC, a élaboré un « panier type » de consommation contenant environ 600 produits représentatifs des dépenses d'un ménage moyen. Chaque mois, cet organisme relève en moyenne 60 000 prix. Certains produits voient leur prix relevé plusieurs fois et en des lieux très différents alors que dans d'autres cas les prix, plus stables, sont relevés moins souvent. Cela permet de calculer, pour chaque produit, un indice élémentaire.

L'IPC est un *indice synthétique*, c'est-à-dire qu'il combine tous ces *indices élémentaires* pour donner, en un seul chiffre, un résumé du niveau des prix. Chacun des produits qui constituent le panier type se voit attribuer une *pondération* conforme au comportement d'un ménage typique. L'indice (élémentaire) du prix des disques est, par exemple, de 121 en 1994 (base 1986 = 100) et la pondération des disques (et cassettes) est de 0,35 %. Cela signifie que le prix des disques a augmenté de 21 % entre 1986 et 1994 et que l'achat de disques compte pour 0,35 % du budget d'un ménage moyen. Comme on peut s'y attendre, les paniers types évoluent avec le temps, et celui de 2011 a remplacé la rubrique *disques et cassettes* par la rubrique *équipement informatique numérique et dispositifs* (avec une pondération de 0,66 %).

Le tableau 5.10 indique les pondérations associées aux principaux groupes de produits. On y remarque, par exemple, que les Canadiens dépensent deux fois plus pour s'habiller que pour assouvir certains vices (tabac et alcool). Plutôt que d'indiquer la pondération de chaque élément du panier (il aurait fallu plusieurs pages), nous les avons groupés par catégories. La pondération des *dispositifs numériques multifonctionnels* (0,11 %, non citée dans le tableau), par exemple, fait partie des 11,26 % alloués au groupe *loisirs*. La somme des pondérations donne évidemment 100 % (colonne 1).

Tableau 5.10 - Pondérations de l'indice des prix au Canada

	Pondérations originales du panier de 2011 (en %) [1]	**IPC juin 2014 (Base 2002 = 100)** [2]	**Moyenne pondérée** [3]	**Pondérations du panier de 1992** (en %) [4]
Aliments	16,35	136,4	22,30	18,04
Logement	25,86	132,2	34,19	27,58
Ameublement et équipement du ménage	12,57	116,4	14,63	10,35
Vêtements et chaussures	6,20	92,7	5,75	6,82
Transports	20,05	133,1	26,69	17,22
Soins de santé et soins personnels	4,95	119,0	5,89	4,35
Loisirs, formation et lecture	11,26	108,2	12,18	10,17
Boissons alcoolisées et produits du tabac	2,76	146,7	4,05	5,47
Ensemble	**100,00**	**125,9**	**125,68**	**100,00**

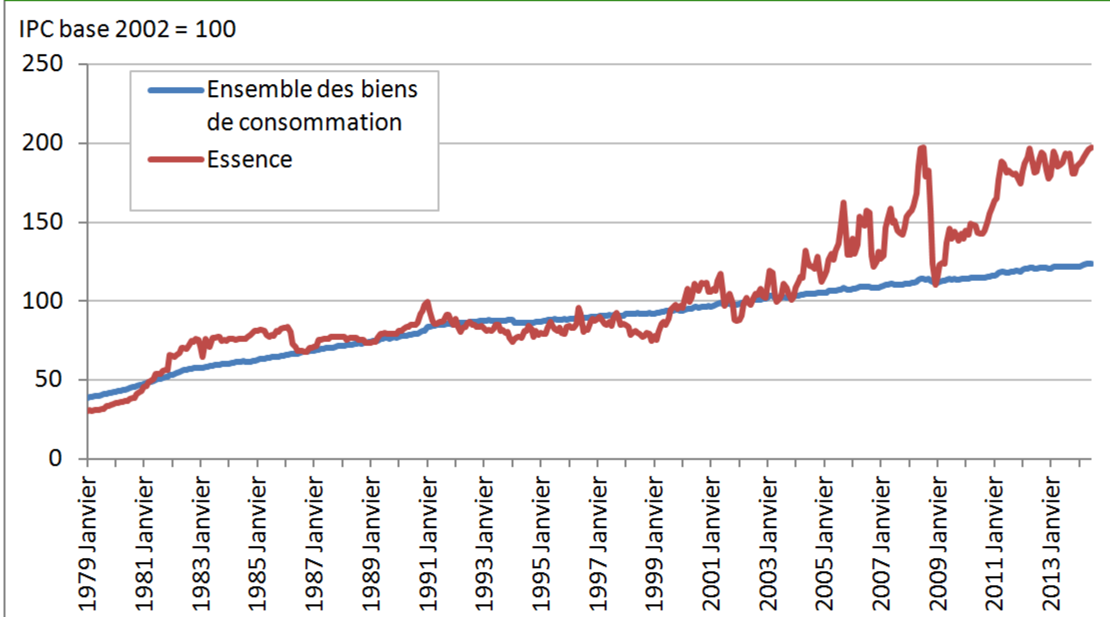
Source : Document de référence de l'indice des prix à la consommation, Statistique Canada; Cansim 326-0020.

Le tableau 5.10 contient aussi l'indice des prix pour juin 2014 (colonne 2). Celui des aliments, par exemple, est de 136,4. Cela signifie que les prix des aliments ont augmenté de 36,4 % depuis l'année de base (qui est 2002). Dans l'ensemble (dernière ligne du tableau), les prix ont augmenté de 25,9 % entre 2002 et juin 2014. L'IPC, qui était de 125,9 en juin 2014, représente la moyenne (pondérée) de tous les indices de la colonne 2.

Aux fins du calcul de l'indice synthétique (l'IPC), les indices de chaque groupe de produits du tableau 5.10 peuvent être considérés comme des indices élémentaires. Pour calculer l'IPC, nous avons donc fait la moyenne pondérée des indices élémentaires (en utilisant les pondérations de la colonne 1). Ainsi, les aliments contribuent à 16,35 % × 136,4 = 22,3 points de l'IPC de juin 2014 (colonne 3). La somme de la colonne 3 donne l'IPC de juin 2014, que l'on peut comparer au chiffre officiel (en bas de la colonne 2). Le léger écart entre les deux chiffres (125,9 pour le chiffre officiel et 125,68 pour notre chiffre calculé) vient du fait que nous travaillons sur des données arrondies (à une seule décimale de pourcentage) alors que Statistique Canada dispose de données plus précises.

Puisque l'IPC représente un indice synthétique, il n'est pas dit que tous les indices élémentaires qui le composent le suivent de près. On remarque sur la figure 5.1, d'une part, que le prix de l'essence est beaucoup plus instable que la moyenne des prix à la consommation et, d'autre part, que ce prix a parfois eu tendance à augmenter plus rapidement que cette moyenne, notamment à partir de l'année 2004. Curieusement, cela n'a pas empêché les Canadiens d'acheter de plus en plus de camionnettes et de camions légers en guise de véhicules de promenade.

FIGURE 5.1 - Évolution de l'indice du prix de l'essence au Québec



Source : Statistique Canada, Cansim 326-0020.

À titre de complément au tableau 5.10, précisons que l'essence représentait 4,85 % de la pondération de l'IPC en 2011, contre 5,57 % pour les autres frais associés à l'utilisation de véhicules de tourisme et 7,64 % pour l'achat ou la location de ces mêmes véhicules. En tout, le coût du transport privé comptait donc pour 18,06 % du budget d'un consommateur typique, contre 1,99 % pour le transport public, et 16,35 % pour les aliments (ce dernier chiffre figure dans le tableau 5.10). Comme nous venons de le constater, l'IPC nous livre volontiers toute une richesse d'informations.

4.2. La mise à jour des pondérations

Avis : On recherche le « Canadien moyen ».

Les pondérations de l'IPC reposent sur le panier type d'un ménage canadien moyen. Mais le Canadien moyen existe-t-il? Est-ce un individu à moitié français et à moitié anglais? Est-il très catholique, assez protestant et un peu musulman? Certaines sociétés insistent sur ce qui unit leurs membres, d'autres sur ce qui les distingue.

Pour contribuer au débat, nous vous proposons de comparer les pondérations des ménages québécois, pauvres et riches (tableau 5.11 ci-après), à celles des ménages canadiens moyens (colonne 4 du tableau 5.10, vu précédemment). Nous utiliserons pour cela les pondérations de 1992, que nous comparerons ensuite à celles de 2011. Les différences sont évidentes, non seulement entre les Québécois et les Canadiens en général, mais aussi entre les riches et les pauvres. Les ménages pauvres affectent une plus grande partie de leur budget à la satisfaction des besoins essentiels (aliments et logement). Ces écarts de pondération ne signifient pas pour autant qu'un groupe de la population subit plus d'inflation que l'autre.

Tableau 5.11 - Les pondérations de l'IPC selon le niveau de vie au Québec en 1992

(Pondérations basées sur les prix de 1992)	Ménages pauvres (1er quartile du revenu)	Ménages riches (4e quartile du revenu)
Aliments	23,0	18,1
Logement	33,5	20,4
Dépenses et équipement du ménage	9,6	10,3
Habillement	5,4	8,6
Transport	10,9	18,5
Santé et soins personnels	5,5	5,5
Loisirs et formation	5,2	10,4
Tabac et alcool	4,9	4,0
Divers	2,0	4,2
Total	**100**	**100**

Source : Le Québec statistique, 1995.

Ces différences entre ménages riches et pauvres tendent par contre à montrer que les pondérations sont susceptibles de changer avec le temps, puisqu'elles dépendent largement du niveau de revenu et que celui-ci évolue sensiblement d'une génération à l'autre. Le fait que la part consacrée à l'alimentation et au logement diminue d'une génération à l'autre (comparer les colonnes 1 et 4 du tableau 5.10) témoigne d'une augmentation moyenne du niveau de vie à long terme. C'est la raison pour laquelle Statistique Canada effectue régulièrement des enquêtes sur la consommation afin de mettre à jour les pondérations.

Il n'y a d'ailleurs pas que les habitudes de consommation qui font changer les pondérations de l'IPC. La hausse du prix de l'essence fera augmenter la pondération correspondante : on ne roulera pas plus, mais l'essence pèsera plus lourd dans le budget. Raison de plus pour réviser périodiquement les pondérations.

4.3. L'IPC et le taux d'inflation

Le taux annuel d'inflation n'est autre que le *taux annuel de variation* des prix. Pour l'obtenir, il suffit de comparer l'IPC de l'année qui nous intéresse à celui de l'année précédente. Ainsi, le taux d'inflation en 1991 sera égal à l'écart entre les IPC de 1991 et de 1990 divisé par l'IPC de 1990 (voir tableau 5.12).

Figure 5.12 - Indice des prix et taux d'inflation

	IPC (base 1986 = 100)	IPC (base 1991 = 100)	IPC (base 2002 = 100)	Taux de croissance annuelle (%)	*Indice du prix des cassettes et disques*
	[1]	[2]	[3]	[4]	*[5]*
1984	92,4	73,2	60,6		..
1985	96,0	76,1	63	3,9	*98,3*
1986	**100,0**	79,2	65,6	4,2	*100,0*
1987	104,4	82,7	68,5	4,4	*106,1*
1988	108,6	86,1	71,2	4,0	*114,3*
1989	114,0	90,3	74,8	5,0	*119,6*
1990	119,5	94,7	78,4	4,8	*120,5*
1991	126,2	**100,0**	82,8	5,6	*127,2*
1992	128,1	101,5	84	1,5	*121,1*
1993	130,4	103,3	85,6	1,8	*121,8*
1994	130,7	103,6	85,7	0,2	*121,0*
1995	133,6	105,9	87,6	2,2	
2002	152,5	120,8	**100,0**		
2013	187,3	148,4	122,8		

Source : Statistique Canada, Cansim.

Taux annuel d'inflation = (IPC de l'année considérée – IPC de l'année précédente)/IPC de l'année précédente

Exemple : Taux d'inflation en 1991 = (126,2 – 119,5)/119,5 = 0,056 = 5,6 %

On pourrait de la même manière calculer un taux d'inflation sur une période différente. Le taux d'inflation entre 1990 et 1994 serait de (130,7 – 119,5)/119,5 = 9,4 %.

Nous nous sommes servis des données de la colonne 1 pour calculer les taux d'inflation qui précèdent, mais nous aurions tout aussi bien pu utiliser les données des colonnes 2 ou 3. L'information contenue dans ces trois colonnes est la même, seul diffère le point de référence (l'année de base).

Lorsque les prix baissent, les calculs restent les mêmes. Cependant, l'expression « inflation négative » peut sembler un peu étrange. On pourra alors parler de déflation.

4.4. Le changement de base

Il est facile de modifier l'année de base de l'IPC. C'est d'ailleurs ce que fait périodiquement Statistique Canada, généralement au moment où elle met à jour les pondérations. Dans le tableau 5.12 ci-dessus, nous avons créé un nouvel IPC basé sur l'année 1991 (colonne 2). On observe que les trois colonnes d'IPC sont proportionnelles. On peut facilement vérifier ce fait en calculant le taux annuel d'inflation à partir des deux séries d'IPC : on obtient la même réponse.

IPC (base 1991) de l'année N = IPC (base 1986) de l'année N/IPC (base 1986) de l'année 1991 × Base

Exemple : IPC (base 1991) de 1994 = 130,7 /126,2 × 100 = 103,6

4.5. La croissance en valeur et la croissance réelle

La croissance réelle mesure la croissance du pouvoir d'achat d'une somme d'argent.

Entre 1985 et 2013, le salaire minimum passe de 4 $ à 10,15 $ au Québec. Le taux de croissance semble appréciable : (10,15 – 4)/4 = 1,65/4,35 = 154 %. Dans le même temps, les prix augmentent de : (122,8 – 63)/63 = 95 %. On serait tenté de dire que le pouvoir d'achat a progressé de 59 % (soit 154 - 95), et on aurait tort! Mais comment calcule-t-on au juste le taux de croissance réelle du salaire?

Dans le tableau 5.13, nous avons converti le salaire officiel ou *nominal* (colonne 1) en dollars de 2002 (colonne 5) à l'aide de l'indice des prix à la consommation (colonne 3).

Figure 5.13 - La croissance réelle du salaire miminum au Québec

	Salaire minimum ($ courants)	Taux de croissance en valeur (%)	IPC (base 2002 = 100)	Taux d'inflation (%)	Salaire minimum ($ de 2002)	Taux de croissance réelle (%)
	[1]	[2]	[3]	[4]	[5]	[6]
1985	4,00		63,0		6,35	
1986	4,35	8,7	65,6	4,1	6,63	4,4
1987	4,55	4,6	68,5	4,4	6,64	0,2
1988	4,75	4,4	71,2	3,9	6,67	0,4
1989	5,00	5,3	74,8	5,1	6,68	0,2
1990	5,30	6,0	78,4	4,8	6,76	1,1
1991	5,55	4,7	82,8	5,6	6,70	-0,8
1992	5,70	2,7	84,0	1,4	6,79	1,2
1993	5,85	2,6	85,6	1,9	6,83	0,7
1994	6,00	2,6	85,7	0,1	7,00	2,4
1995	**6,45**	7,5	87,6	2,2	7,36	5,2
2002	**7,20**	**11,6**	100,0	**14,2**	7,20	**-2,2**
2013	10,15	41,0	122,8	22,8	8,27	14,8

Sources : Commision des normes du travail du Québec (salaire minimum); Statistique Canada, Cansim (IPC).

Note : Le salaire minimum québécois passe de 6,45 $ en 1995 à 7,20 $ en 2002, soit une hausse apparente de 11,6 % (en valeur). Entre-temps, les prix augmentent 14,2 %. Les prix ayant augmenté plus vite que le salaire, le pouvoir d'achat du salarié diminue de 2,2 %. Ce dernier chiffre peut être obtenu en convertissant le salaire officiel (exprimés en dollars courants) en salaire réel (exprimé en dollars d'une année de base) grâce à l'indice des prix à la consommation (IPC).

Salaire d'une année particulière en $ de 2002 = Salaire nominal/IPC (base 2002) × Base

Exemple : Salaire de 2013 en $ de 2002 = 10,15 $/122,8 × 100 = 8,27 $.

Un calcul similaire pourrait être fait pour n'importe quelle année de base.

Le salaire réel (en pouvoir d'achat) passe donc de 6,35 $ en 1985 à 8,27 $ en 2013 (colonne 5). Le taux de croissance réelle du salaire est de (8,27 – 6,35)/6,35 = 0,302 = 30,2 % entre 1985 et 2013. Ça peut paraître beaucoup, mais pendant ces 28 ans, le PIB réel par habitant a augmenté de 75 % au Québec (source : Cansim 384-0038). Paradoxalement, le sort des travailleurs les plus modestes s'améliore moins vite que celui de l'ensemble de la population.

Il existe une autre méthode pour calculer le taux de croissance réelle. Pour cela, il faut utiliser une forme particulière d'indice : ce que nous avons appelé, au chapitre 4, l'indice de variation. Reprenons les données du problème pour la période 1985-2013. Le salaire officiel (ou nominal) augmente de 154 % (il est multiplié par 2,54) et les prix augmentent de 95 % (ils sont multipliés par 1,95). D'habitude, lorsqu'on divise le montant d'argent qu'on a en poche par le prix d'un article, on sait alors combien d'articles on peut acheter. On agira ici de la même manière : en divisant l'indice de variation du salaire officiel par l'indice de variation des prix, on obtient l'indice de variation du pouvoir d'achat : 2,54/1,95 = 1,302. Pour chaque dollar gagné en 1986, on reçoit maintenant 1,302 dollars, soit une hausse de 30,2 %.

Indice de variation réelle = Indice de variation nominale / Indice de variation des prix
Indice de variation = 1 + Taux de variation
Taux de variation = Indice – 1

2,54/1,95 = 1,302, d'où le taux de variation = 1,032 – 1 = 0,302 = 30,2 %.

Cette méthode possède un autre avantage : on n'a même pas besoin de connaître le salaire ou l'IPC pour déterminer le taux de croissance réelle, les taux de variation suffisent.

Cette méthode nous permet de calculer, par exemple, que :

- une hausse de salaire (nominal) de 32 % combinée à 10 % d'inflation équivaut à une hausse réelle de 20 % (car 1,32/1,10 = 1,20);
- un taux d'intérêt bancaire de 15,5 % combiné à 10 % d'inflation rapporte en réalité 5 % (car 1,155/1,10 = 1,05);
- une hausse de 7,2 % de l'aide aux pays en développement combinée à 10 % d'inflation équivaut en pouvoir d'achat à une baisse de 2,5 % (car 1,072/1,10 = 0,975, d'où le taux 0,975 – 1 = –0,025 = –2,5 %).

Cette méthode de calcul du taux de croissance réelle est un peu plus abstraite que la précédente, mais elle n'a rien de sorcier et elle peut facilement s'appliquer à de nombreuses situations.

Pour conclure sur l'IPC, soulignons qu'il existe d'autres indices de prix (des importations, de la production, des biens d'équipement, des matières premières). Le ménage qui veut évaluer son augmentation de salaire ou son taux d'hypothèque se basera sur l'indice des prix à la consommation. L'éleveur de bétail s'intéressera plutôt à l'indice des prix du grain. Ces indices de prix sont relativement faciles à obtenir. Ce qui compte, c'est avant tout d'utiliser l'indice approprié à la situation étudiée.

EXERCICES 4

1. Taux de variation

a) Quel est le taux d'inflation en 1986, 1987 et 1994?

b) Quel est le taux de variation du prix des disques et cassettes en 1986?

Note : utilisez les indices de prix du tableau 5.12.

2. La croissance réelle

a) Vérifiez les taux de croissance réelle du salaire minimum en 1986, 1987 et 1994 qui apparaissent au tableau 5.13.

b) Calculez le taux de croissance réelle du salaire minimum entre 1988 et 1993.

c) La Banque Nationale du Canada offre un taux d'intérêt de 7 % sur ses comptes d'épargne en 1986. Quel est le taux d'intérêt réel?

d) Recherche. Consultez le site de la Commission des normes du travail du Québec pour obtenir les dernières données sur le salaire minimum.

EXERCICES SUPPLÉMENTAIRES

1. Le paradis des retraités

Un Canadien qui songeait à prendre sa retraite en 1995 pourrait constater, à la lecture du tableau 5.14, que les maisons ne sont pas chères en Irlande, mais que le coût de la vie y est plus élevé qu'au Mexique. Pour simplifier les comparaisons, nous vous demandons de convertir tous les chiffres du tableau en indices, en utilisant comme base la ville canadienne qui figure dans le tableau.

Tableau 5.14 - Le coût de la vie dans les villes de rêve pour retraités en 1995

		Achat d'une maison de 3 chambres	Dépenses mensuelles par personne
		(en $US)	
Sarasota	États-Unis	160 000	1500
Okanagan	Canada	115 000	900
San José	Costa Rica	120 000	600
Comtés de l'Ouest	Irlande	90 000	700
Guadalajara	Mexique	100 000	500

Source: Fortune, 24 juillet 1995.

2. Les disparités en Allemagne après la réunification

Dans le tableau 5.15 ci-dessous, le PIB par tête sert d'indicateur du niveau de vie dans les différents *Länder* (provinces, ou régions) qui forment la République fédérale d'Allemagne.

Tableau 5.15 - L'Allemagne unie mais pas homogène

	Population	**Superficie**	**PIB par tête**
	(en millions)	(en milliers de km²)	(en milliers de marks)
Rhénanie-du-Nord-Westphalie	17,8	34,1	43,4
Bavière	11,9	70,5	49,1
Bade-Wurtemberg	10,2	35,8	48,3
Basse-Saxe	7,7	47,6	40,1
Hesse	6,0	21,2	57,1
Saxe*	4,6	18,4	23,7
Rhénanie-Palatinat	3,9	19,8	38,7
Berlin	3,5	0,9	42,6
Saxe-Anhalt*	2,8	20,4	23,4
Schleswig-Holstein	2,7	15,7	40,7
Brandebourg*	2,5	29,5	25,1
Thuringe*	2,5	16,2	23,3
Mecklembourg-Poméranie-Occidentale*	1,8	23,2	22,5
Hambourg	1,7	0,8	78,9
Sarre	1,1	2,6	39,5
Brême	0,7	0,4	58,1

Source : Der Spiegel in Courrier international, 23 mai 1996.

Note : 1 DM (Deutsche Mark) = 0,71 $CAN au 11 juillet 1995.

a) On peut déduire des données du tableau que le PIB total de l'Allemagne est 3 465 milliards de marks et que le pays compte 81,4 millions d'habitants. Calculez le PIB par tête pour l'ensemble de l'Allemagne.

b) En prenant comme base le PIB par tête dans l'ensemble de l'Allemagne (tel que calculé dans la sous-question précédente), calculez l'indice de PIB pour chacun des *Länder* allemands.

c) Qu'avez-vous à dire sur les régions marquées d'un astérisque dans le tableau?

Laboratoire

d) Vérifiez, à l'aide des données du tableau, que le PIB total de l'Allemagne est bien de 3 465 milliards de marks (conseil : utilisez un chiffrier électronique).

e) Tracez un diagramme circulaire représentant la répartition de la population par *Land*.

f) Dessinez une carte d'Allemagne en représentant l'indice de PIB par tête par des couleurs plus ou moins foncées.

g) Recherche. Obtenez des données similaires sur un autre pays et dessinez une carte semblable à celle de la sous-question précédente.

3. Les familles qui ont encore des enfants

Le tableau 5.16 ci-dessous montre comment le revenu des différents types de familles canadiennes varie avec les années. Aux fins de l'étude, on n'a considéré que les familles qui ont un ou plusieurs enfants (il n'existe pas de famille monoparentale sans enfants!), et on n'a pas tenu compte des enfants de 18 ans et plus, ni des familles dont le chef avait 65 ans ou plus.

Tableau 5.16 - Évolution du revenu moyen des familles canadiennes

	Deux parents et enfants	Père et enfants	Mère et enfants		Deux parents et enfants	Père et enfants	Mère et enfants
	(en $ de 1991)				(en $ de 2011)		
1981	56 005	42 923	23 686	2001	79 300	49 200	37 300
1982	54 666	38 496	21 682	2002	80 100	49 900	34 700
1983	54 424	35 641	21 435	2003	80 700	53 100	34 900
1984	54 642	38 153	22 491	2004	83 600	50 500	35 700
1985	56 244	36 405	22 174	2005	81 900	57 700	40 300
1986	57 469	37 707	22 271	2006	84 000	59 000	40 900
1987	58 824	47 024	22 907	2007	88 300	56 200	42 500
1988	59 724	40 248	22 938	2008	89 000	56 900	43 500
1989	61 644	47 062	25 020	2009	89 000	58 500	45 800
1990	60 420	38 790	23 196	2010	91 400	58 600	46 100
1991	59 014	36 669	22 186	2011	93 700	55 100	43 000

Source : Statistique Canada, Enquête sur les finances des consommateurs, SC 13-207; Cansim 202-0603.

Note : Seules les familles ayant des enfants de moins de 18 ans et dont le chef est âgé de moins de 65 ans sont comptabilisées.

a) Calculez l'indice de revenu pour chacune des trois catégories en prenant 1981 pour année de base (1981 = 100).

b) Calculez, pour les familles biparentales, le taux de variation annuel en 1982, 1990 et 1991 en utilisant d'abord les chiffres bruts et en utilisant ensuite les indices.

c) Calculez le taux de croissance total entre 1981 et 1991.

d) Calculez le taux de croissance annuel moyen entre 1981 et 1991.

e) Que pensez-vous de l'affirmation suivante : « En réalité, le revenu moyen des familles a baissé au cours des années 1980, compte tenu de l'augmentation du coût de la vie. »

f) Que pensez-vous de l'affirmation suivante : « Au cours de la décennie 2001-2011, ce sont les familles monoparentales, et notamment celles dont le chef de famille est un homme, qui ont connu la plus faible croissance du revenu familial moyen. »

4. L'université : pas gratuit!

L'indice des frais de scolarité universitaires aux États-Unis (base 1980 = 100) était en 1994 de 360 pour les universités d'État et de 340 pour les universités privées (source : Fortune, 25 juillet 1995). L'indice des prix à la consommation de 1994 était quant à lui d'environ 180 (toujours avec la même base de comparaison).

a) Calculez le taux de croissance des frais entre 1980 et 1994 pour les deux types d'université.

b) Calculez le taux de croissance annuel moyen des frais entre 1980 et 1994 pour les deux types d'université.

c) Calculez le taux de croissance *réel* (en tenant compte de l'inflation) des frais entre 1980 et 1994 pour les deux types d'université.

À titre d'information supplémentaire, les frais de scolarité étaient en 1995-96 de 20 865 $US pour Harvard (université privée) et de 4 354 $US pour Berkeley (université publique). À Berkeley cependant, les étudiants en provenance de l'extérieur de la Californie devaient débourser 12 053 $US. En 2014, les frais de scolarité pour un étudiant de première année étaient de 41 616 $ à Harvard et 16 294 $ à Berkeley.

En juin 2014, l'indice des frais de scolarité aux États-Unis (base 1982-1984 = 100) était de 659,5, contre 238,3 pour l'indice d'ensemble des prix à la consommation. Cela signifie que les prix, en général, avaient augmenté de 138,3 % au cours de ces trois décennies, alors que les frais de scolarité, en particulier, avaient augmenté de 559,5 % (source : *Bureau of Labor Statistics*)! Si les États-Unis ne bénéficiaient pas d'un puissant exode des cerveaux en leur faveur, on pourrait se demander si une telle situation n'est pas suicidaire pour le pays et sa prospérité future.

5. Commentaire de tableau

Commentez le tableau 5.17.

Tableau 5.17 - Indice de production alimentaire par habitant

(1979-81 = 100)	1992
Libye	84
Tunisie	121
Syrie	93
Jordanie	134
Algérie	127
Liban	190
Iraq	84
Égypte	119
Maroc	110
Yémen	74
Soudan	89

Source : Rapport mondial sur le développement humain 1995, PNUD. Données de 1992.

Notes : le tableau comprend les pays arabes, à l'exception des pays riverains du Golfe Persique. Les pays sont classés selon l'indice de développement humain (IDH) de la PNUD.

6. Un IPC pour chacun

L'IPC calculé dans la colonne 3 du tableau 5.10 est basé sur les pondérations du « Canadien moyen ». En utilisant les pondérations du tableau 5.11, calculez l'IPC du Québécois pauvre moyen et du Québécois riche moyen pour la même période. Comparez les résultats.

7. Les pondérations : Québécois et Canadiens, riches et pauvres

Dites si les affirmations suivantes sont vraies ou fausses et expliquez pourquoi. Si nécessaire, faites des hypothèses sur le niveau de vie ou les types de produits consommés. (Utilisez le tableau 5.10 et le tableau 5.11)

a) Les ménages québécois pauvres consomment plus de tabac et d'alcool que les ménages québécois riches.

b) Les Canadiens des provinces anglaises consomment plus de tabac et d'alcool que les ménages québécois.

c) Les ménages pauvres québécois sont mieux logés que les ménages canadiens en général.

8. Votre propre indice

Le cabinet genevois Corporate Resources Group a analysé la qualité de la vie dans 118 métropoles à partir de 42 facteurs tels que la délinquance, la pollution, les conditions économiques et sociales. Genève arrive en tête devant Vancouver, Vienne, Toronto, Luxembourg, Ottawa, Zurich et Montréal. Les pires métropoles sont (en commençant par la dernière) : Alger, Kiev, Lagos, Moscou, Saint-Pétersbourg, Canton, Shanghai et Pékin.

a) Que pensez-vous des résultats? Est-ce un signe que le choix des indicateurs est subjectif? Y a-t-il une seule définition de la qualité de la vie?

b) Construisez votre propre indice de la qualité de vie : quels indicateurs retiendriez-vous? Quelle importance respective leur accorderiez-vous?

c) Recherche. Testez votre indice en obtenant des données sur les différentes villes, quartiers, écoles ou autres lieux de vie.

DOSSIER 5 LA BOMBE DÉMOGRAPHIQUE

Le Nigéria dépasse le Japon

Estimée à seulement 20 millions d'habitants en 1931 (contre 65 millions pour le Japon), la population du Nigéria a, depuis, connu une ascension vertigineuse : 56 millions en 1970, 103 millions en 1993. Dans son *Rapport sur le développement dans le monde 1995*, la Banque mondiale estime même que cette population dépassera 238 millions en 2025 et continuera d'augmenter par la suite, jusqu'à ce qu'elle atteigne un plafond de 382 millions d'habitants. Le Nigéria aurait non seulement distancé largement le Japon, mais il serait alors plus peuplé que les États-Unis, pour un territoire 10 fois plus petit. Le tableau D5.1 montre que les prévisions faites par la Banque mondiale en 1995 sont en bonne voie de se réaliser.

Tableau D5.1 - Population du Nigéria et du Japon

	Nigéria	**Japon**
	(en millions)	
1931	*20*	*65*
1970	56,1	104,3
1993	103,1	124,5
2003	132,6	127,7
2013	173,6	127,3
Prévision 2025	*238*	*122*

Source : Banque mondiale, Rapport sur le développement dans le monde 1995; IDM 2014; Nihon Sōmushō Tokeikyoku.

Essayons d'avancer quelques hypothèses à propos de cette situation. Pour cela, il n'est pas nécessaire de bien connaître le Nigéria ni le Japon. Faisons plutôt appel à notre logique. Nous irons chercher plus tard les données qui nous permettront d'évaluer nos hypothèses.

Tout d'abord, il est probable que les mères nigérianes donnent naissance à plus d'enfants que les mères japonaises. Il se pourrait aussi que le Nigéria accueille relativement plus d'immigrants que le Japon. Ce sont les deux explications les plus évidentes, mais on peut en proposer d'autres. Ce n'est pas tout d'avoir de nombreux enfants, encore faut-il qu'ils atteignent l'âge de procréer pour que la boucle soit bouclée : se pourrait-il que les enfants nigérians meurent moins facilement que les enfants japonais? Par ailleurs, avez-vous songé que si les gens vivaient deux fois plus longtemps, la population d'un pays doublerait sans l'aide d'autres facteurs : serait-il possible que les chances de vivre vieux soient en train d'augmenter au Nigéria?

Nous avons donc formulé quatre hypothèses, qu'il va falloir accepter ou rejeter à la lumière des chiffres.

- 1. Les mères nigérianes ont plus d'enfants que les mères japonaises.
- 2. L'immigration est plus importante au Nigéria qu'au Japon.
- 3. Moins de Nigérianes que de Japonaises meurent avant de mettre des enfants au monde.

- 4. Les chances de vivre de plus en plus vieux augmentent plus rapidement au Nigéria qu'au Japon.

Comment mesurer la fécondité d'une femme?

Commençons par explorer le concept du nombre d'enfants par femme. Une Nigériane a-t-elle généralement plus d'enfants qu'une Japonaise? Pour être plus précis, quel est le nombre moyen d'enfants d'une femme nigériane (nous prendrons l'année 1993 comme base de nos calculs)? Il est difficile de répondre directement à de telles questions, car il peut s'écouler un certain nombre d'années entre la naissance de l'aîné et celle du petit dernier. Et pendant ce temps-là, les mœurs familiales peuvent évoluer rapidement.

Supposons que, dans un village typique, les femmes de 80 ans aient eu en moyenne 9 enfants, les femmes de 60 ans aient eu 6 enfants et les femmes de 40 ans en aient eu 3. Il est clair que le nombre d'enfants par femme est en baisse d'une génération à l'autre. Mais quel chiffre retenir pour mesurer l'état *actuel* des choses? À quelle génération appartient une femme de 50 ans? La femme de 40 ans a-t-elle fini de donner naissance à des enfants? Il y a un moyen bien simple de réconcilier tous ces faits. Il suffit d'observer le nombre de naissances par femme pour chaque âge où cette dernière peut procréer. Nous combinerons le tout en dressant un portrait type fabriqué à partir des caractéristiques des mères de tous âges. Examinons l'extrait de tableau suivant.

Tableau D5.2 - Calcul de l'indice synthétique de fécondité

Âge de la femme	..	20	21	22	..
Nombre de femmes	..	100	100	80	..
Naissances cette année-là	..	10	8	6	..
Nombre moyen d'enfants par femme pour cet âge-là	**..**	**0,1**	**0,08**	**0,075**	**..**

Le nombre moyen d'enfants par femme (ou *taux de fécondité*) est obtenu en divisant le nombre de naissances par le nombre de femmes. Ainsi, même si la formulation peut sembler cocasse, une femme a en moyenne 0,1 enfant entre son 20^{e} et son 21^{e} anniversaire. Pour les 3 âges inscrits au tableau (de 20 à 22 ans), le nombre moyen d'enfants serait de 0,1 + 0,08 + 0,075 = 0,255. En additionnant ainsi tous les taux de fécondité pour chaque âge de la vie d'une mère, on obtiendra le nombre moyen d'enfants par femme à un moment donné. Ce chiffre (qu'on appelle *indice de fécondité*) représente donc une synthèse, un instantané, de toutes les générations de femmes en âge de procréer à une époque donnée.

Il nous suffit maintenant d'obtenir des données sur l'indicateur choisi, ce qui est relativement facile dans le domaine de la démographie.

En 1993, cet indice de fécondité est de 6,4 pour le Nigéria, ce qui signifie qu'en moyenne, une Nigériane met au monde environ 6 enfants. C'est beaucoup plus qu'il n'en faut pour assurer la survie de la génération suivante. On peut aussi parier que les familles de 12 enfants ou plus ne doivent pas être rares (étant donné que certaines femmes ont peu d'enfants, ou pas du tout). La famille « moyenne » compte d'ailleurs plus de 24 enfants lorsque le mari a quatre femmes! Même si l'indice de fécondité des Nigérianes est en baisse (il atteignait 6,9 en 1970, mais n'était redescendu qu'à 6,0 en 2013), il n'en demeure pas moins très supérieur à celui du Japon (1,5 enfant en 1993, et 1,4 en 2013). (*Note :* ces chiffres, ainsi que ceux qui suivent, proviennent des *Indicateurs de développement dans le monde* de la Banque mondiale.)

Hypothèse 2 : l'immigration

Notre deuxième hypothèse pour expliquer la progression plus rapide de la population du Nigéria (par rapport à celle du Japon) était l'importance de l'immigration. Mais, contrairement au Canada où l'immigration contribue pour un bon tiers à la croissance de la population, le Nigéria et le Japon accueillent relativement peu d'immigrants. Cette hypothèse ne peut donc servir à éclairer la situation qui nous intéresse.

Hypothèse 3 : la mortalité infantile

Le taux de mortalité infantile pourrait nous aider à vérifier notre troisième hypothèse. En 1993, au Nigéria, sur 1000 enfants, 84 meurent avant l'âge d'un an et 183 meurent avant l'âge de 5 ans. Au Japon, les taux sont respectivement de 5 pour 1000 (avant 1 an) et de 6 pour 1000 (avant 5 ans). Même si les mères japonaises mettent moins d'enfants au monde, ceux-ci ont plus de chances d'arriver à l'âge où ils pourront à leur tour se reproduire. En 2012, le taux de mortalité infantile (avant 5 ans) atteint encore 124 pour 1000 au Nigéria, tandis qu'il est passé à 3 pour 1000 au Japon. Cette hypothèse joue nettement à l'avantage du Japon.

Hypothèse 4 : l'espérance de vie

Il nous reste une dernière hypothèse, celle qui concerne la durée de la vie. En moyenne, un Nigérian qui vient au monde peut espérer vivre 52 ans en 2012 (contre 83 ans au Japon). Si les Nigérians vivaient aussi longtemps que les Japonais, ils pourraient donc être beaucoup plus nombreux qu'ils ne le sont actuellement. Or, l'espérance de vie à la naissance des Nigérians ne cesse d'augmenter : elle n'était que de 41 ans en 1970, contre 46 ans en 1993 et 52 ans en 2012. Entre 1970 et 2012, l'espérance de vie des Japonais augmente également de 11 ans, pour atteindre 83 ans, ce qui n'est pas un mince exploit compte tenu du fait que la barre était déjà très haute! Cependant, toutes proportions gardées, la performance relative du Nigéria (52 – 41/41 = +27 %) lui donne ici un avantage sur le Japon (83 – 72/72 = +15 %)

Une *méthode quantitative*

Dans cette courte et modeste étude, nous avons cherché à analyser un phénomène humain en utilisant des chiffres de façon méthodique. Pour cela, nous avons commencé par réfléchir sur la mécanique qui sous-tend l'évolution des effectifs d'une population. Nous avons alors élaboré quatre hypothèses susceptibles d'expliquer la situation : si les deux premières hypothèses étaient plutôt évidentes, les deux dernières étaient par contre relativement subtiles. Il ne nous restait plus qu'à choisir des indicateurs précis correspondant à ces hypothèses et à obtenir des chiffres sur ces indicateurs.

Nous pouvons conclure que la hausse rapide de la population nigériane s'explique surtout par le nombre élevé d'enfants par femme (hypothèse 1) et, dans une moindre mesure par l'allongement de l'espérance de vie (hypothèse 4). Or, même si les différences entre le Nigéria et le Japon venaient à s'estomper rapidement, la machine nigériane ne s'arrêterait pas de sitôt, ne serait-ce parce que les nombreuses petites filles d'aujourd'hui deviendront les nombreuses mères de demain.

DÉSAMORCER LA BOMBE : L'EXPÉRIENCE CHINOISE

En 1995, la population chinoise atteint 1,2 milliard d'habitants, soit 21,1 % du total mondial. À cause du fort taux de natalité enregistré dans les décennies précédentes, et notamment à l'époque de la Révolution culturelle (1966-1976), la Chine compte alors une proportion importante de femmes et d'hommes en âge de procréer. Si on se contentait de limiter le nombre d'enfants à deux par famille, la croissance de la population ne pourrait être freinée avant la prochaine génération. Le gouvernement chinois juge qu'un tel niveau de surpopulation pose non seulement un énorme obstacle au développement économique, mais qu'il menace l'environnement et la paix mondiale. Il n'est donc pas question d'abandonner la politique de planning familial, dite de « l'enfant unique », mise en route en 1979.

Les joies de la piscine en Chine

Pour désamorcer la bombe, la politique de planning familial chinoise a été orientée en fonction des paramètres suivants (*source : Jinri Zhongguo*, Pékin, mars 1996) :

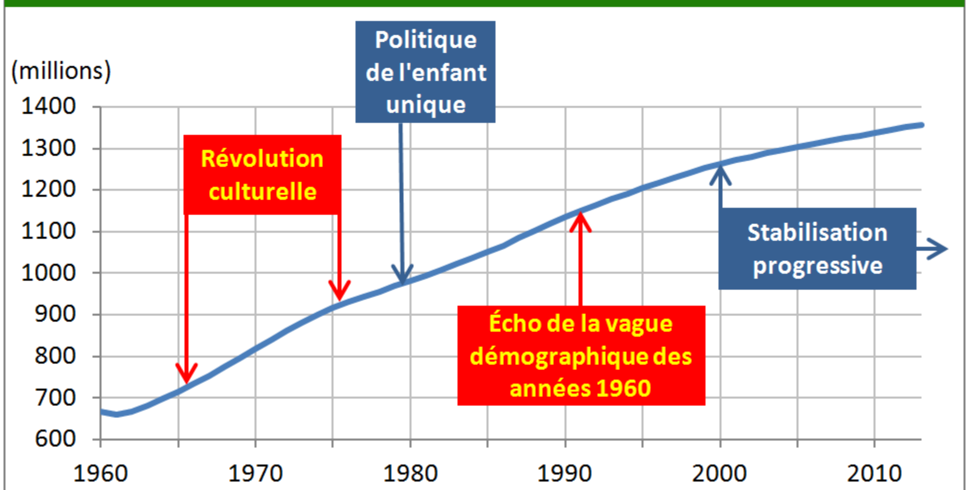
- Mariage tardif (25 ans pour les garçons et 23 ans pour les filles).
- Procréation tardive (24 ans pour les filles).
- Limitation des naissances (1 enfant par couple, sauf exception : enfants handicapés, minorités nationales, situations familiales particulières).
- Accouchement de qualité (la santé du nouveau-né et de la mère doivent être garantis).

On le devine, ces mesures s'accompagnent d'une infrastructure appropriée (contraception, cliniques d'avortement) et d'une série d'incitatifs matériels (primes et congés pour le mariage et la procréation tardive, pertes d'avantages sociaux pour un trop grand nombre de naissances dans la même famille).

En 2013, le taux de croissance annuel de la population chinoise n'est plus que de 0,5 % (contre 2,8 % au plus fort de la Révolution culturelle). En 2013, la population de la Chine représente « seulement » 19,1 % du total mondial (contre 22,6 % en 1974, record historique).

Beaucoup, en Occident, se sont mis à critiquer la Chine, au moment où ce pays semblait prendre ses « responsabilités démographiques ». Cela est d'autant plus paradoxal que la Chine était plutôt adulée au moment de la soi-disant Révolution culturelle. Or, on s'aperçoit que les mesures de la politique familiale chinoise s'appuient sur une étude méthodique des données démographiques. Toute critique de cette politique doit faire preuve d'autant de rigueur. N'est-ce pas d'ailleurs le but des méthodes quantitatives en sciences humaines?

Figure D5.1 - Évolution de la population chinoise



Source des données brutes : Banque mondiale, IDM.

QUESTIONS

1. Le taux et l'indice de fécondité

a) Sur 100 femmes de 15 ans, 4 mettent un enfant au monde au cours de l'année. Le taux de fécondité est le même pour les femmes âgées de 16 à 24 ans. Les femmes âgées de 25 ans à 34 ans ont un taux de fécondité de 8 pour 100 et celles âgées de 35 à 44 ans ont un taux de fécondité de 6 pour 100. Enfin, les femmes âgées de moins de 15 ans et de plus de 45 ans n'ont jamais d'enfants. Calculez l'indice synthétique de fécondité.

b) Sachant que dans un pays, les femmes sont fécondes pendant 30 ans et que le taux de fécondité de chaque groupe d'âge est de 8 naissances pour 100 femmes, quel est l'indice synthétique de fécondité? (Chiffres fictifs)

2. Recherche

a) Choisissez deux pays différents en matière d'évolution démographique, formulez des hypothèses pour expliquer ces différences et trouvez des chiffres pour confirmer ou infirmer vos hypothèses.

b) Obtenez des chiffres plus anciens ou plus récents sur le Nigéria et le Japon et interprétez-les.

CHAPITRE 6 L'ANALYSE DE TABLEAUX

TABLE DES MATIÈRES

1. Trois clés : Définir, observer, interpréter
2. Analyser un tableau : faut-il se faire des cheveux blancs?
3. Analyser un graphique : les filles en pantalon
4. Construire ses propres tableaux

• Exercices supplémentaires

• Dossier

La personne qui s'intéresse aux sciences humaines demeure parfois perplexe devant un tableau présentant des chiffres. Il est pourtant facile de lire et interpréter un tableau si on respecte quelques règles de base. Ce sont ces principes que nous allons étudier dans ce chapitre. Nous profiterons aussi de l'occasion pour faire la synthèse de plusieurs notions vues jusqu'ici, avant de nous lancer dans la deuxième partie de ce manuel.

Nous commencerons par présenter, dans une première section, les principes simples qui permettent de faire de l'analyse de tableau une activité riche d'enseignements. Puis nous proposerons, dans les trois autres sections, trois situations typiques, toutes reliées à différents domaines des sciences humaines. La première situation sera présentée sous forme de tableaux, la deuxième sous forme de graphiques (une manière d'« habiller » un tableau) et la dernière sous forme de texte (une façon de « noyer » un tableau). Nous verrons que ce sont à peu près toujours les mêmes principes qui s'appliquent dans tous les cas.

Vous disposerez alors d'une arme redoutable pour comprendre (en lisant des tableaux) et faire comprendre (en produisant vos propres tableaux) les aspects quantitatifs de la réalité humaine.

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

- Comment peut-on tirer le maximum de l'analyse d'un tableau?
- Pourquoi est-il primordial de bien cerner la population et les caractéristiques de la population sur laquelle portent les observations consignées dans un tableau?
- Comment peut-on transformer les données d'un tableau pour les rendre plus explicites?
- Comment un tableau peut-il être générateur d'hypothèses?
- Comment peut-on construire ses propres tableaux?

1. TROIS CLÉS : DÉFINIR, OBSERVER, INTERPRÉTER

L'analyse d'un tableau se fait en trois temps. On commence par *définir* les éléments qui constituent le tableau. Puis on se met à *observer* les données elles-mêmes. Enfin, on cherche à *interpréter* les données observées. Pour parler plus simplement, on se pose successivement les trois questions suivantes : de quoi parle-t-on? Quels sont les chiffres? Que signifient-ils?

Ces trois étapes sont décrites de façon détaillée dans les paragraphes qui suivent. Toutefois, il est bien connu qu'on est souvent plus pressé d'écouter de la musique que de lire le mode d'emploi de l'appareil audio qu'on vient d'acheter. Si c'est votre cas, vous pouvez passer directement aux applications qui nous servent à illustrer la démarche, quitte à vous y référer au fil des commentaires.

1.1. La démarche à suivre

A. Définir les éléments du tableau

À cette étape-ci, on considère surtout la structure du tableau et non pas ce qui s'y trouve.

- 1. À quelle population le tableau fait-il référence et sous quel angle (quelles caractéristiques) la population est-elle étudiée? Quelle est l'échelle de mesure de ces caractéristiques? Dans quelle unité les données sont-elles exprimées? Les données sont-elles brutes ou dérivées? En d'autres termes, résultent-elles d'un traitement quelconque? Si elles sont dérivées, s'agit-il de proportions, d'autres rapports, de moyennes, de taux de variation, d'indices?
- 2. Y a-t-il des *relations* entre les variables? Si oui, qui dépend de quoi? Pour répondre à ces deux questions, il faut déterminer les *variables*, compte tenu de ce que l'on cherche à comprendre. On peut, par exemple, considérer chaque caractéristique de la population comme une variable. Dans d'autres cas, on peut trouver plus intéressant de considérer chaque « colonne » de chiffres comme une variable.
- 3. Quel est le *domaine* étudié (le sujet, le lieu, l'époque)? Les données sont-elles chronologiques (évolution sur une période donnée, et avec quelle fréquence) ou statiques (photographie à un moment précis)? Les données sont-elles homogènes (y compare-t-on les mêmes années, les mêmes réalités)? Quelles sont les sources des données? Ces sources sont-elles fiables? Peut-on les consulter?

B. Observez les données

Ce n'est qu'à partir d'ici qu'on commence réellement à regarder les chiffres proprement dits.

- 1. *Décrivez* l'évolution d'une des variables. Quelle est son allure générale (valeurs habituelles, tendance à long terme, changement radical) et quelles sont ses spécificités (évolution à court terme, éléments particuliers)? Y a-t-il des valeurs ou des variables inattendues?
- 2. *Comparez* les différentes variables (même si on ne croit pas qu'elles sont nécessairement reliées). Se suivent-elles? Vont-elles en sens inverse? Où s'arrêtent leurs relations?
- 3. S'il y a lieu, *transformez* les données (calculez des rapports, des moyennes, des taux de variation, des indices) pour faciliter l'observation ou mettre certains éléments en évidence. À l'inverse, on peut essayer de reconstituer les données originales (qui se cachent derrière des rapports ou des taux de variation, par exemple).

C. Interprétez les données

Après avoir défini et observé les données, on essaie de comprendre ce que les chiffres nous disent et ce qu'ils pourraient nous laisser supposer.

- 1. Émettez des *hypothèses* qui pourraient expliquer les faits observés.
- 2. *Distinguez* dans les hypothèses ce qui est démontrable (d'après les données du tableau) de ce qui est spéculatif (c'est-à-dire ce qui est seulement possible ou probable).
- 3. *Recherchez* d'autres données qui pourraient éclairer (confirmer ou infirmer) les hypothèses.

1.2. Première application : Les fermes canadiennes

Ces étapes vous semblent sans doute un peu abstraites présentées ainsi. Voyons, avec deux exemples concrets, comment on peut les appliquer à des données réelles tirées d'un rapport sur l'activité humaine et l'environnement.

Notre premier tableau a trait aux fermes canadiennes (tableau 6.1a).

Tableau 6.1a - Les fermes canadiennes

	Superficie totale (millions hectares) [1]	**Nombre de fermes** (milliers) [2]	**Superficie moyenne** (hectares) [3]
1901	25,7	551,1	46,6
1911	44,1	682,8	64,6
1921	57,0	711,1	80,2
1931	66,0	728,6	90,6
1941	70,2	732,9	95,8
1951	70,4	623,1	113,0
1961	69,8	480,9	145,1
1971	68,7	366,1	187,7
1981	65,9	318,4	207,0
1991	67,8	280,0	242,1
2001	67,5	246,9	273,4
2011	64,8	205,7	315,0

Source des données brutes : Statistique Canada, L'activité humaine et l'environnement 1994; Recensement de l'agriculture 3438.

Note : 1 ha (hectare) = 100 m × 100 m = 10 000 m².

A. Définir

La population à laquelle le tableau fait référence est l'ensemble des fermes canadiennes. La caractéristique étudiée est la superficie de chaque ferme (on aurait pu étudier d'autres caractéristiques comme le revenu ou la grosseur du troupeau de vaches de chaque ferme). Théoriquement, le tableau aurait pu être construit à partir d'une liste contenant pour chaque ferme sa superficie respective. Le nombre de fermes constitue l'effectif de la population et la superficie de chaque ferme est la seule variable étudiée. La superficie est mesurée sur une échelle de rapport et son unité de mesure est le million d'hectares (1 hectare = 100 mètres × 100 mètres = 10 000 m^2).

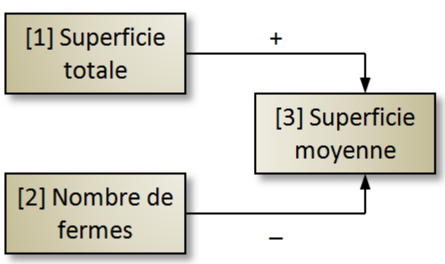
Il y a cependant une autre manière de voir les choses. Comme nous ne possédons pas cette fameuse liste des fermes (qui n'a peut-être d'ailleurs jamais existé comme telle) et puisqu'on nous donne des chiffres sur l'évolution de trois éléments qui semblent reliés entre eux, nous pouvons décider de considérer chaque colonne comme une variable : 1. superficie totale (en millions d'hectares), 2. nombre de fermes (en milliers d'unités), 3. superficie moyenne (en hectares). On remarque que les variables 1 et 3 utilisent la même unité de mesure de base, mais avec un facteur différent, ce qui permet de mieux visualiser les données*.

De façon similaire, on compte généralement les habitants d'un pays en millions et les membres d'une famille en unités, même si dans les deux cas on additionne des individus.

On note que notre variable 3 est une donnée dérivée (alors que les autres variables sont des données brutes). Plus précisément, la variable 3 représente le *rapport** entre la variable 1 et la variable 2. Le schéma de la figure 6.1 illustre la relation entre les trois variables. La variable 1 influence la variable 3 de façon *directe* : plus la variable 1 est grande, plus la variable 3 est grande. La variable 2 influence la variable 3 de façon *inverse* : plus la variable 2 est grande, plus la variable 3 est petite. Sur le schéma, nous indiquons le sens de la relation par le signe plus (+) pour la relation directe et par le signe moins (–) pour la relation inverse.

C'est également une *moyenne*, forme particulière de rapport, qu'on obtient en divisant le total (variable 1) par le nombre d'unités (variable 2). Ainsi, même si on ne possède pas la liste de toutes les fermes avec leur superficie respective, on est en mesure de calculer leur superficie moyenne.

Figure 6.1 - Relations entre variables



Pour terminer cette première étape (*définir*), voici quelques mots sur le *domaine* couvert. Le sujet étudié est l'agriculture. Les données sont chronologiques. Elles portent sur une période d'environ un siècle et l'intervalle entre chaque relevé (la périodicité) est de 10 ans. Il s'agit d'une période et d'intervalles relativement longs (par rapport à une vie humaine par exemple) : on nous présente une évolution à long terme. Les données portent sur l'ensemble du Canada. Le document est publié par Statistique Canada et peut être consulté facilement, si nécessaire.

B. Observer

La superficie totale cultivée au Canada (variable 1) augmente rapidement au début du XXe siècle. Par la suite, elle plafonne puis elle décline légèrement à partir de 1951. Le nombre de fermes (variable 2) augmente aussi au début du XXe siècle, quoique de façon apparemment moins rapide que pour la variable 1, plafonne dès les années 1930 et décroît ensuite de façon systématique, surtout après la Deuxième Guerre mondiale. La superficie moyenne des fermes (variable 3) ne cesse d'augmenter tout au long de la période considérée.

Pour mieux observer la vitesse à laquelle se produisent les changements, on pourrait calculer des taux de variation pour chaque décennie (tableau 6.1b). On a alors la confirmation que la superficie totale (variable 1) a augmenté plus vite que le nombre de fermes (variable 2) jusqu'en 1931. Le calcul des taux de variation permet de mettre clairement en évidence une valeur particulière : de 1931 à 1941, la superficie moyenne a très peu augmenté (colonne 3).

Tableau 6.1b - Le tableau 6.1a vu sous un autre angle

	Taux de variation décennale (en %)			**Indice de variation 1901 = 100**		
	Superficie totale	**Nombre de fermes**	**Superficie moyenne**	**Superficie totale**	**Nombre de fermes**	**Superficie moyenne**
	[1]	[2]	[3]	[1]	[2]	[3]
1901				100	100	100
1911	71,6	23,9	38,5	172	124	138
1921	29,3	4,1	24,1	222	129	172
1931	15,8	2,5	13,0	257	132	194
1941	6,4	0,6	5,7	273	133	205
1951	0,3	-15,0	18,0	274	113	242
1961	-0,9	-22,8	28,5	272	87	311
1971	-1,6	-23,9	29,3	267	66	402
1981	-4,1	-13,0	10,3	256	58	444
1991	2,9	-12,1	17,0	264	51	519
2001	-0,4	-11,8	12,9	263	45	586
2011	-4,0	-16,7	15,2	252	37	676

C. Interpréter

Que peut-on conclure de ces observations? Il va de soi que cette partie est la plus délicate, car émettre des hypothèses, c'est jouer sur un terrain glissant. C'est pourtant ce que nous avons fait tout au long de ce manuel, non pas en prétendant découvrir une vérité absolue, mais dans le but de stimuler la réflexion. Après tout, pour un chercheur débutant, il faut commencer par avoir de nombreuses idées avant de détenir quelques rares certitudes.

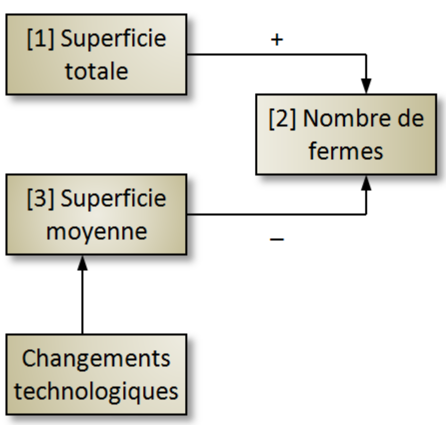
Commençons par ce qui est démontrable par la simple logique. Il est clair que l'évolution de la variable 3 s'explique par la variation des 2 autres (revoir la figure 6.1). La superficie moyenne augmente d'abord à cause d'une augmentation plus rapide de la superficie totale que du nombre de fermes (jusqu'en 1941). Par la suite, la variable 3 augmente à cause de la baisse de la variable 2.

On peut imaginer que, jusque dans les années 1930, une partie du territoire cultivable canadien était encore vierge. Cette hypothèse pouvoir se vérifier facilement, et si jamais on s'apercevait du contraire, les choses s'avèreraient encore plus intéressantes. Le léger déclin après 1951 peut s'expliquer par l'empiètement des villes sur les terres agricoles et l'abandon de terres peu fertiles. Tout au long du siècle, les progrès techniques (machines agricoles, recherche biologique, mise en marché, informatisation) ont à la fois permis et exigé l'exploitation de surfaces de plus en plus grandes avec de moins en moins de travailleurs. Certains fermiers ont dû vendre et quitter le métier, d'autres se sont agrandis. Comme la plupart des fermes sont des entreprises familiales (hypothèse à vérifier aussi), la superficie moyenne augmente, mais pas trop rapidement : il faut que le cultivateur demeure capable de gérer son espace.

On serait tenté de croire, à première vue, que le ralentissement observé dans les années 1930 (la superficie moyenne augmentant très peu comparativement aux autres périodes) est causé par la dépression économique qui sévissait alors. On pourrait faire l'hypothèse que les gens s'accrochaient à la ferme comme à un refuge dans une époque de chômage endémique, d'où une hausse du nombre de fermes. On pourrait aussi émettre l'hypothèse que beaucoup de fermiers endettés devaient céder leur propriété aux créanciers (d'où une baisse du nombre de fermes). Ces deux hypothèses, dont les effets sont contradictoires, mériteraient une recherche plus approfondie. Cependant, n'y a-t-il pas une explication plus simple? Au cours des années 1930, la superficie cultivée a déjà cessé d'augmenter (colonne 1 du tableau 6.1b) alors que le nombre de fermes ne commencera à baisser que dans la décennie suivante (colonne 2). On est assis entre deux chaises : il n'est alors pas étonnant que la hausse de la superficie moyenne marque un léger temps d'arrêt pendant cette période charnière (colonne 3).

Nous avons affirmé précédemment que la superficie moyenne (variable 3) dépend des deux autres variables (superficie totale et nombre de fermes). Cela paraît évident sur le plan mathématique parce que la superficie *moyenne* s'obtient en divisant la superficie *totale* par le *nombre* de fermes. Toutedois, sur le plan humain, n'est-ce pas le contraire qui se produit? N'est-ce pas l'augmentation systématique de la superficie moyenne des fermes, en raison des progrès de la mécanisation, de la gestion, de la recherche génétique, etc., qui est la cause de l'accroissement des exploitations agricoles et, par conséquent, de la disparition ou de la fusion d'un grand nombre de fermes (voir la figure 6.2)?

Figure 6.2 - Relations revues et corrigées



Au terme de cette première tentative, certains lecteurs pourraient se dire qu'analyser un tableau est une chose bien compliquée : nous semblons hésiter sur le choix des variables et la façon dont elles s'influencent. C'est vrai, il n'y a pas de recette magique, mais n'avez-vous pas l'impression que cette démarche vous a permis de comprendre bien des choses? Derrière la froideur des chiffres, on peut découvrir un univers humain riche et passionnant.

1.3. Seconde application : La morue se fait rare

Quel rapport entre la morue et les sciences humaines, diront certains? Cela nous paraît pourtant évident. La pêche, tout comme l'agriculture, illustre les relations entre l'être humain et la nature. Nous nous intéresserons particulièrement au tournant des années 1990, lorsque les stocks de morue, réputés inépuisables depuis une époque antérieure à la découverte du Canada, s'effondrèrent brusquement.

A. Définir

Dans le tableau 6.2, on pourrait considérer qu'il n'y a qu'une variable (les prises de morues) et que cette variable (mesurée en milliers de tonnes) est distribuée sur une échelle nominale qui contient deux catégories : pêcheurs canadiens et pêcheurs étrangers.

Tableau 6.2 - Les prises de morue dans l'Atlantique canadien

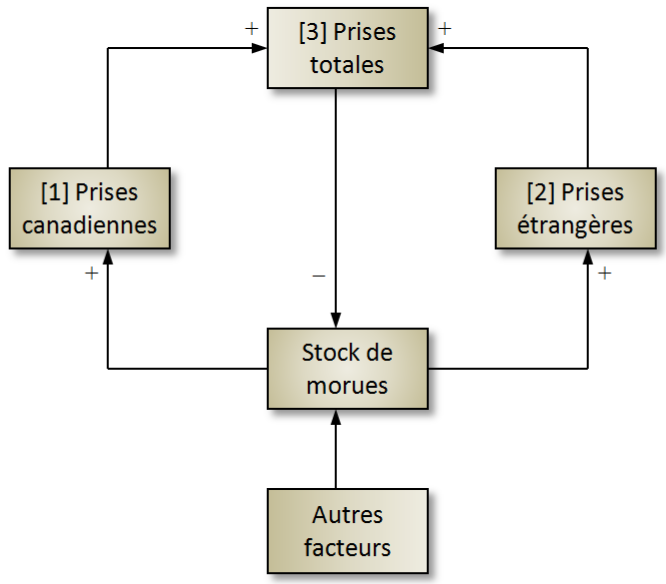
	Prises canadiennes	Prises étrangères	Total
		(en milliers de tonnes)	
	[1]	[2]	[3]
1980	418,0	90,0	508,0
1981	432,0	97,0	529,0
1982	510,0	102,0	612,0
1983	507,0	86,0	593,0
1984	492,0	82,0	574,0
1985	474,0	119,0	593,0
1986	452,0	166,0	618,0
1987	452,0	102,0	554,0
1988	468,0	77,0	545,0
1989	426,0	118,0	544,0
1990	384,0	111,0	495,0
1991	304,0	106,0	410,0
1992	192,0	47,0	239,0

Source des données : Statistique Canada, L'activité humaine et l'environnement 1994.

Ici encore, on pourrait voir les choses autrement dans l'espoir de mettre en évidence les relations entre divers éléments. Nous aurions alors trois variables : les prises canadiennes (variable 1), les prises étrangères (variable 2) et les prises totales (variable 3). Ne perdons pas de vue que la variable 3 est la somme des deux autres. Cette manière de voir peut se justifier si nous considérons que ce qui sépare les pêcheurs canadiens et étrangers (traditions, lois, conditions économiques, etc.) est plus important que ce qui les unit (le stock commun de morues).

La figure 6.3 illustre cette façon de relier nos trois variables. Il est clair que les prises (*flux*) diminuent les *stocks* de morue et que la baisse des stocks de morue entraîne une baisse (volontaire ou non) des prises. D'autres facteurs (température de l'eau, pollution, courants) peuvent aussi influencer la façon dont les stocks se reproduisent.

Figure 6.3 - Relations entre variables



Nous avons ici des données chronologiques. Compte tenu des circonstances (saisons de pêche, cycle annuel de reproduction des poissons, utilisation des bateaux sur plusieurs années), la période (10 ans) et la périodicité (1 an) sont particulièrement bien choisies. Elles nous permettent de mettre à la fois en relief les fluctuations annuelles et l'évolution générale des prises à moyen terme.

On pourrait compléter le tableau 6.2 en calculant les proportions respectives des prises canadiennes et étrangères dans le total et en calculant des indices simples basés sur l'année 1980 = 100. Nous vous laissons ce soin.

B. Observer

Les trois variables (les trois colonnes) fluctuent autour d'une même valeur jusqu'en 1990. En d'autres termes, ça varie d'une année à l'autre (phénomène à court terme), mais ça reste dans les mêmes ordres de grandeur (phénomène à long terme). Puis on observe une chute (qui devient spectaculaire en 1992).

C. Interpréter

Une façon de noyer le poisson serait d'une part pour le Canada d'accuser les étrangers (les chalutiers espagnols) de ne pas avoir diminué leurs prises assez rapidement (en 1991). Pour les étrangers, il s'agirait de mettre en évidence le fait que le Canada (dont les prises sont trois fois plus élevées que celles de l'étranger) n'avait besoin de personne pour l'aider à exterminer les poissons. Ou peut-être que tout cela est la faute de Brigitte Bardot, dont les campagnes anti-fourrure ont causé une prolifération du plus gros des consommateurs de morue, le vorace phoque!

EXERCICES 1

1. Les poissons chevauchants

a) Quelle est la population étudiée dans le tableau 6.3? Quelle est la variable? Quelle est son échelle de mesure? Les données sont-elles brutes ou dérivées? Quelle est leur unité de mesure? Dans le cas de la morue (tableau 6.2), la comparaison se faisait dans le temps. Qu'en est-il ici?

b) Calculez la répartition (en proportion du total) des prises de chaque zone. Évaluez, à l'aide d'un rapport, l'importance des stocks chevauchants dans la pêche mondiale.

c) Quel est l'intérêt d'étudier les poissons chevauchants?

Tableau 6.3 - Prises sur stocks de poissons chevauchants (ou pouvant le devenir)

Mer ou océan	**1991** (en milliers de tonnes)
Atlantique	2 001
Méditerranée	156
Océan indien	131
Pacifique	10 186
Total	12 474
Prises mondiales	**81 749**

Source : FAO dans L'Observateur de l'OCDE, août-septembre 1995.

Note : Les zones de pêches sont divisées en zones nationales (jusqu'à 200 milles des côtes) et zones hauturières (en haute mer). Mais certains poissons ignorent les frontières et chevauchent les deux zones. Par ailleurs, à titre d'information, la population mondiale était de 5,356 milliards d'habitants en 1991.

2. Les amateurs de poisson

a) Quelle est la population étudiée dans le tableau 6.4? Quelle est la variable? Quelle est son échelle de mesure? Les données sont-elles brutes ou dérivées? Quelle est leur unité de mesure?

b) Comparez la consommation mondiale par habitant de poissons et coquillages (dernière ligne du tableau 6.4) à la production mondiale par habitant (que vous calculerez à partir des données du tableau 6.3). Commentez.

c) Observation. Quel genre de pays retrouve-t-on en tête (ou en queue, ou au milieu)? Ont-ils des points communs?

c) Interprétation. Essayez de proposer des hypothèses historiques, culturelles ou géographiques qui permettraient, une fois vérifiées, d'expliquer les différences entre pays. Expliquez en particulier le cas de l'Islande, du Portugal et de l'Autriche.

Tableau 6.4 - Consommation annuelle de poissons et de coquillages dans les pays de l'OCDE

	Moyenne 1988-90
	(en kg par habitant)
Islande	92,1
Portugal	60,2
Norvège	41,1
Espagne	38,0
France	31,1
Finlande	30,6
Nouvelle Zélande	28,9
Suède	26,9
Canada	24,3
États-Unis	21,3
Danemark	21,2
Italie	20,7
Royaume-Uni	19,9
Grèce	19,1
Australie	18,8
Belgique	18,8
Irlande	15,9
Suisse	13,3
Allemagne	12,2
Mexique	11,0
Pays-Bas	9,8
Autriche	8,8
Turquie	6,3
Monde	**13,4**

Source : Ministère américain du commerce extérieur dans L'Observateur de l'OCDE, août-septembre 1995.

2. ANALYSER UN TABLEAU : FAUT-IL SE FAIRE DES CHEVEUX BLANCS?

La population vieillit en Occident, à ce qu'il paraît. Ce refrain ne date pas d'aujourd'hui, puisqu'on l'entendait déjà dans les années 1990. Comment voyait-on les choses à l'époque? Quelle fut la réalité 20 ans plus tard? Pour répondre à ces questions, retournons d'abord quelques décennies en arrière et plaçons-nous en pensée, au mois de février 1996, date à laquelle a été élaborée l'analyse de tableau qui va suivre.

Ce mois-là, le gouvernement canadien s'aperçoit soudainement que le régime national de pension s'en va à vau-l'eau. Bref, les réserves seront à sec en 2015. Le gouvernement canadien, face à cette triste découverte, laisse entendre qu'on devra peut-être baisser le montant des pensions distribuées ou encore retarder l'âge de la retraite. À Québec, tout en n'étant pas d'accord avec Ottawa sur les solutions à apporter, on semble tout aussi surpris. C'est comme si un météorite venait subitement de faire son apparition dans l'orbite terrestre. Le peuple est inquiet et les élites promettent seulement d'adoucir le choc inévitable.

Devant un aussi beau concert, il y a de quoi rester sceptique (on a l'impression que les musiciens ont répété en cachette). Il est peu probable que nos dirigeants aient attendu un jour de février 1996 pour se rendre compte qu'une personne née en 1950 aura pas mal de chances d'atteindre les 65 ans en l'an 2015. Alors, il nous faut des chiffres, s'il vous plaît! Compte tenu de ce que nous connaissons de la situation en 1996, nous voulons savoir, par exemple, si c'est vrai qu'il y aura « trop » de personnes âgées en 2015. Par la même occasion, nous tenterons de vérifier quelles seront alors les chances des futurs grands-pères et grands-mères, de dénicher un éventuel partenaire du sexe opposé. Place au tableau 6.5; définissons, observons et interprétons.

Tableau 6.5 - Les personnes âgées au Québec jusque dans les années 1990

	Population de 65 ans et plus (en milliers)			Rapport de masculinité (en %)		
	Hommes	Femmes	**Total**	55-64 ans	65-74 ans	75 ans et plus
	[1]	[2]	[3]	[4]	[5]	[6]
1961	144,7	161,6	**306,3**	97	92	85
1971	183,5	235,8	**419,3**	94	82	70
1981	237,6	336,3	**573,9**	89	78	59
1991	316,1	465,1	**781,2**	92	78	55
1993	334,6	492,6	**827,2**	93	78	54

	Rapport de dépendance (en %)			
	65 ans et plus/ Population	65 ans et plus/ 0 à 14 ans	65 ans et plus/ 15-64 ans	[65 ans et plus et 0 à 14 ans]/ 15-64 ans
	[7]	[8]	[9]	[10]
1961	5,8	16,4	9,9	70,2
1971	6,8	23,3	10,7	56,5
1981	8,7	40,7	12,5	44,3
1991	11,0	55,7	16,0	44,6
1993	11,5	58,3	16,7	45,3
2036	*22*	*129*	*36*	*64*

Source des données brutes : Le Québec statistique 1995; Statistique Canada, Cansim 051-0001.

Note : tous les chiffres concernent le Québec, sauf les données de 2036 qui ont trait au Canada. Les données de 2036 sont déduites des prévisions de Statistique Canada dans Statistiques démographiques annuelles 1993.

2.1. Définir

Le tableau 6.5 traite de la population du Québec à différentes périodes. Cette population est vue sous deux angles : le groupe d'âge et le sexe, que nous considérerons comme nos variables, et que nous étudierons à travers le temps. À l'origine, l'âge se mesurait sur une échelle de rapport, mais il a ici été découpé en classes et l'échelle est devenue ordinale. Ce découpage peut d'ailleurs varier : selon ce qui nous intéresse, on regroupe ou on subdivise les classes. Si, par exemple, on se préoccupe de la pression exercée par les personnes à charge sur celles qui sont actives, on pourrait diviser l'âge de la population en trois catégories : les « enfants », les « personnes actives » et les « retraités ». Il va de soi que ces catégories ne sont pas tranchées au couteau, par contre, il faut bien leur attribuer des bornes précises. Pour le sexe, les choses sont plus simples (il n'y en a que deux) : l'échelle est nominale.

Nos variables retenues : l'*âge* et le *sexe*.

Nous aurions pu traiter chaque colonne du tableau comme une variable séparée, ainsi que nous l'avons suggéré dans les exemples précédents. Nous ne le ferons pas, car nous nous intéressons en premier lieu à l'âge des gens, et accessoirement à leur sexe. Si vous examinez le tableau 6.5, vous verrez en effet que le sexe a manifestement une influence sur la longévité. Nos deux variables, l'âge et le sexe, s'avèrent donc particulièrement bien choisies.

Examinons maintenant les dix colonnes du tableau 6.5. Pour mieux comprendre comment lire chacune de ces colonnes, imaginons ce qui se cache derrière chaque ligne de ce tableau 6.5, et pour cela regardons le tableau 6.6, qui ne contient aucun chiffre, mais seulement une structure.

Tableau 6.6 - Données brutes qui sous-tendent le tableau 6.5

	Hommes	**Femmes**	**Total**
0 à 14 ans	H1	F1	**Total 1**
15 à 55 ans	H2	F2	**Total 2**
55 à 64 ans	H3	F3	**Total 3**
65 à 74 ans	H4	F4	**Total 4**
75 ans et plus	H5	F5	**Total 5**
Total	**Total H**	**Total F**	**Total global**

Seules les trois premières colonnes du tableau nous fournissent des données brutes. Chaque chiffre de la colonne 1 du tableau 6.5 correspond par exemple à (H4 + H5) dans le tableau 6.6.

Dans les colonnes 4 à 6, nous avons des rapports entre les valeurs prises par les catégories d'une même variable (Homme/Femme). Chaque chiffre de la colonne 4 du tableau 6.5 correspond par exemple à (H3 / F3) dans le tableau 6.6. Les données sont exprimées en pourcentage (comme d'ailleurs pour les colonnes suivantes).

Dans les colonnes 8 à 10, nous avons encore des rapports entre catégories, mais cette fois il s'agit l'autre variable (l'âge). Chaque chiffre de la colonne 8 du tableau 6.5 correspond par exemple à (H4 + F4 + H5 + F5)/(H1 + F1) dans le tableau 6.6.

Nous n'avons pas oublié la colonne 7 qui représente la *proportion* de personnes de 65 ans et plus (la partie) dans la population (le *tout*). Chaque chiffre de la colonne 7 du tableau 6.5 correspond à (H4 + F4 + H5 + F5)/(Total global) dans le tableau 6.6.

2.2. Observer

Au lieu de parler de telle ou telle classe d'âge, nous emploierons parfois des termes plus imagés (*enfants*, *personnes âgées*, *retraités*, etc.). On se référera au tableau 6.5 pour la définition précise de chaque classe d'âge.

Colonnes 1 à 3 : une augmentation qui commence à ralentir

Le nombre de personnes âgées (de 65 ans et plus) augmente rapidement entre 1961 et 1991 et cette croissance va en s'accélérant. La croissance ralentit par contre entre 1991 et 1993 (et là, il faut considérer que l'intervalle n'est que de 2 ans au lieu de 10 ans : voir les calculs ci-dessous). Nous n'avons aucune indication sur l'augmentation du reste de la population, même si nous pouvons nous douter qu'elle a été moins rapide (ça fait déjà un point qui mériterait d'être vérifié). Parmi les personnes âgées, c'est le nombre de femmes qui augmente le plus vite. En effet, l'écart relatif se creuse entre hommes et femmes âgés au cours des années ou, vu sous un autre angle et de façon très approximative, le nombre de femmes âgées triple en 30 ans alors que le nombre d'hommes âgés ne fait que doubler.

Quelques calculs pour vérifier tout cela.

Taux de variation entre 1961 et 1971 : (419,3 – 306,3)/ 306,3 = 36,9 %

Taux de variation entre 1971 et 1981 : (573,9 – 419,3)/ 419,3 = 36,9 %

Taux de variation entre 1981 et 1991 : (781,2 – 573,9)/ 573,9 = 43,6 %

Entre 1991 et 1993, la population âgée est multipliée par 827,2/ 781,2 = 1,059 (soit un taux de variation de 5,9 %, car 1,059 – 1 = 0,059 = 5,9 %). Si la même tendance devait se maintenir pendant 10 ans (c'est-à-dire si la population est multipliée par 1,059 tous les deux ans cinq fois de suite), la population serait multipliée par $1{,}059^5$ = 1,332 au bout de 10 ans. En d'autres termes, le taux de variation entre 1991 et 2001 serait de 33,2 %, car 1,332 – 1 = 0,332 = 33,2 %. (Pour le rapport entre le taux de variation et l'indice de variation, revoir le chapitre 4.)

Taux de variation entre 1991 et 2001 (« si la tendance se maintient ») : 33,2 %.

Nous disions aussi que l'écart se creuse entre hommes et femmes âgés de 65 ans et plus. Cela peut être vérifié de deux façons. En 1961, les femmes représentent 52,8 % des personnes âgées (161,6/306,3) et en 1993 cette *proportion* est passée à 59,6 % (492,6/827,2). On peut également dire qu'il y avait 90 hommes pour 100 femmes en 1961 (144,7/161,6 × 100) et que ce *rapport* (ou *ratio de masculinité*) a baissé à 68 hommes pour 100 femmes en 1993 (334,6 / 492, 6 × 100).

Colonnes 4 à 6 : de moins en moins d'hommes

Plus l'âge avance et moins il y a d'hommes par rapport aux femmes. D'autre part, cet écart augmente avec le temps. En 1993, il n'y a plus, chez les gens âgés de 75 ans ou plus, que 54 hommes pour 100 femmes (contre 85 pour 100 en 1961).

Toutefois, la délimitation des classes d'âge peut perdre de sa pertinence sur une aussi longue période. L'espérance de vie a passablement augmenté entre 1961 et 1993, surtout pour les femmes. Qui sait si avoir 75 ans en 1961 n'« équivaut » pas à avoir 80 ans en 1993?

Colonnes 8 à 10 : les personnes à charge changent

Pour caricaturer un peu, disons qu'il y a de plus en plus de grands-parents par enfant (colonne 8) et qu'il y a de plus en plus de retraités par personne active (colonne 9). Par contre, le nombre de personnes à charge (enfants et retraités) pour chaque personne active commence à peine à augmenter à partir de 1991 et devrait être encore inférieur en 2036 à ce qu'il était en 1961. Cela dit, il est clair que la « charge » occasionnée par une personne âgée n'est pas la même que celle occasionnée par un enfant.

Colonne 7 : on ne rajeunit pas

Les personnes âgées occupent une part de plus en plus importante dans la population du Québec. Cela confirme l'intuition que nous avions après avoir examiné les colonnes 1 à 3, à savoir que la catégorie 65 ans et plus augmente plus vite que le reste de la population. D'ailleurs, en combinant les colonnes 7 et 3, nous sommes en mesure de calculer la population totale pour chaque année envisagée. De plus, en combinant les colonnes 8, 9 et 3, on peut reconstituer les effectifs des trois principales classes d'âge.

Population totale en 1993 :
les 827,2 (milliers) de personnes âgées (colonne 3) représentent 11,5 % de la population totale (colonne 7).

827,2/Population totale = 11,5 %

Population totale = 827,2/11,5 % = 827,2/0,115 = 7 193 (milliers d'habitants)

Vérification : 827,2 (gens âgés)/7 193 (total) = 0,115 = 11,5 %

Personnes de 14 ans ou moins (« jeunes ») en 1993 :
Il y a 58,3 personnes âgées pour 100 jeunes (ou 0,583 pour 1 si on préfère) (colonne 8), et 827,2 (milliers) de personnes âgées.

Personnes âgées/Jeunes = 0,583

Jeunes = Personnes âgées/0,583 = 827,2/0,583 = 1 419 (milliers)

Vérification : 827,2 (gens âgés)/1 419 (jeunes) = 0,583 = 58,3 %

Personnes de 15 à 64 ans (« actifs ») en 1993 :
Il y a deux façons de calculer ce nombre. On peut utiliser la même méthode que pour la classe d'âge précédente:

On a alors 827,2/0,167 = 4 953,3 (milliers) de personnes de 15 à 64 ans.

On peut aussi y aller par soustraction :

Actifs = Population totale – Jeunes – Gens âgés = 7193 – 1 419 – 827,2 = 4947

Étant donné que tous nos chiffres sont arrondis, l'écart de quelques milliers peut être considéré comme négligeable. Si par contre nous avions observé un écart important, nous aurions pu en déduire soit que nos calculs étaient faux, soit que les données du tableau étaient fausses (la première éventualité est généralement la bonne!).

2.3. Interpréter

Colonnes 1 à 3 : naissances et longévité

L'accroissement du nombre de personnes âgées entre 1961 et 1991 peut s'expliquer de deux façons :

1. La cohorte de gens nés entre 1896 et 1926 est relativement grande par rapport à la génération précédente ou à la génération actuelle. En passant, il est difficile de mettre cela sur le dos des *baby-boomers*.

2. L'espérance de vie augmente, puisqu'une proportion de plus en plus grande de personnes dépasse l'âge de 65 ans).

Il serait intéressant de vérifier ces hypothèses (en obtenant des chiffres) et d'évaluer comment ces deux facteurs influenceront la population à l'avenir.

Même si cela n'est pas indiqué dans le tableau, sachant que les années 1945 à 1960 ont été particulièrement fécondes au Québec, on devrait peut-être s'attendre à une forte augmentation de la classe d'âge des 65 ans et plus entre 2010 et 2025. Et comme par hasard, l'année fatidique 2015 tombe en plein milieu de cette *vague*. Si on estime, de plus, que l'espérance de vie continuera d'augmenter, la progression relative du groupe de personnes âgées devrait même de s'accélérer.

Colonnes 4 à 6 : un monde sans hommes

Eh oui, l'écart s'est creusé au cours des années entre hommes et femmes (quoiqu'il tende finalement à se stabiliser). La population très âgée est de plus en plus dominée par des femmes (et par des veuves). Cela aura sûrement des implications sur le plan matériel, culturel, psychologique, etc. Voilà un bon sujet d'étude.

Colonne 7 : des vieillards minoritaires

On a beau dire, la proportion de personnes âgées plafonnera à un niveau relativement faible. Étant donné que l'espérance de vie est d'environ 77 ans au Québec (en 1993) et qu'elle augmente très lentement, la population de 64 ans et moins sera toujours beaucoup plus nombreuse que celle de 65 ans et plus… à moins qu'on arrête carrément de faire des enfants, auquel cas le problème finira par être réglé définitivement.

Colonnes 8 à 10 : pas plus de dépendants qu'en 1961

Si on estime (en simplifiant beaucoup) que les personnes de moins de 15 ans et de plus de 64 ans sont à la charge des autres personnes, on s'aperçoit que le poids de cette charge a baissé depuis 1961 (colonne 10). Ce poids commence à peine à remonter dans les années 1990 et, en 2036, il sera encore inférieur au niveau de 1961. Au début de cette section, nous nous inquiétions du financement des futures retraites. La colonne 10 nous donne ici un bon éclairage. C'est finalement plus une question d'organisation que de fardeau réel. Il n'y a donc pas de quoi s'alarmer, pourvu que le problème soit traité honnêtement et intelligemment.

Il n'en demeure pas moins que les personnes à charge ne sont plus les mêmes : les personnes âgées remplacent les enfants. C'est ce qu'on remarque clairement dans la colonne 8. Pour chaque jeune, il y a de plus en plus de personnes âgées : 16 personnes âgées pour 100 jeunes en 1961 et 58 en 1993. Il y a aura bientôt plus de grands-parents que de petits enfants! Ces derniers seront particulièrement gâtés (et peut-être mal élevés, ce qui risque de mal les préparer à leurs responsabilités d'adultes ayant un nombre croissant de personnes à charge).

Par curiosité, nous avons tenté d'obtenir de l'information supplémentaire. Le tableau 6.7 montre que le Régime des rentes du Québec est justement devenu déficitaire en 1991. Il n'y a rien d'anormal à ce que les dépenses d'un tel régime dépassent ses revenus à certaines périodes. Après tout, il y a des années de vaches grasses (arrivée d'une grosse vague de gens en âge de travailler) et des années de vaches maigres (la vague atteint l'âge de la retraite). Par contre, si le déficit devait persister à plus long terme, il serait approprié de prendre des mesures immédiates.

Tableau 6.7 - Le budget du régime des rentes du Québec

	Revenus	Dépenses	Solde
		(en milliards de $)	
1971	0,391	0,051	0,340
1981	1,730	0,859	0,871
1991	3,604	3,566	0,038
1992	3,596	3,932	-0,336
1993	3,657	4,203	-0,546
1994	3,885	4,466	-0,581

Source : L'observateur économique canadien.

2.4. Et qu'en est-il, 20 ans plus tard?

Le texte qui précédait a été écrit en 1996. Le moment est venu de mettre nos données à jour (lignes en bleu dans le tableau 6.5b figurant ci-après). Nous constaterons d'emblée que nos prévisions de 1996 y sont largement confirmées. Il n'y a là rien d'étonnant compte tenu du sujet étudié (la démographie, phénomène hautement prévisible) et de la façon d'aborder ce sujet (les méthodes quantitatives). Nous sommes loin des scénarios catastrophiques évoqués par les politiciens de 1996.

Tableau 6.5b - Les personnes âgées au Québec jusque dans les années 2010

	Population de 65 ans et plus (en milliers)			Rapport de masculinité (en %)		
	Hommes	Femmes	**Total**	55-64 ans	65-74 ans	75 ans et plus
	[1]	[2]	[3]	[4]	[5]	[6]
1961	144,7	161,6	**306,3**	97	92	85
1971	183,5	235,8	**419,3**	94	82	70
1981	237,6	336,3	**573,9**	89	78	59
1991	316,1	465,1	**781,2**	92	78	55
2001	398,1	563,6	**961,6**	95	84	56
2011	548,8	707,3	**1256,1**	100	91	63

	Rapport de dépendance (en %)			
	65 ans et plus/ Population	65 ans et plus/ 0 à 14 ans	65 ans et plus/ 15-64 ans	[65 ans et plus et 0 à 14 ans]/ 15-64 ans
	[7]	[8]	[9]	[10]
1961	5,8	16,4	9,9	70,2
1971	6,8	23,3	10,7	56,5
1981	8,7	40,7	12,5	44,3
1991	11,0	55,7	16,0	44,6
2001	13,0	73,6	18,7	44,2
2011	15,7	101,8	22,8	45,1
2036	*22*	*129*	*36*	*64*

Source des données brutes : Le Québec statistique 1995; Statistique Canada, Cansim 051-0001.

Quelles sont les grandes lignes qui se dégagent du tableau 6.5b? Tout d'abord, le nombre de personnes âgées de 65 ans et plus a effectivement continué à progresser rapidement depuis 1991 (colonne 3). Cependant, il ne faut pas oublier que la population totale du Québec a également augmenté entre-temps. En définitive, le rapport de dépendance de ce groupe d'âge par rapport à l'ensemble de la population passe de 11 % en 1991 à 15,7 % en 2011, ce qui est substantiel, mais pas dramatique (colonne 7).

Quant au rapport « Personnes à charge/Personnes actives », qui correspond très approximativement à la colonne 10, il est resté particulièrement stable depuis 1981, et il se situe nettement en dessous de sa valeur de 1961, époque des familles nombreuses. Cependant, en 2011, le nombre de personnes âgées à charge a rejoint — et dépassé — celui des enfants à charge (observer l'écart entre la colonne 9 et la colonne 10).

Le ratio calculé à la colonne 8 confirme ce qui vient d'être dit, puisqu'il a franchi le cap des 100/100. Les enfants seront donc de plus en plus gâtés, comme nous l'avions d'ailleurs prédit. En gros, le ratio « grands-parents/petits-enfants » est passé de 1 pour 6 en 1961 à 1 pour 1 en 2011.

Comme nous l'avions entrevu en 1996, le taux d'accroissement du nombre de personnes âgées a effectivement ralenti durant les années 1990. Entre 1991 et 2001, ce taux est de (961,6 – 781,2)/781,2 = 23,1 % (contre 36,1 % pour la décennie précédente). Ce n'est qu'à l'approche de la décennie 2010 que ce taux se remet à grimper : l'effet *baby-boomers* commence seulement à se manifester. Il n'est en effet pas difficile de vérifier que 1945 + 65 = 2010!

Entre 1991 et 2011, un phénomène inattendu s'est produit au Québec, et dans de nombreux pays développés : l'espérance de vie des hommes a rattrapé une partie du terrain perdu par rapport à celle des femmes. En conséquence, après un long déclin, les rapports de masculinité se sont redressés pour les hommes de 55 ans et plus (colonnes 4 à 6). Les femmes d'un certain âge auront donc de meilleures chances que prévu de se trouver cavalier pour danser le tango (ou le triple swing).

EXERCICES 2

1. Qui fait les travaux à la maison?

Une enquête de Statistique Canada sur la famille et les amis indique comment les hommes et les femmes perçoivent leur participation aux travaux domestiques au Canada. L'enquête, qui est basée sur un échantillon de 13 000 personnes environ, révèle que la participation des hommes à la préparation des repas, à la vaisselle, au ménage et au lavage est la plus élevée chez les couples non mariés, chez les jeunes, et chez les gens les plus scolarisés.

a) Définir. D'après le texte qui précède, quelles sont la population et les variables étudiées? Parmi ces variables, lesquelles figurent dans le tableau 6.8 ci-dessous? Sur les 6411 hommes de l'échantillon interrogé, 996 prétendent faire la vaisselle tout seuls et 1115 avec leur conjointe. Vérifiez ces chiffres à partir du tableau 6.8.

Tableau 6.8 - Qui fait les travaux à la maison?

		Moi-même	Conjoint(e)	Partage égal	Autres et refus	Total
			(en %)			
Préparation des repas	Hommes	11,8	74,0	11,6	2,6	**100,0**
	Femmes	81,4	8,4	7,8	2,3	**100,0**
Vaisselle	Hommes	15,5	60,3	17,4	6,7	**100,0**
	Femmes	70,9	10,9	11,7	6,4	**100,0**
Ménage et lavage	Hommes	8,7	73,3	12,9	5,1	**100,0**
	Femmes	79,2	5,6	9,7	5,6	**100,0**
Entretien intérieur et extérieur	Hommes	75,1	5,4	3,4	16,1	**100,0**
	Femmes	9,2	68,5	5,7	16,6	**100,0**

Source : Statistique Canada, Enquête sociale générale : La famille et les amis, 1994.

b) Observer. Rédigez, sous forme de phrases, les 3 ou 4 faits saillants qui se dégagent du tableau.

c) Analyser. Quelles hypothèses pouvez-vous faire à la lecture du tableau 6.8? Distinguez ce qui est démontrable de ce qui est spéculatif.

d) Supplément d'information. En quoi le tableau 6.9, figurant ci-après, présente-t-il un intérêt pour la question étudiée? Commentez-le.

Tableau 6.9 - Êtes-vous satisfait(e) du partage des travaux ménagers?

	Hommes	**Femmes**
	(en %)	
Satisfait(e)	89,1	81,8
Insatisfait(e)	4,2	10,6
Sans opinion ou non déclarés	6,7	7,6

Source : Statistique Canada, Enquête sociale générale : La famille et les amis, 1994.

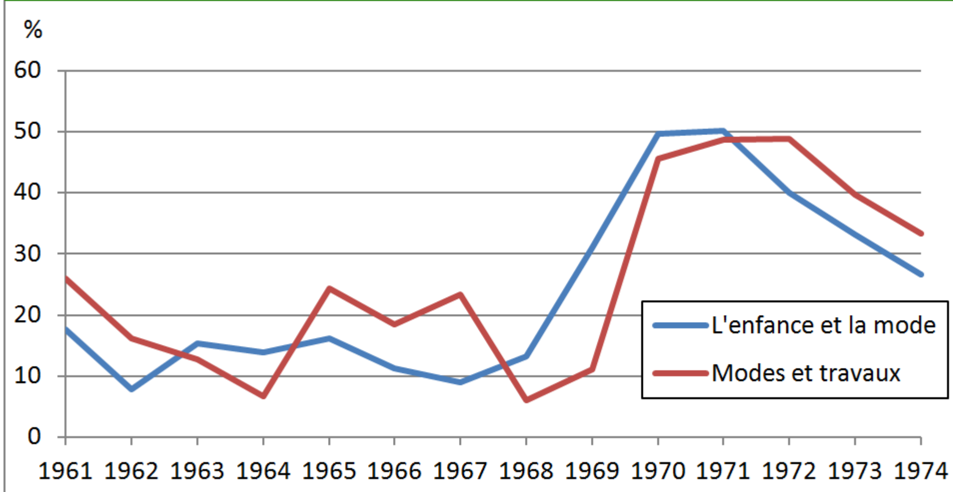
3. ANALYSER UN GRAPHIQUE : LES FILLES EN PANTALON

Les années 1966 à 1968 ont marqué un tournant dans l'évolution des mentalités. Les étudiants envahissaient les rues de Paris, les hippies californiens se réfugiaient à la campagne et les G.I. désertaient au Canada, les seins nus apparaissaient sur les plages de la Côte d'Azur, les jeunes Chinois coiffaient leurs professeurs du bonnet d'âne, les parents, nourris de Luis Mariano, Bing Crosby ou Fernand Gignac, ne comprenaient rien à la musique de Pink Floyd. Et nous pourrions en rajouter… Ah oui, autre chose : les filles se mirent à porter le pantalon.

3.1. Un phénomène de société

Nous allons étudier ici le port du pantalon (c'est notre variable) des jeunes Françaises (c'est notre population) et l'évolution de cette variable autour de 1966-1968 (les années charnières). La variable est observée, de façon indirecte, à travers l'image fournie par deux revues de mode parisiennes : *L'enfance et la mode* et *Modes et travaux*. La variable *port du pantalon* est mesurée selon une échelle nominale comprenant deux catégories : absence ou présence du pantalon dans l'habillement. Les données représentent des proportions (filles en pantalon par rapport à l'ensemble de filles) et sont exprimées en pourcentage. La figure 6.4 illustre le résultat de l'enquête.

Figure 6.4 - Les filles adoptent le pantalon



Source : Psychologie de la mode, Marc-Alain Descamps, PUF.

Avant 1969, le port du pantalon par les filles reste marginal. À partir de 1969 (après le tournant culturel dont nous parlions plus haut), le pantalon devient un habit féminin à part entière, qui peut se porter aussi bien pendant les loisirs qu'au travail. On remarque également, en dehors de ce grand virage, des fluctuations constantes : la mode peut varier d'une année à l'autre. Les deux phénomènes (long terme et court terme) se superposent.

On observe une très forte *corrélation* entre les deux courbes (en d'autres mots, les courbes se suivent de très près). Cela renforce l'idée que l'image projetée par les revues colle pas mal à la réalité. Par ailleurs, on constate qu'une des courbes précède l'autre de façon presque systématique. *L'enfance et la mode* a une longueur d'avance sur *Modes et travaux*. Cela peut s'expliquer par le fait que la première revue s'adresse aux gens du métier (qui sont à l'avant-garde) alors que la seconde est destinée au grand public.

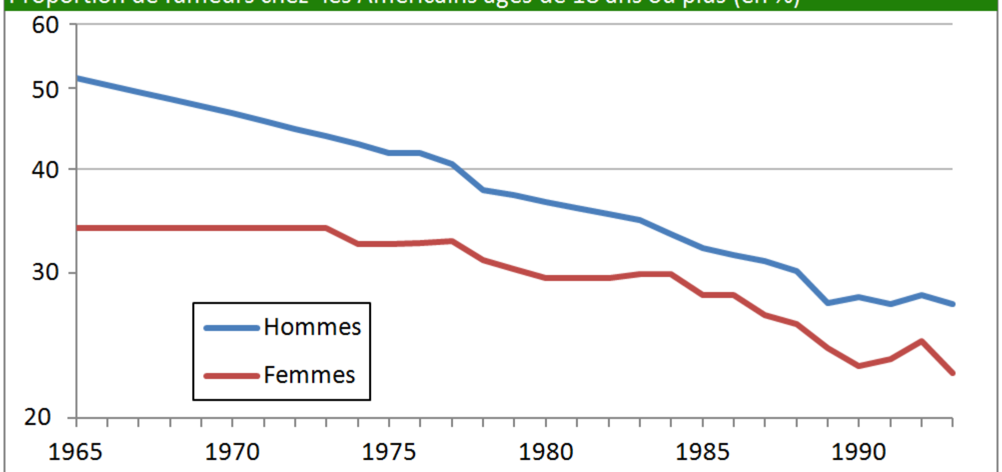
Si les revues reflètent le changement des mœurs, elles ont elles-mêmes une influence sur la manière de s'habiller. Dans ce sens, il y a une influence mutuelle entre le phénomène proprement dit (la façon dont les filles veulent s'habiller) et l'image projetée dans les revues.

EXERCICES 3

1. La cigarette en perte de vitesse

Dans les années 1960 et 1970, les amphis ressemblaient parfois à des tripots clandestins tant la fumée des cigarettes y était épaisse les jours d'examens. Une bonne moitié des hommes et un tiers des femmes s'adonnaient alors au « vice » du tabagisme. Mais le destin de la cigarette était déjà scellé, et son usage avait déjà amorcé son inexorable déclin, suite aux multiples campagnes de santé publique. En 2011, l'enquête annuelle du CDC (organisme du gouvernement américain) évaluait la proportion de fumeurs à seulement 19 % de la population adulte.

Figure 6.5 - Le tabagisme aux États-Unis
Proportion de fumeurs chez les Américains âgés de 18 ans ou plus (en %)



Source : Health United States 1995.

a) Quelle est la population étudiée? Quelles sont les deux caractéristiques de cette population (ou variables) qui nous intéressent ici?

b) Expliquez le choix de l'échelle sur l'axe vertical.

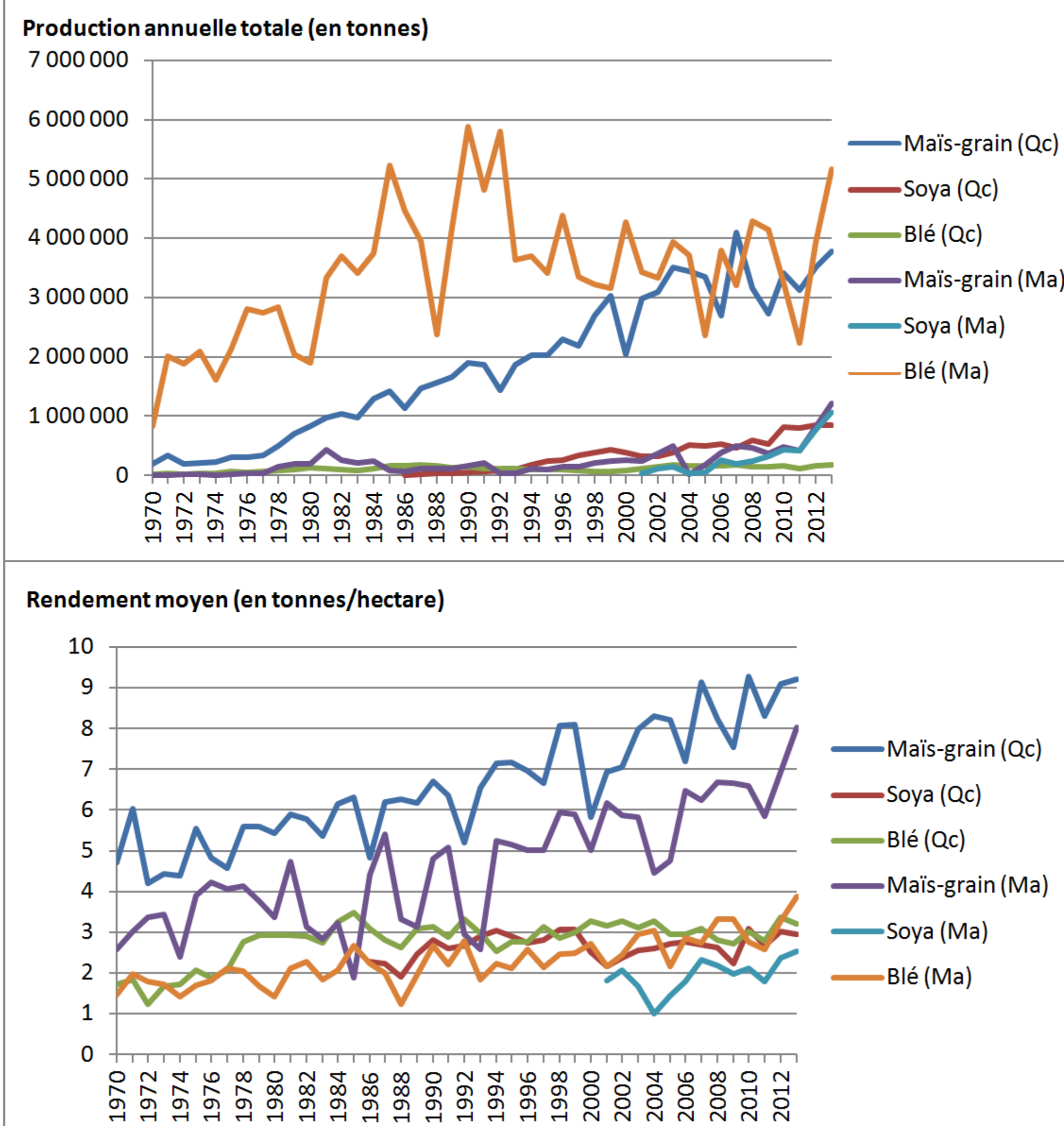
c) Commentez les courbes.

2. Un parfum d'OGM

La figure 6.6 montre l'évolution de la production et du rendement de trois grandes cultures au Québec et au Manitoba. Vous vous baserez sur la grille d'analyse présentée au début de ce chapitre pour répondre aux questions suivantes.

Complément d'information : Au premier semestre 2014, les prix mondiaux moyens de la tonne de maïs, de blé et de soya étaient respectivement de 211 $, 348 $ et 549 $ (source : FAO).

Figure 6.6 - Évolution de la production agricole au Québec et au Manitoba



Source : Statistique Canada, Cansim 001-0010.

a) Définissez. Énoncez les éléments qui ont servi à la construction des figures. Expliquez notamment la relation entre les deux séries de courbes (production et rendement).

b) Observez.

c) Interprétez.

4. CONSTRUIRE SES PROPRES TABLEAUX

De nos jours, il est facile d'accéder à un nombre considérable de bases de données statistiques gérées par des organismes publics ou internationaux tels que la Banque mondiale, l'ONU, Eurostat, Statistique Canada, l'INSEE, etc.

Le tableau brut est la mine ou on puise l'information chiffrée.

Il faut reconnaître que bon nombre de tableaux publiés par les organismes statistiques s'avèrent fort complexes : ils contiennent beaucoup de variables (par exemple, le partage des travaux ménagers selon le sexe, le niveau de scolarité, l'état matrimonial, l'âge et la langue), elles-mêmes divisées en un grand nombre de catégories (les dizaines de langues parlées dans chacune des 98 divisions de recensement du Québec). Certaines variables sont de plus affublées de noms à coucher dehors (par exemple, « taux de croissance mensuel de la consommation en dollars enchaînés de 2011 aux pondérations de 2002 convertis sur une base annuelle »). Ces tableaux « bruts » ont cependant un double mérite : ils sont souvent exhaustifs (tout est couvert) et les définitions sont précises (même si elles sont parfois longues). Ils sont la matière première, la mine d'or dans laquelle le chercheur peut puiser son information.

Les chiffres intéressants peuvent être extraits du tableau brut et présentés dans un plus petit tableau, clair et agréable à consulter.

Les chiffres extraits d'un tableau brut peuvent être présentés dans un nouveau tableau, facile à saisir, ou à l'intérieur d'un texte. Dans le premier cas, seule l'information essentielle (en fonction de la recherche) est retenue. Le tableau brut que nous avons utilisé au sujet des nationalités dans l'ex-Yougoslavie comptait 8 pages (une pour chaque république et région autonome) et portait sur 12 nationalités différentes (qui pouvaient varier d'une région à l'autre). Étant donné que nous nous intéressions plus particulièrement au conflit bosniaque, nous avons réduit le tableau original à quatre nationalités ou groupes de nationalités et à trois républiques.

Quant aux tableaux ou aux graphiques tout faits destinés au grand public, ils possèdent un triple inconvénient pour le chercheur. Les données y sont souvent déjà transformées (dans les tableaux) ou imprécises (dans les graphiques). De plus, ces données transformées sont plus difficiles à traiter que les données brutes. Enfin, utiliser ces données déjà transformées serait faire preuve de peu d'originalité puisqu'elles sont le fruit du travail d'un autre chercheur. Mieux vaut mettre la main sur les données brutes, que l'on peut ensuite transformer et enrichir selon ses besoins.

Pour construire un tableau à partir d'une base de données, on sélectionne trois dimensions : les *variables*, *l'époque* et la *zone géographique* couvertes.

Si la lecture de tableaux déjà construits peut servir de source d'inspiration, la meilleure façon de procéder consiste encore à construire ses propres tableaux. Dans une première étape, le chercheur sélectionne les variables pertinentes ainsi que l'époque et la zone géographique qui l'intéressent. Dans une seconde étape, il « fait parler » les données en transformant les données brutes en rapports, moyennes, taux de variation, etc. C'est ce que nous verrons dans cette section à travers quelques exemples reliés à la criminologie.

4.1. Les gendarmes et les voleurs

Le jeu des gendarmes et des voleurs, autrefois très populaire dans les camps de vacances et les mouvements de jeunesse, est sans doute appelé à disparaître, dans un monde où les relations entre les individus se « dématérialisent » de plus en plus. L'équipement nécessaire au jeu des gendarmes et des voleurs est on ne peut plus simple sur le plan technologique : il suffit d'avoir sous la main une balle et un groupe d'enfants, que l'on divise en deux camps placés face à face. Les gendarmes doivent capturer les voleurs, en les frappant avec la balle, tandis que les voleurs cherchent à intercepter cette balle au vol et à la faire parvenir, par-dessus le camp des gendarmes, jusqu'aux voleurs déjà en prison. Ces derniers peuvent alors s'évader s'ils réussissent à faire rebondir la balle sur un gendarme. Ce jeu passionnant, qui pourrait durer une éternité, s'arrête le plus souvent quand un des joueurs portant des lunettes, qu'il soit gendarme ou voleur, reçoit la balle en pleine figure.

Où l'on voit que les règles de l'exclusivité et de l'exhaustivité sont omniprésentes dans les statistiques reliées aux sciences humaines.

Au Québec, un gendarme est une personne « responsable d'effectuer la surveillance du territoire, de répondre aux appels des citoyens, de procéder aux enquêtes de base, d'assurer le transport des prévenus, d'effectuer des activités de prévention ainsi que d'appliquer les lois et les règlements ». En 2012, 63 % des policiers permanents du Québec travaillaient à la gendarmerie, tandis que 17 % d'entre eux s'occupaient des enquêtes, et 13 % étaient affectés au soutien opérationnel (détention des suspects, identité judiciaire, les pièces à conviction, renseignements criminels, liaison judiciaire, etc.). Les 7 % restants se répartissaient entre les relations communautaires et prévention (seule catégorie où les femmes étaient majoritaires), la direction et l'administration.

Source de la définition du mot *gendarme*, ainsi que des chiffres cités dans le paragraphe : Ministère de la Sécurité publique du Québec, Profil organisationnel 2012.

Les policiers du Québec ont — à tort ou à raison — la réputation d'être grassement payés. Or, au moment de la grande réforme des fusions municipales au Québec, la question des salaires versés aux policiers provenant de municipalités différentes et regroupés au sein d'un même corps n'a pas manqué de se poser. Fallait-il aligner les salaires sur le niveau le plus élevé, sur le niveau le plus bas ou sur un niveau intermédiaire? Il ne s'agit bien sûr que d'une des multiples questions soulevées par la réforme. Pour y voir plus clair, nous avons décidé de construire un petit tableau contenant quelques données de base, propres à permettre une amorce de réflexion. Nous avons identifié six variables clés et nous avons choisi six villes du Québec (dimension *zones géographiques*), peu avant les fusions (dimension *époque*). Les résultats sont rassemblés dans le tableau 6.10, que nous vous demandons d'examiner attentivement. Nous vous demandons également de réfléchir sur le choix des six variables.

Tableau 6.10 - Données relatives aux corps de police municipaux du Québec, 1998

	Population desservie (1999-01-01)	Policiers permanents	Budget municipal (milliards $)	Richesse foncière (milliards $)	Coût du corps de police (millions $)	Rémunération des policiers (millions $)
Aylmer	34 901	45	37,5	1,515	4,789	3,459
Gatineau	120 369	153	130,1	4,670	15,301	11,062
Hull	62 339	128	110,1	3,979	12,248	9,493
Montréal CUM	1 775 846	4157	3081,0	88,700	376,369	271,656
Québec	270 651	444	445,6	11,426	40,267	32,467
Rouyn-Noranda	29 797	41	33,5	1,232	3,991	3,122
Total Québec	**5 665 839**	**9031**	**7506,7**	**248,447**	**870,1**	

Source : Ministère de la Sécurité publique du Québec, Rapport sur les forces policières municipales 2001-2002.

À première vue, le tableau 6.10 ne paraît pas fournir d'information extraordinaire. On y voit bien que Montréal est une grosse métropole, avec un gros budget, une grosse population et beaucoup de policiers, mais cela n'a rien d'original. Il suffirait pourtant de quelques calculs élémentaires pour que le tableau « se mette à table ». Pour défricher le terrain, nous avons donc calculé les proportions en colonnes (tableau 6.10b). Notons, avant d'aller plus loin, un petit détail important : la population de Québécois qui nous intéresse ici est celle qui est desservie par un corps de police municipal, soit 5,666 millions de personnes sur un total de 7,310 millions au 1er janvier 1999.

Tableau 6.10b - Données dérivées : Proportions par rapport à l'ensemble du Québec

	Population desservie	Policiers permanents	Budget municipal (en %)	Richesse foncière	Coût du corps de police
Aylmer	0,6	0,5	0,5	0,6	0,6
Gatineau	2,1	1,7	1,7	1,9	1,8
Hull	1,1	1,4	1,5	1,6	1,4
Montréal CUM	31,3	46,0	41,0	35,7	43,3
Québec	4,8	4,9	5,9	4,6	4,6
Rouyn-Noranda	0,5	0,5	0,4	0,5	0,5
Total Québec	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**

Le tableau 6.10b devrait se lire de la façon suivante : quelle est la proportion occupée par chaque municipalité par rapport à l'ensemble du Québec, et ce pour chacune des variables? Nous voyons d'emblée que 46,0 % de tous les policiers du Québec œuvrent à Montréal, ville dans laquelle n'habitent que 31,3 % des Québécois desservis par une police municipale. À première vue, on pourrait affirmer que Montréal compte trop de policiers. Certains pourraient même conclure hâtivement que les policiers, comme les médecins, répugnent à exercer en province. Mais n'allons pas si vite! Les policiers se trouvent principalement là où se trouvent les criminels, ou ceux jugés comme tels. Or, il est clair que les grandes villes, et les centres de ces grandes villes, constituent généralement un terrain plus intéressant pour les criminels que les banlieues-dortoirs et les villages perdus au fond de la campagne. Sans compter que, si Montréal ne compte que 31,3 % des Québécois la nuit, il en va autrement pendant la journée, quand les banlieues-dortoirs, justement, se déversent sur la métropole. En fin de compte, il ne semble pas anormal de trouver tant de policiers à Montréal. L'ancienne ville de Hull, flanquée de ses deux banlieues, Gatineau et Aylmer, se trouve dans une position similaire, à une échelle réduite. Hull possédait d'ailleurs autrefois le surnom de « Petit Chicago ».

Le déséquilibre fiscal entre villes centrales et banlieues est un phénomène bien documenté. Étant donné qu'une partie importante des recettes municipales provient des impôts fonciers (bâtiments et terrains), il est utile de comparer la richesse foncière de Montréal (35,7 % du total québécois) à ses effectifs policiers (46,0 % du total). En d'autres termes, on pourrait considérer que le fardeau (46,0 %) de la métropole est plus considérable que ses capacités (35,7 %). Il n'est alors pas étonnant que le chiffre du budget municipal soit situé entre ces deux valeurs.

Examinons maintenant le coût du corps de police de Montréal (43,3 %) et comparons-le à l'ensemble du budget (41 %), dont ce coût fait évidemment partie : Montréal consacre une plus grande proportion de son budget à sa police que la moyenne du Québec. Si nous comparons maintenant le coût du corps de police (43,3 %) à la proportion de policiers (46 %), nous pouvons en déduire que Montréal dépense moins par tête de policier que le reste du Québec. Il est possible que les policiers de Montréal soient moins bien équipés que leurs confrères, mais il est surtout probable qu'ils sont moins bien payés.

Pour en avoir le cœur net, nous avons calculé quelques rapports, c'est-à-dire que nous avons divisé certaines variables par d'autres variables, opération à la portée du premier écolier venu, mais immensément féconde (tableau 6.10c).

Tableau 6.10c - Données dérivées : Rapports entre variables

	Policier / 1000 Habitants	**Rémunération / Policier** (en $)	**Coût du corps / Habitant** (en $)	**Coût du corps / Budget municipal** (en %)	**Coût du corps / 100 $ Richesse foncière** (en $)
Aylmer	1,3	76 867	137,2	12,8	0,3
Gatineau	1,3	72 301	127,1	11,8	0,3
Hull	2,1	74 164	196,5	11,1	0,3
Montréal CUM	2,3	65 349	211,9	12,2	0,4
Québec	1,6	73 124	148,8	9,0	0,4
Rouyn-Noranda	1,4	76 146	133,9	11,9	0,3
Total Québec	**1,6**		**153,6**	**11,6**	**0,4**

Une première constatation saute aux yeux, c'est l'importance du salaire versé à un policier en 1998 (et, probablement, de nos jours). Cette année-là, alors que la rémunération moyenne d'un policier de la ville de Québec tournait autour de 73 000 $, le revenu moyen d'un Québécois atteignait à peine les 28 000 $, et le revenu médian s'établissait seulement à quelque 20 100 $ (source : Statistique Canada, Cansim 202-0402; chiffres convertis en dollars courants). Les policiers des quatre coins du monde, qui sont généralement moins scolarisés et moins bien payés que la moyenne de la population qu'ils desservent, demeureraient surpris, voire incrédules, en prenant connaissance des salaires obtenus par leurs confrères québécois.

Le ratio *Policier pour 1000 Habitants* est également révélateur. Ce ratio est bien plus élevé dans les grandes villes (Montréal) et dans les centres-villes (le « Petit Chicago ») qu'en province (Rouyn-Noranda) ou en banlieue (Aylmer, Gatineau). La ville de Québec se trouve dans une position intermédiaire.

Le travail n'est pas encore terminé. Avant de conclure que les policiers sont trop nombreux à Montréal et Hull, par exemple, il faut se demander ce qui détermine la présence de policiers sur un territoire. Un des éléments à ne pas négliger est certainement le nombre de crimes commis. Le site du Ministère de la Sécurité publique du Québec contient justement d'autres données sur le sujet. On y constate que Montréal a connu 20 004 crimes avec violence en 1998, ce qui équivaut à 11,26 crimes pour 1000 habitants. C'est plus que dans l'Outaouais (7,35) et dans la ville de Québec (4,69).

Il y aurait bien d'autres commentaires à faire sur les tableaux dérivés. Il a suffi de quelques calculs très simples et surtout d'un peu de méthode pour découvrir la richesse de l'information disponible.

4.2. La base de données sur le crime aux États-Unis

Les chiffres qui nous ont permis de créer le tableau précédent provenaient d'un rapport riche en données brutes. Cette fois-ci, nous irons chercher directement les chiffres dans une base de données.

Les statistiques sur la criminalité, comme bien des statistiques reliées aux sciences humaines, sont parfois difficiles à analyser à cause de disparités dans la définition des variables. Avant de comparer les données sur les vols de véhicules, par exemple, que ce soit dans le temps et dans l'espace, il faut s'assurer que tout le monde s'entend, et s'entendait au point de départ, sur ce qu'est un véhicule. Simple, direz-vous, et pourtant… Doit-on considérer un bateau à moteur comme un véhicule? Un wagon de chemin de fer? Bon, admettons que la définition mérite d'être clarifiée, mais en ce qui concerne les homicides, la situation est nette et tranchée, n'est-ce pas?

Pourtant, seriez-vous capable de répondre sans hésiter aux questions suivantes? L'exécution d'un condamné à mort constitue-t-elle un homicide? Le policier qui descend un criminel pendant un holdup commet-il un homicide? Et qu'en est-il d'un commerçant qui tue, en état de légitime défense, le cambrioleur qui s'est introduit dans son magasin? Et le chauffard qui écrase des écoliers, pendant qu'il consulte sa page *Facebook* sur son téléphone? Et les victimes des attentats du 11 septembre 2001? Ont-elles été victimes d'un homicide? Le ministère de la Justice des États-Unis, par exemple, ne classera aucun des cas cités dans les homicides. Il est donc important de bien cerner les définitions des variables avant de les interpréter. Il faut aussi, avant de se livrer à des comparaisons, tenir compte du fait que certaines définitions peuvent changer d'une époque à l'autre ou d'un organisme à l'autre. D'où l'intérêt de mettre sur pied un système uniformisé, tel que fichier DUC au Canada et le fichier *Uniform Crime Statistics* aux États-Unis.

Le tableau 6.11 provient de ce dernier fichier. Nous avons commencé par sélectionner les variables sous forme de données brutes, puis nous avons choisi quelques États aux quatre coins de l'Union (afin de construire la figure qui accompagne ce tableau). Dans un premier temps, nous nous sommes contentés de résumer la situation dans ses grandes lignes (première partie du tableau 6.11). Comme notre étude couvre un demi-siècle, et que la population des États-Unis s'est accrue entre-temps, nous avons calculé des taux de crimes par habitant, une opération qui ne prend que quelques secondes avec un chiffrier (seconde partie du tableau 6.11).

Tableau 6.11 - Les statistiques uniformisées du crime aux États-Unis

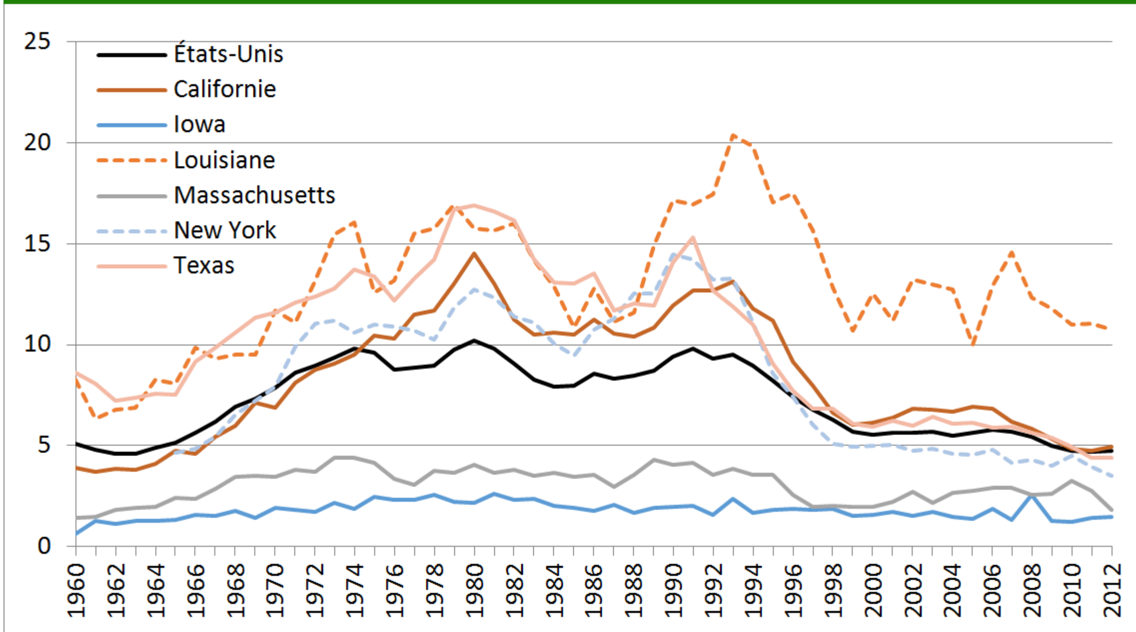
Nombre (en milliers)	**1960**	**1970**	**1980**	**1990**	**2000**	**2010**	**2012**
Population	179 323	203 235	225 349	249 464	281 422	309 330	313 914
Homicides	9,1	16,0	23,0	23,4	15,6	14,7	14,8
Viols	17,2	38,0	83,0	102,6	90,2	85,6	84,4
Vols avec violence	107,8	349,9	565,8	639,3	408,0	369,1	354,5
Voies de fait graves	154,3	335,0	672,7	1 054,9	911,7	781,8	760,7
Cambriolages	912,1	2 205,0	3 795,2	3 073,9	2 051,0	2 168,5	2 103,8
Vols simples	1 855,4	4 225,8	7 136,9	7 945,7	6 971,6	6 204,6	6 150,6
Vols de véhicules	328,2	928,4	1 131,7	1 635,9	1 160,0	739,6	721,1
Taux pour 100 000 habitants							
Homicides	5,1	7,9	10,2	9,4	5,5	4,8	4,7
Viols	9,6	18,7	36,8	41,1	32,0	27,7	26,9
Vols avec violence	60,1	172,1	251,1	256,3	145,0	119,3	112,9
Voies de fait graves	86,1	164,8	298,5	422,9	324,0	252,8	242,3
Cambriolages	508,6	1 084,9	1 684,1	1 232,2	728,8	701,0	670,2
Vols simples	1 034,7	2 079,3	3 167,0	3 185,1	2 477,3	2 005,8	1 959,3
Vols de véhicules	183,0	456,8	502,2	655,8	412,2	239,1	229,7

Source : Ministère de la Justice des États-Unis, Uniform Crime reporting statistics.

Les infractions inscrites dans le tableau 6.11 sont regroupées en deux catégories : les crimes avec violence et les crimes contre la propriété. Comme on peut le constater d'emblée, les années 1970 à 1990 ont été marquées par une vague de la criminalité, qui semblait alors « inarrêtable ». Et pourtant, toutes les catégories de crimes se sont mises à reculer par la suite, parfois dans des proportions substantielles.

Le fichier uniformisé américain contient également des données sur les 50 États et sur des milliers de villes. Afin de nous livrer à une première étude de l'évolution de la criminalité dans le temps et dans l'espace, nous avons choisi la variable *Homicides* et nous avons retenu six États, mais les possibilités d'étude étaient presque infinies (figure 6.7)

Figure 6.7 - Taux d'homicides pour 100 000 habitants aux États-Unis



Source : Ministère de la Justice des États-Unis, *Uniform Crime reporting statistics*.

La moyenne nationale est représentée par la courbe en noir sur la figure. On y voit que le taux d'homicides progresse rapidement à partir de 1964 (ère du rock), pour atteindre un sommet en 1974 (ère du disco). La situation est plus ou moins stable jusqu'au début des années 1992, où s'amorce la décrue.

On ne sera pas étonné de trouver l'Iowa, et même le Massachusetts, bien en dessous de la moyenne. Par contre, la Californie, terre des hippies, dépasse largement la moyenne américaine dans les années 1975 à 1997. L'État de New York, réputé pour ses violences à la même époque, se classe aujourd'hui au-dessous de la moyenne.

Ce n'est qu'un début, une première prise de contact. Il nous resterait à continuer notre recherche en fouillant dans cette mine d'informations que constitue une telle base de données.

4.3. Les homicides à New York

Pour comprendre la criminalité, il est important d'examiner d'autres variables que les crimes eux-mêmes. Qui sont les criminels? Qui sont les victimes? Sont-ils jeunes ou vieux, hommes ou femmes, Blancs ou Noirs, récidivistes ou néophytes? Le tableau 6.12, extrait d'un rapport du Service de police de la ville de New York, fournit justement ce type d'éclairage. Nous vous demandons, avant de poursuivre votre lecture, de bien décortiquer ce tableau en utilisant la grille de lecture habituelle, comme s'il s'agissait d'une question d'examen.

Tableau 6.12 - Homicides à New York en 2005

(en %)

Selon la race	Victimes	Meurtriers	Population	Policiers
Noirs	59	57	25	15
Hispaniques	27	32	28	22
Blancs	8	7	35	60
Asiatiques	5	4	11	3
Selon le sexe	**Victimes**	**Meurtriers**	**Population**	**Policiers**
Hommes	82	93	47	85
Femmes	18	7	53	15
Selon l'âge	**Victimes**	**Meurtriers**	**Population**	
Moins de 18 ans	8	10	32	
18 à 24 ans	28	37	11	
25 à 40 ans	40	38	24	
41 à 59 ans	19	12	21	
60 ans et plus	4	1	12	
Selon le casier judiciaire	**Victimes**	**Meurtriers**		
Sans casier	45	3		
Avec casier	55	97		

Source : Service de police de la ville de New York, décembre 2005, NYC-CCRB, et Bureau du recensement des États-Unis.

Note : Données portant sur les 508 homicides enregistrées dans l'année jusqu'au 18 décembre 2005 (proportions de l'année 2000 pour la population).

Voici maintenant un exemple des points que l'on pourrait relever en examinant ce tableau.

A. Définir

- Variables principales : Homicides (Meurtrier, Victime) × (Race, Sexe, Âge, Casier)
- Autres variables (pour référence) : Population, Policiers
- Proportions selon la deuxième série de variables (Race, Sexe, Âge, Casier)
- Données d'une grande métropole

B. Observer

- Les meurtriers sont surtout des Noirs et des Hispaniques, des hommes, des gens relativement jeunes (18 à 40 ans), des personnes avec casier judiciaire (comparer aux proportions respectives dans la population).
- Les victimes sont aussi surreprésentées selon les mêmes variables, sauf pour le casier judiciaire (comparer aux proportions dans la population). On remarque quand même qu'une bonne moitié des victimes a aussi un casier judiciaire.
- Les policiers sont surtout des hommes et des Blancs (toutes proportions gardées).
- La plupart des meurtriers sont des Noirs, mais la plupart des Noirs ne sont pas des meurtriers (loin de là!). Idem pour les hommes, etc. Il ne faut pas oublier que les meurtriers et les victimes constituent une infime minorité de la population (il y avait environ 8 millions d'habitants dans la ville de New York en 2005).
- Si les femmes commettent peu d'homicides, elles sont aussi moins souvent assassinées que les hommes.

C. Interpréter

- Les meurtriers et les victimes appartiennent souvent au même milieu (Race, Sexe, Âge, Casier). Pour caricaturer, on pourrait dire, par exemple, que les voyous se tuent surtout entre eux, etc.
- En ce qui concerne les effectifs policiers, on assiste peut-être à une diversification progressive. On embauche sans doute une bonne proportion de femmes et de non-Blancs, mais comme la carrière d'un policier dure plusieurs décennies, cela ne se reflète pas immédiatement dans les proportions (effet de génération).

EXERCICES 4

1. D'où provient la violence?

Le texte qui suit est adapté du rapport annuel du Département de la santé et des services humains des États-Unis (*Health United States 1995*) et traite du taux de crimes violents perpétré sur des personnes âgées de 12 ans et plus.

En 1992-93, les femmes avaient 6,6 fois plus de chances d'être victimes d'un crime violent perpétré par leur conjoint (ou ex-conjoint) que les hommes (9,3 crimes pour mille femmes de 12 ans et plus contre 1,4 pour mille). On compte chaque année, en moyenne, 1 008 000 crimes de ce genre.

Les femmes ont à peu près autant de chances de subir des crimes violents perpétrés par leurs conjoints ou d'autres parents (37 %) que par des amis ou connaissances (40 %), et beaucoup moins de chances d'être victimes d'inconnus. Chez les hommes, par contre, la situation est très différente. En effet ,les crimes violents perpétrés contre des hommes sont susceptibles de provenir de connaissances (44 %) ou d'étrangers (49 %) plutôt que de parents (7 %).

a) Complétez le tableau suivant :

Tableau 6.13 - Taux de crimes violents perpétrés sur des personnes âgées de 12 ans et plus

(Taux pour 1000 individus selon le sexe de la victime et l'origine du criminel)

	Conjoint	Autre parent	Connaissance	Inconnu	Total
Femmes	9,3		37,5	7,4	
Hommes		1,2			**38,8**

Source : Ministère de la Santé des États-Unis.

b)

Taux de crimes violents perpétré sur des personnes âgées de 12 ans et plus.

b) Quelles sont les deux variables que l'on retrouve dans le tableau ci-dessus?

c) Quel est le nombre de femmes dans la population étudiée?

d) Commentez les affirmations suivantes :

- La plupart des hommes commettent un crime violent contre leur conjointe (ou ex-conjointe).
- Les femmes devraient se méfier de leur conjoint, les hommes devraient se méfier des inconnus et tout le monde devrait se méfier de ses amis.
- Les femmes subissent plus de crimes violents que les hommes.

2. La récidive en Angleterre

Commentez le tableau 6.14 en vous servant de la grille d'analyse présentée au début de ce chapitre.

Tableau 6.14 - Récidive criminelle en Angleterre et au Pays de Galles - 2002-2004

Après un an

	Femmes			Hommes		
	Récidivistes	Non récidivistes	**Total**	Récidivistes	Non récidivistes	**Total**
18-20 ans	394	432	**826**	4 013	2 906	**6 919**
21-24 ans	583	566	**1 149**	4 506	3 671	**8 177**
25-34 ans	981	1 106	**2 087**	6 687	6 555	**13 242**
35 ans et plus	427	1 060	**1 487**	3 081	6 116	**9 197**
Total	**2 385**	**3 164**	**5 549**	**18 287**	**19 248**	**37 535**

Après deux ans

	Femmes			Hommes		
	Récidivistes	Non récidivistes	**Total**	Récidivistes	Non récidivistes	**Total**
18-20 ans	471	355	**826**	4 809	2 110	**6 919**
21-24 ans	678	471	**1 149**	5 405	2 772	**8 177**
25-34 ans	1 160	927	**2 087**	8 197	5 045	**13 242**
35 ans et plus	520	967	**1 487**	3 973	5 224	**9 197**
Total	**2 829**	**2 720**	**5 549**	**22 384**	**15 151**	**37 535**

Source : Ministère de l'Intérieur du Royaume-Uni (Home Office), 2005.

EXERCICES SUPPLÉMENTAIRES

1. Qui travaille le plus : les hommes ou les femmes?

Première partie : Tableau 6.15

a) Quelle est la population étudiée? Quelles sont les variables et quelles sont leur échelle de mesure? Les données sont-elles brutes ou dérivées? Quelle est leur unité de mesure? Qu'avez-vous à dire sur les années choisies?

b) Calculez, pour chaque pays étudié, l'indice du temps de travail pour les hommes et les femmes en prenant pour base 100 la moyenne des 14 pays industrialisés.

c) Vérifiez les chiffres de la colonne 4.

d) Commentez le tableau 6.15.

Tableau 6.15 - La charge de travail par sexe dans certains pays

		Temps de travail			Charge de travail des femmes comparée à celle des hommes
		Femmes	Hommes	Moyenne	
		(en minutes par jour)			(différence en %)
		[1]	[2]	[3]	[4]
Pays en développement : milieu urbain					
Colombie	1983	399	356	378	12,0
Indonésie	1992	398	366	383	9,0
Kenya	1986	590	572	581	3,0
Pays en développement : milieu rural					
Bangladesh	1990	545	496	521	10,0
Kenya	1988	676	500	588	35,0
Philippines	1975-77	546	452	499	21,0
Pays industrialisés					
Canada	1992	429	430	430	-0,2
France	1985-86	429	388	409	10,6
Italie	1988-89	470	367	419	28,1
Moyenne de 9 pays en développement		544	483	514	13,0
Moyenne de 14 pays industrialisés		430	428	441	5,8

Source : PNUD, Rapport mondial sur le développement humain 1995.

Note : Les moyennes figurant au bas du tableau portent sur les pays sélectionnés pour l'étude. Le choix de l'échantillon s'est fait en fonction de la disponibilité et de la fiabilité des sources. Il vaut mieux avoir un échantillon moins grand et moins représentatif que des données de base moins fiables.

Deuxième partie : Tableau 6.16

e) Quelle est la population étudiée? Quelles sont les variables et quelles sont leur échelle de mesure? Les données sont-elles brutes ou dérivées? Quelle est leur unité de mesure?

f) Commentez le tableau 6.16 (en le reliant au tableau 6.15 si nécessaire.)

Tableau 6.16 - La répartition du temps de travail marchand et non-marchand

		Femmes		Hommes	
	Année	Marchand	Non-marchand	Marchand	Non-marchand
		(proportion du temps de travail total en %)			
		[1]	[2]	[3]	[4]
Pays en développement : milieu urbain					
Colombie	1983	24	76	77	23
Indonésie	1992	35	65	86	14
Kenya	1986	41	59	79	21
Pays en développement : milieu rural					
Bangladesh	1990	35	65	70	30
Kenya	1988	42	58	76	24
Philippines	1975-77	29	71	84	16
Pays industrialisés					
Canada	1992	39	61	65	35
France	1985-86	30	70	62	38
Italie	1988-89	22	78	77	23
Moyenne de 9 pays en développement		34	66	76	24
Moyenne de 14 pays industrialisés		34	66	66	34

Source : PNUD, Rapport mondial sur le développement humain 1995.

Note : Le temps est divisé en deux parties : les activités privées (que personne ne peut faire à notre place, comme manger ou dormir) et les activités productrices (qui peuvent être effectuées par une tierce personne: quelqu'un peut cuisiner à votre place, par exemple). Le tableau indique la répartition des activités productrices selon qu'elles ont une valeur marchande (travail salarié, produits d'un jardin même s'ils sont autoconsommés) ou non marchande (travaux ménagers, activités communautaires volontaires).

2. La mort aux trousses

a) Commentez le tableau 6.17 en vous servant de la grille d'analyse présentée au début de ce chapitre.

b) Représentez les données selon le sexe et l'origine ethnique à l'aide d'un diagramme en bâtons.

Tableau 6.17 - Taux d'homicide selon l'âge, le sexe et l'origine ethnique aux États-Unis

Nombre de morts pour 100 000 individus en 1991-93

Selon l'âge		Selon le sexe et l'origine ethnique	
1 à 4 ans	2,8	Hommes hispaniques	29,5
5 à 14 ans	1,6	Autres hommes blancs	5,9
15 à 24 ans	22,7	Hommes noirs	70,4
25 à 34 ans	17,6	Femmes hispaniques	4,8
35 à 44 ans	11,3	Autres femmes blanches	2,6
45 à 54 ans	7,6	Femmes noires	13,5
55 à 64 ans	5		
65 à 74 ans	3,8		
75 à 84 ans	3,9		
85 ans et plus	4,1	**Total de la population**	**10,7**

Source : Health United States 1995, Département de la santé et des services humains des États-Unis.

3. L'avortement au pays de Billy Graham

a) Commentez le tableau 6.18 en vous servant de la grille d'analyse présentée au début de ce chapitre.

b) Représentez les données à l'aide d'un ou plusieurs graphiques.

Tableau 6.18 - Taux d'avortement légal selon l'âge, la race et l'état civil de la femme aux États-Unis

Nombre d'avortements pour 100 naissances vivantes

	1973	1992		1973	1992
Total	19,6	33,5	Moins de 15 ans	123,7	79
			15 à 19 ans	53,9	44
Blanches	32,6	23,6	20 à 24 ans	29,4	37,6
Noires	42	51,8	25 à 29 ans	20,7	22,2
			30 à 34 ans	28	18,3
Mariées	7,6	8,4	35 à 39 ans	45,1	25,6
Célibataires	139,8	79	40 ans et plus	68,4	45,4

Source : Health United States 1995, Département de la santé et des services humains des États-Unis.

4. L'endémie

a) Commentez le tableau 6.19 en vous servant de la grille d'analyse présentée au début de ce chapitre.

b) Tracez deux courbes montrant l'évolution de l'incidence du SIDA selon le sexe. Note : utilisez une échelle logarithmique pour l'axe vertical.

Tableau 6.19 - Évolution de l'incidence du SIDA chez les Américains de 13 ans et plus selon le sexe

Nombre estimé de nouveaux cas

Semestre	**Femmes**	**Hommes**	Semestre	**Femmes**	**Hommes**
1-1985	360	4 700	1-1990	2 600	19 200
2-1985	460	6 000	2-1990	2 600	18 700
1-1986	700	7 800	1-1991	3 200	21 500
2-1986	750	9 300	2-1991	3 400	21 600
1-1987	1 100	12 000	1-1992	4 000	24 000
2-1987	1 400	13 000	2-1992	4 100	23 400
1-1988	1 700	14 900	1-1993	4 600	24 100
2-1988	1 800	15 200	2-1993	4 600	23 500
1-1989	2 100	17 500	1-1994	5 200	25 600
2-1989	2 200	17 500	2-1994	5 200	24 800

Source : Health United States 1995, Département de la santé et des services humains des États-Unis.

DOSSIER 6 LE PEUPLE RÉCLAME DU PAIN

On piétine un drapeau.

À Versailles, le 1er octobre 1789, la famille royale assiste à un banquet en l'honneur du régiment des Flandres. Pendant que l'orchestre attaque *Ô Richard, ô mon roi, l'univers t'abandonne*, les invités, excités par le vin, piétinent le drapeau tricolore. Pendant ce temps, le pain, source presque unique de nourriture pour bien des gens, manque à Paris.

Parfois, une banale provocation peut, dans le contexte approprié, avoir des conséquences historiques. Quelques jours après l'incident, les Parisiennes marchent sur Versailles pour aller réclamer du pain au roi. Les méchantes langues disent que la reine s'étonna que, faute de pain, le peuple n'eût pas pensé à manger de la brioche. Le 6 octobre, la foule en cortège ramène la famille royale, et quelques charrettes de pain, à Paris.

Le 21 janvier 1793, à 10 h 22, en même temps que Louis XVI, c'est la monarchie et le pouvoir absolu qui sont guillotinés.

Au Siècle des lumières, la faim n'a pas encore disparu.

Au XVIIIe siècle, le Siècle des lumières, le peuple français souffre encore périodiquement de la faim. Pour en savoir plus long, nous avons extrait quelques chiffres du livre *Machinisme et bien-être* de Jean Fourastié. Le tableau D6.1 montre que la variable clé n'est pas tant le salaire du chef de famille (qui est stable) que le prix du pain (qui fluctue beaucoup). Nous avons complété ce tableau par quelques données prises dans le corps du texte de Jean Fourastié (voir tableau D6.2).

Tableau D6.1 - Salaires et pain en France au siècle des Lumières

Valeurs en livres (ancien nom du franc)

	Salaire horaire d'un journalier	**Prix du setier de blé à Alençon**
1760	0,080	16,85
1765	0,085	15,95
1770	0,090	34,40
1775	0,090	27,85
1780	0,090	21,80
1785	0,095	24,55
1790	0,100	27,90

Source : Jean Fourastié, machinisme et bien-être, Les Éditions de minuit, Paris, 1962.

Note : le setier de blé pèse 240 livres ou 117,5 kg.

Notre hypothèse est que le salaire du travailleur ne lui permet pas toujours de bien nourrir sa famille : à la veille de la Révolution de 1789, le peuple français n'est toujours pas à l'abri de la famine. Nous supposerons, pour notre mini-étude, que la famille typique se compose de 5 personnes (voir tableau D6.2), que le salaire est principalement consacré à l'alimentation (le « pain quotidien »), et qu'un kilogramme de blé équivaut à un kilogramme de pain.

Tableau D6.2 - Besoins caloriques d'une famille typique

(en calories)	**Au XVIIIe siècle**	**Aujourd'hui**
Enfant 1	1900	
Enfant 2	1900	
Adolescent	3400	
Mère	2800	2200
Père	4300	3000
Total	**14300**	
Moyenne par personne	**2860**	

Pour vérifier notre hypothèse, nous devons convertir le « gagne-pain » (le salaire) en calories. C'est ce que nous faisons, étape par étape dans le tableau D6.3 pour deux années typiques : 1785 (année difficile) et 1765 (année relativement faste). La figure D6.1 (schéma de variables) reprend la même démarche sans les chiffres.

Tableau D6.3 - Conversion du salaire en calories

(On suppose que le salaire du travailleur journalier est consacré uniquement à l'achat de pain de blé.)

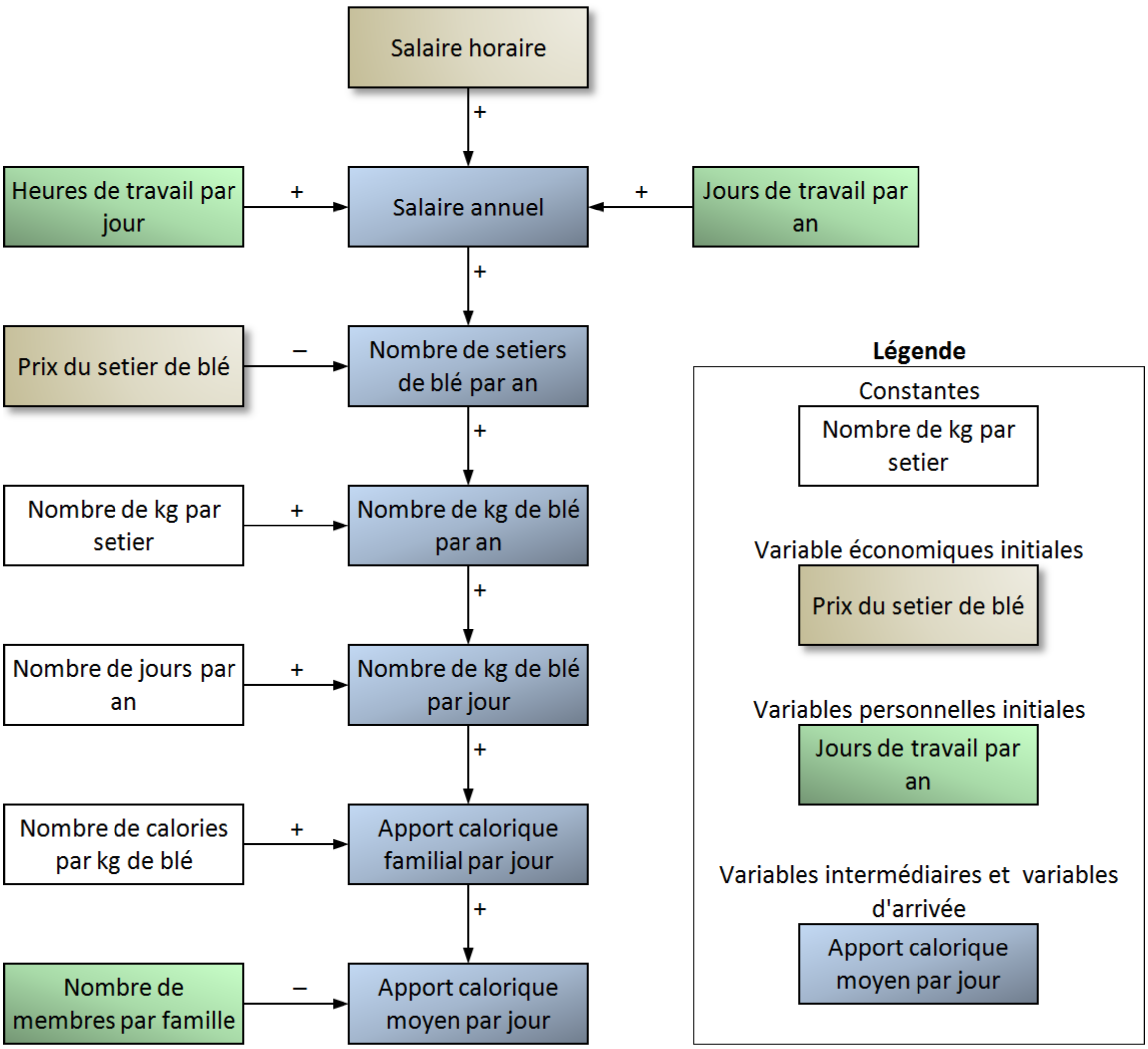
	1785	**1765**	
1	**0,095**	**0,085**	Salaire horaire (en livres)
2	**10**	**10**	Nombre d'heures travaillées par jour
3	**290**	**290**	Nombre de jours travaillés par an
4 = (1 × 2 × 3)	275,5	246,5	Salaire annuel (en livres)
5	**24,55**	**15,95**	Prix du setier de blé (en livres)
6 = (4 / 5)	11,22	15,45	Nombre de setiers consommés par an
7	*117,5*	*117,5*	Poids du setier de blé (en kg)
8 = (6 × 7)	1319	1816	Nombre de kg de blé par an
9	*365*	*365*	Nombre de jours où il faut manger dans l'année
10 = (8 / 9)	3,613	4,975	Quantité de blé disponible par jour (en kg)
11	*2380*	*2380*	Nombre de calories par kg de pain
12 = (10 × 11)	8598	11841	Apport calorique journalier pour la famille (en calories)
13	**5**	**5**	Nombre de membres dans la famille
14 = (12 / 13)	1720	2368	Apport calorique moyen par individu (en calories)

Note : les chiffres figurant en italiques sont des constantes, les chiffres en gras sont les variables initiales, les autres chiffres sont des données déduites des précédentes.

Explication du tableau

Le tableau D6.3 montre bien que c'est le prix du pain qui a la plus grande influence sur le niveau de vie de la famille du travailleur. En 1765, année où le prix du blé est très bas, chaque membre de la famille mange à sa faim (2368 calories) sans toutefois atteindre le niveau recommandé (2860 : voir tableau D6.2). En 1785, quelques années à peine avant la Révolution, la consommation est bien inférieure aux besoins.

Figure D6.1 - Schéma de variables : Influence du salaire horaire et du prix du blé sur l'alimentation de la famille du travailleur



Explication du schéma

L'alimentation d'une famille moyenne (*apport calorique moyen par jour* : case bleue au pied du schéma) dépend du *salaire horaire* du chef de famille et du *prix du blé* (cases beiges : nos deux variables clés). La hausse du salaire améliore l'alimentation (signe positif sur la flèche), alors que la hausse du prix la détériore (signe négatif sur la flèche).

Nous n'avons pas considéré le nombre de *membres par famille*, d'*heures de travail par jour* et de *jours de travail par an* (cases vertes) comme de véritables variables. Dans le premier cas, il s'agit de situations particulières alors que nous nous intéressons à la vie du peuple en général. Dans les deux autres, la marge de manoeuvre est faible compte tenu du temps disponible et des coutumes (la messe du dimanche et les fêtes religieuses). Comme pour les cases beiges, le signe indique le sens de la relation (directe ou inverse).

Les cases blanches sont de simples constantes de conversion et n'ont pas d'influence sur la situation. En effet, même si on inventait le kilo de 1200 grammes, ça ne ferait pas plus de pain sur la table.

L'*apport calorique moyen par jour* (dernière case du schéma, dont il ne part aucune flèche) représente la variable d'arrivée. Les autres cases bleues sont des variables intermédiaires qui nous permettent d'arriver au résultat final.

QUESTIONS

Le prix du setier

En sachant que le prix du setier de blé s'établissait alors à 27,90 livres et que le salaire horaire était de 0,10 livre, calculez l'apport calorique moyen pour le membre d'une famille vivant en 1790.

LECTURES

Jean FOURASTIÉ, *Machinisme et bien-être*, Les Éditions de minuit, Paris, 1962.
Albert SOBOUL, *La Révolution française*, PUF, Paris.

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 partisanes, 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 ce 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).

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 %.

Taux de réponse = Nombre d'individus qui répondent/Nombre d'individus dans l'échantillon

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

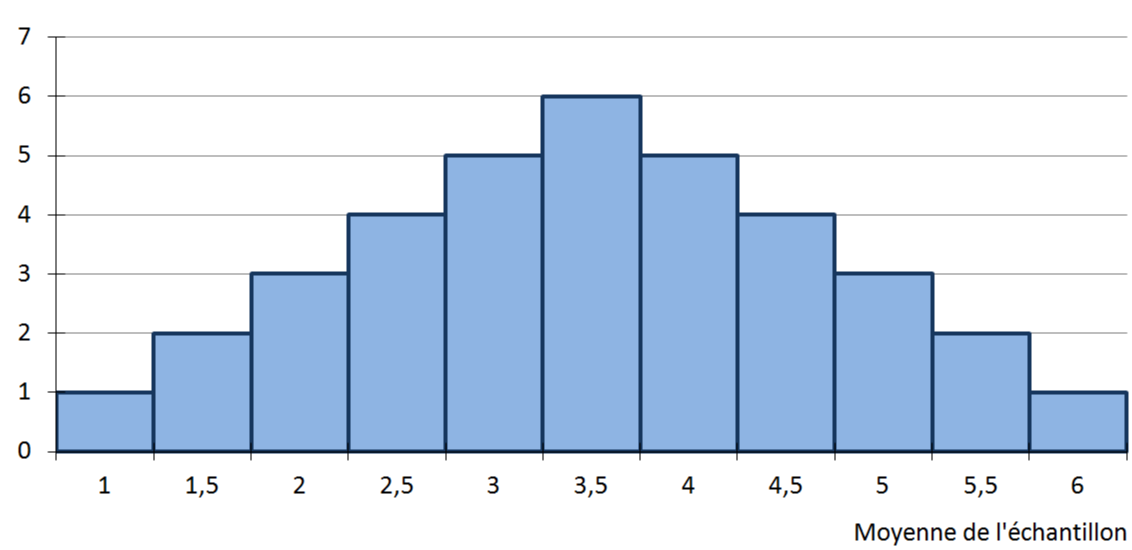
Él. 1	Él. 2	Moy.	Él. 1	Él. 2	Moy.	Él. 1	Él. 2	Moy.
6	6	6,0	4	6	5,0	2	6	4,0
6	5	5,5	4	5	4,5	2	5	3,5
6	4	5,0	4	4	4,0	2	4	3,0
6	3	4,5	4	3	3,5	2	3	2,5
6	2	4,0	4	2	3,0	2	2	2,0
6	1	3,5	4	1	2,5	2	1	1,5
5	6	5,5	3	6	4,5	1	6	3,5
5	5	5,0	3	5	4,0	1	5	3,0
5	4	4,5	3	4	3,5	1	4	2,5
5	3	4,0	3	3	3,0	1	3	2,0
5	2	3,5	3	2	2,5	1	2	1,5
5	1	3,0	3	1	2,0	1	1	1,0

Résumé du tableau

Moyenne de l'échantillon	Nombre d'échantillons
6,0	1
5,5	2
5,0	3
4,5	4
4,0	5
3,5	6
3,0	5
2,5	4
2,0	3
1,5	2
1,0	1
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.

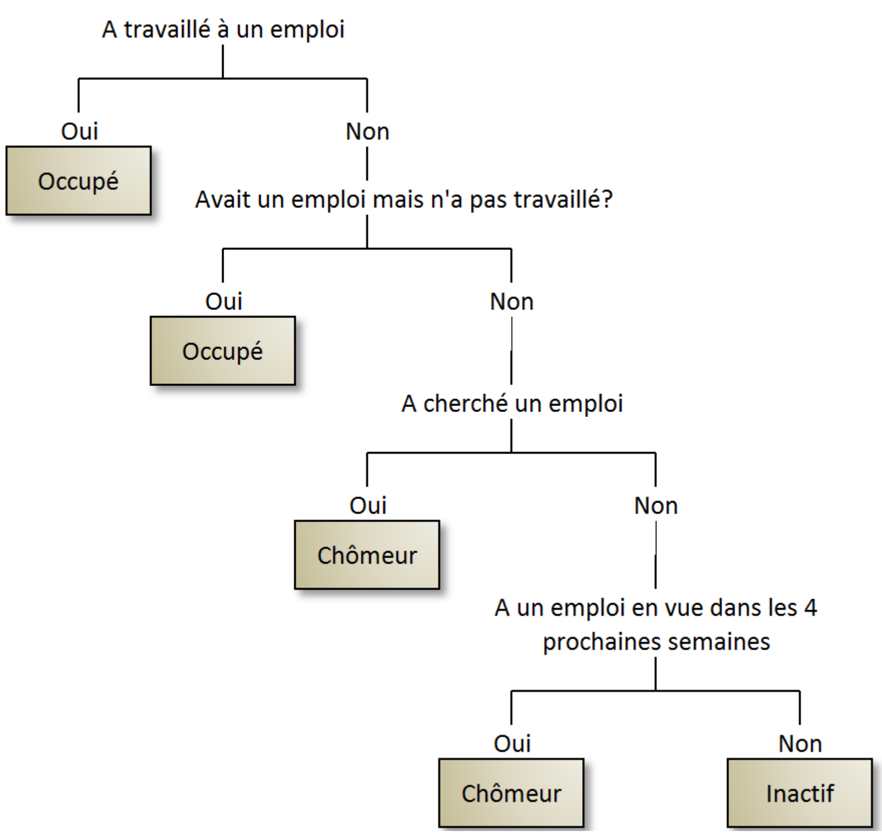
L'échantillonnage par grappes n'est possible que lorsque la population étudiée se présente sous forme d'une hiérarchie (par exemple : pays, régions, villes, immeubles, ménages, individus). Lorsque cette condition est remplie, l'échantillonnage par grappes (qui permet dans le même temps ou pour le même 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.

Poids d'échantillonnage = Taille de la population/Taille de l'échantillon

Poids d'échantillonnage au Canada = 23 millions/62 000 ≈ 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.

Taille de l'échantillon = Taille de la population/Poids d'échantillonnage

Taille de l'échantillon au Québec = 5 730 000/environ 400 ≈ 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.

Résultat corrigé = (Score de la catégorie 1 × Poids de la catégorie 1) + (Score 2 × Poids 2) + …

Résultat corrigé = (48 × 0,5) + (38 × 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-grand-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, fausse, 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.

a) Quelle est la proportion de Québécois de 15 ans et plus résidant en Gaspésie?

b) Utilisez la proportion calculée en *a* pour déterminer approximativement la taille de l'échantillon interrogé en Gaspésie.

c) 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).

a) D'après vous, est-ce que le « oui » a des chances de passer? Justifiez votre réponse avec des chiffres.

b) 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	(en %)
Les médecins	24
Les scientifiques	21
Les professeurs d'université	21
Les gens d'affaire	12
Les journalistes	10
Les avocats	4
Les politiciens	2
Autres	6
Taille de l'échantillon	*1500*

États-Unis 2013	(en %)
Les infirmières	82
Les instituteurs	70
Les médecins	69
Les membres du clergé	47
Les banquiers	27
Les journalistes	21
Les vendeurs d'automobiles	9
Les membres du Congrès	8

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 (IMC inférieur à 20)	**Poids recommandé** (IMC entre 20 et 25)	**Excès de poids possible** (IMC entre 25 et 27)	**Excès de poids** (IMC supérieur à 27)	**Non disponible**	**Total**
	(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	(en %)	**Selon la tranche de revenu du consommateur**	(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. Stratifions

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'infirmer — 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.

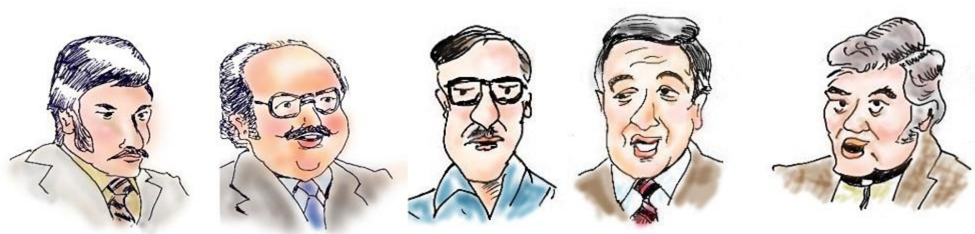
Cinq Québécoises et cinq Québécois des années 2000

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 résidants 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 parait 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.

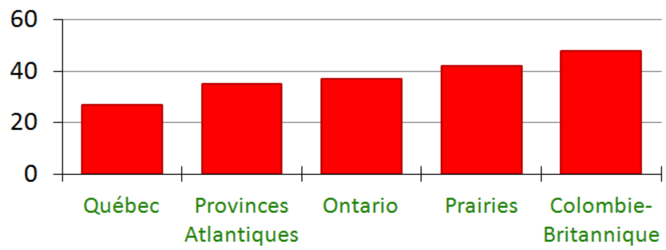
Le nouveau pari pascalien.

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?

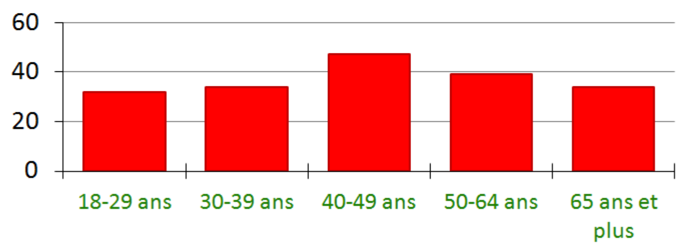
Figure D7.1 - Qui croit encore au Père Noël?

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 plait, 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 résidants 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.

Question	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.

CHAPITRE 8 L'ESTIMATION ET LE TEST D'HYPOTHÈSE

TABLE DES MATIÈRES

1. Estimer des proportions
2. Estimer une moyenne
3. Vérifier une hypothèse à l'aide d'un échantillon
4. Test d'hypothèse sur deux moyennes

- Exercices supplémentaires
- Dossier

Lorsqu'on veut connaître les caractéristiques ou les opinions d'une population, on a souvent recours à un échantillon. Cependant, comme le hasard s'en mêle parfois, toute estimation faite à partir d'un échantillon comporte une marge d'erreur plus ou moins grande. Nous verrons dans ce chapitre comment évaluer cette marge d'erreur. On dira, par exemple, que 79 % des Canadiens trouvent les tribunaux trop complaisants envers les criminels (avec une marge d'erreur de plus ou moins 3 %, 19 fois sur 20), ou que la taille moyenne des Américains de sexe masculin âgés est comprise entre 174,7 cm et 175,7 cm (avec un niveau de confiance de 95 %). Les deux premières sections de ce chapitre seront consacrées à ce genre d'estimation.

Les échantillons peuvent aussi nous permettre de vérifier des hypothèses. Peut-on affirmer, par exemple, que les criminels ont une capacité crânienne inférieure à la moyenne, ou encore que les jeunes filles ont une estime de soi inférieure à celle des jeunes garçons? Là encore, puisque nous utilisons des échantillons, le hasard peut nous jouer des tours. Nous verrons, dans les sections 3 et 4 du présent chapitre, comment évaluer les chances de nous tromper en acceptant ou en rejetant une hypothèse.

Pour nous aider dans notre démarche, nous utiliserons des outils familiers : la moyenne, l'écart type et la courbe de distribution normale. Comme il arrive parfois que les échantillons soient très petits, surtout lorsqu'on étudie un groupe d'individus bien particulier, nous aurons parfois recours à une autre distribution, bien connue des chercheurs en laboratoire : la distribution de Student. Nous pourrons alors, sans grand effort supplémentaire, soutirer plus d'information à l'échantillon.

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

- Comment peut-on, à partir d'un échantillon, estimer la proportion d'individus possédant telle ou telle caractéristique dans une population donnée?
- Comment peut-on, à partir d'un échantillon, estimer la valeur moyenne d'une caractéristique des membres d'une population?
- Quelle est, lorsque l'on fait une estimation, la marge d'erreur possible correspondant à un niveau de confiance donné?
- Comment peut-on vérifier une hypothèse fondée sur une caractéristique d'une population?
- Comment peut-on vérifier une hypothèse fondée sur la comparaison d'une même caractéristique dans deux populations différentes?

1. ESTIMER DES PROPORTIONS

Grâce à un échantillon, nous commencerons par déterminer comment la population se répartit entre les diverses catégories d'une variable. Il nous restera ensuite à évaluer les risques de faire ainsi une estimation éloignée de la réalité.

1.1. Calcul de l'erreur type

Le sondage dont il est question dans le tableau 8.1 porte sur l'opinion des Canadiens concernant la criminalité dans les années 1990. On constate qu'une forte majorité de gens trouvent que les tribunaux manquent de sévérité envers les criminels. En guise de consolation pour le Québec, signalons que, selon la même enquête, ceux qui croient que le taux de criminalité violente est en augmentation dans leur localité s'avèrent moins nombreux à Montréal (45 %) qu'à Toronto (57 %). À vous de vérifier si les gens d'aujourd'hui sont aussi sévères que leurs aînés.

Tableau 8.1 - Échec au crime

Les tribunaux canadiens…	**Opinion des Canadiens** (en %)
ne sont pas assez sévères envers les criminels	79
traitent les criminels de façon correcte	16
traitent les criminels d'une façon trop sévère	2
(Autres réponses)	3
	100

	Canada	**Québec**
Marge d'erreur à 95 %	3,1 %	6 %
Taille de l'échantillon	1016	269

Source : Sondage Gallup paru dans La Presse du 17 juillet 1995.

Nous voulons savoir dans quelle mesure les caractéristiques de l'échantillon observé risquent d'être éloignées de celles de la population que nous cherchons à connaître.

Pour savoir que 79 % de la *population* pense que les tribunaux manquent de sévérité, les sondeurs ont observé un *échantillon* de 1016 individus : ils ont fait une *estimation*. Il reste maintenant à quantifier le degré de fiabilité (ou la *marge d'erreur*) de cette estimation. C'est grâce au calcul de cette marge d'erreur que le chercheur pourra déterminer si l'estimation est suffisamment précise pour les besoins de sa recherche. Inversement, cette méthode permettra d'évaluer la taille minimale de l'échantillon à observer compte tenu de la précision recherchée.

Plus les échantillons sont grands, plus ils sont groupés autour de leur valeur moyenne, ou, si l'on préfère, moins ils sont dispersés.

Nous savons déjà un certain nombre de choses sur les échantillons choisis au hasard. Il est possible de tirer d'une population une quantité considérable d'échantillons. Certains de ces échantillons sont tout à fait représentatifs de la population et d'autres s'en écartent plus ou moins. Lorsque la taille des échantillons est suffisamment grande*, les différentes valeurs des paramètres mesurés à partir de ces échantillons se distribuent selon le modèle de la courbe normale : cela signifie que les échantillons correspondent, en moyenne, à la population, et que l'on peut évaluer leur degré de dispersion.

Cela peut varier d'un minimum de 10 à une centaine d'éléments. Le tout dépend de la façon dont les proportions sont distribuées.

L'erreur type mesure le degré de dispersion des échantillons autour de leur valeur moyenne.

Ce degré de dispersion des échantillons, qui se nomme erreur type, est relativement facile à calculer et à interpréter (c'est le pendant de l'écart type, vu au chapitre 3). Étant donné que l'erreur type est très utilisée, notamment dans les sondages électoraux et les enquêtes sociales, nous allons maintenant expliquer comment on doit s'en servir.

Essayons donc d'évaluer le degré de fiabilité de l'opinion exprimée dans notre sondage du tableau 8.1 (« une proportion de 79 % des gens croit que… ») en obtenant tout d'abord son erreur type. La formule ci-dessous indique comment calculer l'erreur type et, malgré les apparences, elle est plutôt simple.

$$\textbf{Erreur type d'une proportion} = \sqrt{\frac{p(1-p)}{n}}$$

La valeur *p* représente la proportion de la catégorie étudiée (estimée) et *n* la taille de l'échantillon. En principe, on devrait utiliser les proportions de la population, mais comme on ne connaît pas ces proportions on doit se contenter de celles de l'échantillon.

- On calcule d'abord le produit des deux proportions (celle de la catégorie qui nous intéresse et celle des autres catégories réunies). Pour simplifier, on peut écrire les deux proportions sous leur forme décimale (la somme des deux proportions est alors égale à 1).

  Ici, les deux proportions sont respectivement 0,79 (« une proportion de 79 % des gens croit que… ») et 0,21. Leur produit est égal à 0,79 x 0,21 = 0,1659

- On divise le résultat obtenu par le nombre d'éléments dans l'échantillon.

Ici, l'échantillon est de 1016 individus. On obtient donc 0,1659/1016 = 0,0001633

- On calcule la racine carrée du résultat obtenu.

  √(0,0001633) = 0,0128 = 1,28 % (disons 1,3 %)

1.2. Interprétation de la marge d'erreur

La véritable proportion se situe, avec un certain degré de *probabilité*, dans un *intervalle* plus ou moins large autour de la valeur estimée.

Rappelons que pour connaître l'opinion des gens sur la criminalité, les sondeurs ont tiré au hasard un échantillon de 1016 individus. Or, il existe une quantité innombrable d'échantillons de 1016 individus, et certains d'entre eux peuvent, *par malchance*, donner des résultats *relativement éloignés* de la réalité. Grâce à l'erreur type que nous avons calculée, sa valeur s'établissant à 1,3 %, nous allons maintenant pouvoir chiffrer avec précision ce que nous entendons par les expressions *relativement éloignés* et *par malchance*.

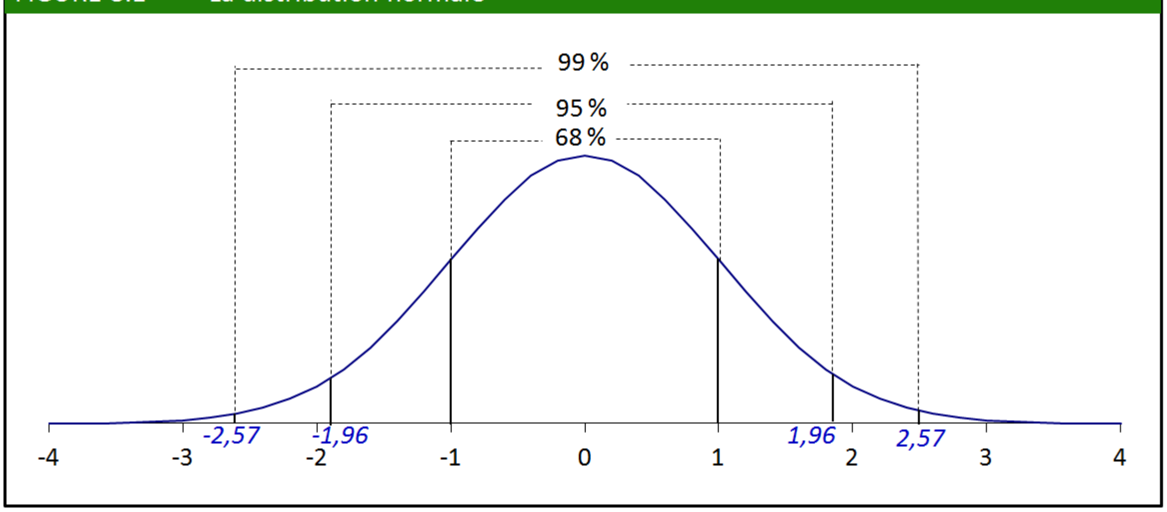
Cette probabilité se nomme niveau de confiance et cet intervalle se nomme intervalle de confiance.

On se rappellera que, dans une distribution normale, 68 % (ou 0,68) des éléments sont contenus à une distance de la moyenne ne dépassant pas un écart type. Ici, les éléments sont les différents échantillons que nous aurions pu tirer, et l'écart type de cet ensemble d'échantillons s'appelle l'erreur type. Cela signifie que la proportion estimée grâce à notre échantillon a 68 % de *chances* de ne pas *s'écarter* de plus de 1,3 % de la proportion véritable.

En ratissant plus large, on augmente le niveau de confiance.

En pratique, on cherche à ratisser plus large de façon à faire des affirmations un peu plus dignes de confiance. En augmentant le nombre d'erreurs types à 1,96, on récupère 95 % de tous les échantillons (voir la figure 8.1). De façon similaire, 99 % de tous les échantillons sont compris entre -2,57 et +2,57 erreurs types. Dans la phrase précédente, le premier chiffre est le niveau de confiance (99 %) et les deux autres chiffres représentent l'intervalle de confiance (entre -2,57 et 2,57).

FIGURE 8.1 - La distribution normale



Marge d'erreur à un niveau de confiance de 95 % = 1,96 × Erreur type

Marge d'erreur à un niveau de confiance de 99 % = 2,57 × Erreur type

Dans notre exemple, 1,96 erreur type représente environ 2,5 % (on multiplie 1,96 par l'erreur type, soit 1,96 x 1,3 % = 2,5 %). On peut donc dire, en simplifiant un peu, que dans 95 % des cas, la proportion véritable se situera entre [79 % – 2,5 %] et [79 % + 2,5 %], ou entre 76,5 % et 81,5 %. Pour nous donner des allures de journaliste, nous dirions aussi que, 19 fois sur 20*, le résultat peut varier au maximum de plus ou moins 2,5 %.

L'expression *19 fois sur 20* représente le niveau de confiance de 95 %, car 19/20 = 0,95 = 95 %.

> Étant donné que la proportion véritable est fixe et que c'est en réalité notre estimation qui a des chances de coïncider plus ou moins bien avec la réalité, on devrait plutôt dire que cet intervalle de confiance (situé entre 76,5 % et 81,5 %) a 95 % de chances de contenir la proportion véritable.

Dans le tableau 8.1 vu précédemment, il y a trois catégories d'opinion et donc trois proportions. Étant donné que l'erreur type dépend de la valeur de la proportion, chaque proportion possède sa propre erreur type et sa propre marge d'erreur. Voilà qui est un peu fastidieux pour le chercheur et ennuyeux pour le lecteur. Pourquoi ne pas mettre les choses au pire en calculant l'erreur type d'une proportion de 50 % (c'est cette proportion qui donne l'erreur type la plus élevée)? C'est ce que font la plupart des instituts de sondage lorsqu'ils publient les résultats de leur enquête dans les journaux. Vérifions si c'est bien le cas dans le tableau 8.1, où la marge d'erreur indiquée est de 3,1 % pour le Canada.

Erreur type = √(0,5 × 0,5)/1016 = √(0,25/1016) = √(0,000246) = 0,0157 = 1,57 %

Marge d'erreur (pour un niveau de confiance de 95 %) = 1,96 x 1,57 % = 3,1 %

Voici une autre manière de calculer l'erreur type :

√(0,5 × 0,5)/√(1016) = √(0,25)/√(1016) = 0,5/31,8 = 0,0157 = 1,57 %

1.3. Le coefficient de variation

Parfois, la marge d'erreur d'une estimation s'avère trop grande pour jouir de la moindre valeur. Lorsqu'une proportion estimée à 40 % possède une erreur type de 1 %, on ne peut pas se plaindre. Par contre, une estimation de 1 % avec une erreur type de 0,5 % ne vaudrait pas grand-chose. On comprend intuitivement que l'erreur type doit pouvoir être comparée à la proportion à laquelle elle se rapporte : ce *rapport* s'appelle le coefficient de variation.

$$\textbf{Coefficient de variation d'une proportion} = \frac{\text{Erreur type}}{\text{Proportion}}$$

Selon l'Enquête sociale générale de Statistique Canada, on comptait 20 525 000 parents adoptifs potentiels au Canada en 1990. On peut constater, par exemple, dans le tableau 8.2 que 4 447 000 d'entre eux n'avaient jamais vécu en couple et que 163 000 d'entre eux ont adopté au moins deux enfants au cours de l'année. Ces données ont été estimées à partir d'un échantillon portant sur 13 495 individus. On peut en déduire que chaque parent sondé représente 20 535 000/13 495 = 1521 parents dans la population. C'est ce qu'on appelle le *poids d'échantillonnage* (voir chapitre 7).

Tableau 8.2 - Nombre d'adoptions selon le type d'union des parents adoptifs				
	Aucune	**1 enfant**	**2 enfants ou +**	**Total**
Type d'union	(Quantités estimées en milliers)			
Aucune union	4 446	1	0	**4 447**
Une union	12 625	*307*	126	**13 058**
Deux unions ou plus	2 830	74	35	**2 939**
Non déclaré	79	0	2	**81**
Total	**19 980**	**382**	**163**	**20 525**

Taille de l'échantillon (n)	13 495
Poids d'échantillonnage	1 521

Source : Statistique Canada, Enquête sociale générale 1994, données de 1990.

Note : Dans le premier sous-tableau, toutes les estimations inférieures à 14 (milliers) devraient être effacées, car elles sont trop sujettes à caution pour être publiées (leur coefficient de variation dépasse 33 %). Les estimations comprises entre 14 et 55 (milliers) devraient être interprétées avec précaution. Quant aux estimations qui dépassent 55 (milliers), elles peuvent être considérées comme suffisamment fiables (leur coefficient de variation est inférieur à 16 %).

Dans le tableau 8.2, les estimations sont exprimées en quantités. Mais il s'agit en réalité de proportions déguisées. Sur un total de 20 525 milliers de parents adoptifs potentiels, on compte, par exemple, 307 milliers de parents qui ont été « mariés » une seule fois dans leur vie et qui ont adopté un seul enfant : la proportion est de 307/20 525 = 1,5 %.

Calculons d'abord l'erreur type de cette proportion en utilisant la formule vue un peu plus haut :

p = 1,5 % = 0,015 (la proportion des parents qui appartiennent à cette catégorie)

(1 – p) = 0,985 (la proportion des parents qui n'appartiennent *pas* à cette catégorie)

n = 13 495 (la taille de l'échantillon interrogé)

Erreur type = √[(0,015 x 0,985)/13 495] = √[0,0148/13 495] = 0,00105 = 0,105 %

Il ne nous reste plus qu'à calculer le coefficient de variation:

Coefficient de variation = 0,105 %/1,5 % = 0,07 = 7 %

Lorsque le coefficient de variation dépasse 16 % (environ), on doit interpréter les estimations avec prudence. Lorsque le coefficient dépasse 33 %, les estimations n'ont tout bonnement aucune valeur. Dans un tel cas en effet, et avec un niveau de confiance de 95 % (soit 1,96 erreur types… disons 2 pour simplifier), cela signifie que les bornes de l'intervalle de confiance seraient situées à plus ou moins 66 % (2 x 33 %) de l'estimation calculée. Si l'estimation était de 3 %, par exemple, cela signifierait qu'un intervalle compris entre 1 % (soit 3 % – [66 % x 3 %] et 5 % (soit 3 % + [66 % x 3 %]) aurait 95 % de chance de contenir la proportion véritable. Ce serait beaucoup trop vague.

EXERCICES 1

1. Échantillon réduit, erreur type accrue

Dans le tableau 8.1, l'institut Gallup évalue que, dans 95 % des cas, la marge d'erreur ne dépasse pas 6 % pour les estimations concernant le Québec.

a) Calculez l'erreur type pour le Québec en vous basant sur une proportion de 50 %.

b) Vérifiez s'il est exact que, 19 fois sur 20, la marge d'erreur ne dépasse pas 6 % pour le Québec.

2. Coefficient de variation

Calculez le coefficient de variation des résultats pour le Québec à partir du tableau 8.1, toujours en vous basant sur une proportion de 50 %.

2. ESTIMER UNE MOYENNE

Dans la section précédente, nous avons vu comment un échantillon permet, avec un degré plus ou moins grand de fiabilité, d'estimer une proportion. Nous avions par exemple estimé que 79 % des Canadiens trouvaient leurs tribunaux trop indulgents, ou que 163 000 parents sur 20 525 000 (soit 7,9 % d'entre eux) avaient adopté deux enfants ou plus.

Un principe similaire s'applique lorsqu'il s'agit d'évaluer une moyenne à partir d'un échantillon. Quelle est la taille moyenne des quinquagénaires américains? Quel est le QI moyen des gens originaire de l'Extrême Orient? Quel est le niveau moyen d'estime de soi des élèves québécois? Dans tous ces cas, on pourrait faire une estimation de la moyenne en observant un simple échantillon de la population.

La taille de l'échantillon n'a pas besoin d'être très grande, car tout dépend du degré de précision désiré et des informations dont on dispose sur la population sondée. Lorsque l'échantillon contient au moins 30 éléments, on utilise encore une fois la distribution normale pour calculer la marge d'erreur. Lorsque l'échantillon est plus petit (ce qui est souvent le cas lors d'expériences de laboratoire), il y a encore moyen de s'en sortir grâce à une autre distribution : le t de Student.

Évidemment, il n'est pas interdit de travailler avec des échantillons relativement grands. Nous allons voir dans l'exemple qui suit que quelques centaines d'individus bien choisis nous permettent d'évaluer avec une précision étonnante une caractéristique d'une population de plusieurs millions de personnes.

2.1. Un gros échantillon : la taille des anciens combattants

En Europe, les Américains ont la réputation d'être un grand peuple (en centimètres), surtout depuis un certain 6 juin 1944. Pour vérifier la chose et faire des comparaisons, nous avons demandé aux fonctionnaires de Washington, Québec et Ottawa de nous fournir des données précises. Les fonctionnaires américains ont été les premiers à nous répondre, sans nous poser de questions ni nous demander d'argent, et ils nous ont fourni toute une série de paramètres sur la taille de leurs concitoyens par groupe d'âge et origine ethnique : moyenne, écart type, mode, centiles, taille des échantillons correspondants pour chaque groupe étudié. Une véritable aubaine pour le chercheur en sciences humaines. Certains de ces paramètres sont reproduits au tableau 8.3.

Tableau 8.3 - Taille des Américains

	Taille de l'échantillon (n)	Moyenne	Écart type (en cm)	Médiane
Ensemble				
18-24 ans	988	177,0	7,1	177,0
25-34 ans	1067	176,8	6,9	176,8
35-44 ans	745	176,3	7,4	176,5
45-54 ans	690	175,3	6,6	175,3
55-64 ans	1227	173,7	6,9	173,7
65-74 ans	1199	171,2	7,1	171,5
Total: 18-74 ans	**5916**	**175,5**	**7,1**	**175,5**
Blancs				
18-24 ans	846	177,3	7,1	177,0
25-34 ans	901	177,0	6,6	177,0
35-44 ans	653	176,8	7,4	176,8
45-54 ans	617	175,5	6,6	175,3
55-64 ans	1086	173,7	6,9	173,7
65-74 ans	1045	171,7	6,9	171,7
Total: 18-74 ans	**5148**	**175,8**	**7,1**	**175,8**
Noirs				
18-24 ans	121	176,8	6,9	178,1
25-34 ans	139	176,8	6,9	177,0
35-44 ans	70	176,5	6,4	175,3
45-54 ans	62	174,2	6,6	173,0
55-64 ans	129	174,2	6,9	174,5
65-74 ans	128	171,2	7,1	171,7
Total: 18-74 ans	**649**	**175,5**	**6,9**	**175,5**

Source : US Department of Health and Human Services. Enquête effectuée en 1976-1980.

Ainsi, 690 hommes âgés de 45 à 54 ans ont été examinés à travers le pays entre 1976 et 1980, ce qui a permis d'estimer la taille moyenne des gens de ce groupe à 175,2 cm. Essayons d'établir la marge d'erreur de cette estimation, en prenant un niveau de confiance de 0,95. Pour cela, nous devons calculer, ici aussi, l'erreur type de la distribution des moyennes d'échantillons.

$$\textbf{Erreur type d'une moyenne} = \frac{\text{Écart type}}{\sqrt{n}}$$

En principe, on devrait utiliser l'écart type de la population (qui est généralement inconnu) et le diviser par la racine carrée de *n*, soit le nombre d'éléments dans l'échantillon. En pratique, on peut se contenter de l'écart type de l'échantillon étudié, pourvu que l'échantillon contienne au moins 30 éléments. On divise alors l'écart type par √(n), comme dans la formule ci-dessus.

> Pour calculer l'écart type d'un échantillon, on utilise la formule étudiée au chapitre 3 à un détail près : on doit diviser la somme des écarts au carré par (n – 1) et non par n comme c'est le cas pour l'écart type d'une population (n représente le nombre d'éléments dans l'échantillon). Dans l'exemple illustré ici, l'écart type de l'échantillon avait déjà été calculé par ceux qui nous ont communiqué les données. Nous donnerons un exemple détaillé du calcul de l'écart type d'échantillon au tableau 8.4, un peu plus loin.

L'écart type de l'échantillon est de 6,6 cm. L'erreur type est donc ici égale à 6,6/√(690) = 6,6/26,25 = 0,25 cm

Notre intervalle de confiance, pour un niveau de 0,95, se situera à 1,96 erreur type de part et d'autre de la moyenne (revoir la figure 8.1 si nécessaire).

Intervalle de confiance : entre [175,2 – (1,96 x 0,25)] et [175,2 + (1, 96 x 0, 25)]

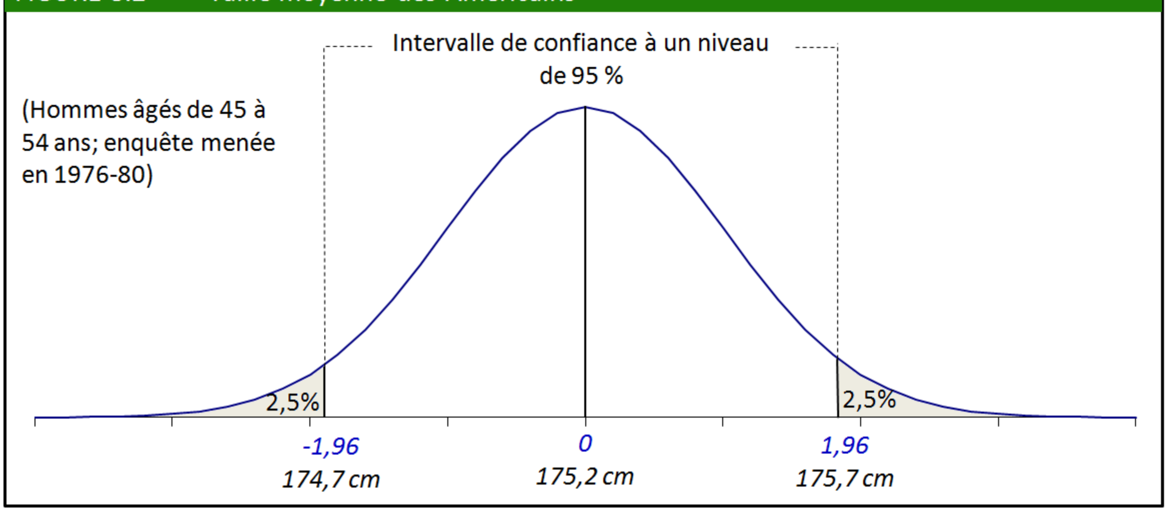
entre (175,2 – 0,5) et (175,2 + 0,5)

entre 174,7 cm et 175,7 cm

Formulé autrement : la marge d'erreur est de 1,96 x 0,25, soit 0,5 cm.

Cela signifie qu'il y a 95 % de chances pour que l'intervalle compris entre 174,7 cm et 175,7 cm contienne la véritable taille moyenne des Américains de ce groupe-là. Dans la plupart des cas, on pourra considérer que la moyenne estimée (à 175,2) est suffisamment précise. La figure 8.2 illustre ce résultat.

FIGURE 8.2 - Taille moyenne des Américains



Étapes de l'estimation : échantillon contenant au moins 30 éléments tirés d'une population normale

- 1. Choisir un niveau de confiance.
- 2. Trouver la valeur correspondante à ce niveau de confiance dans la table de distribution normale.
- 3. Calculer l'erreur type de l'échantillon.
- 4. Calculer la marge d'erreur.
- 5. Calculer les deux bornes de l'intervalle de confiance

  Nous allons maintenant estimer la taille moyenne des hommes de race noire âgés de 45 à 54 ans en suivant les étapes que nous venons de résumer. Cette fois, l'échantillon est beaucoup plus réduit (n = 62) et la moyenne légèrement différente (174,2 cm), mais l'écart type est identique, soit 6,6 cm (revoir le tableau 8.3).

- 1. Niveau de confiance : nous choisissons 0,95. Si jamais nous passons à côté de la cible 1 fois sur 20 (ou 0,05), ce ne sera pas tragique dans un cas semblable.
- 2. Valeur correspondante dans la table : en prenant une surface de 0,475 de part et d'autre de la moyenne, on obtient une surface totale de 0,95 (soit 0,475 x 2); cela correspond à une valeur de 1,96 dans la table de distribution normale.
- 3. Erreur type = Écart type/√(n) = 174,2 cm/√(62) = 6,6 cm/7,87 = 0,84 cm.
- 4. Marge d'erreur : 1,96 x 0,84 cm = 1,6 cm.
- 5. Intervalle de confiance : entre [174,2 – 1,6] et [174,2 + 1,6], soit entre 172,6 et 175,8 cm.

On constate que cet intervalle de confiance est plus large que le précédent. Cet échantillon, parce que plus petit, est moins fiable que le précédent.

On constate également que la borne supérieure de cet intervalle dépasse la borne équivalente pour l'ensemble de la population (qui était de 175,7 cm). Si, en apparence, la moyenne est moins élevée pour les Noirs que pour l'ensemble, il est toujours possible que cet écart observé soit l'effet du hasard d'échantillonnage.

2.2. Un petit échantillon : le prix des abonnements

Les spécialistes de la publicité connaissent fort bien les ficelles quantitatives du métier.

Depuis que les produits sont devenus plus difficiles à vendre qu'à fabriquer, les agences de mise en marché ont inventé toutes sortes de formules pour écouler disques, livres et revues. La maison *Publisher Clearing House* propose par exemple plus d'une centaine d'abonnements annuels à prix réduit sous la forme d'un dépliant coloré accompagné d'une feuille de timbres. Le client n'a qu'à détacher le timbre correspondant à la revue désirée et à coller ce timbre sur le bon de commande pour bénéficier d'une réduction substantielle. Voilà pour le consommateur une façon amusante d'avoir l'impression d'économiser. Pour renforcer cette perception, il va de soi que le montant de la réduction est imprimé en plus gros caractères que le prix effectif de l'abonnement. C'est sans doute la raison pour laquelle les réductions sont libellées en dollars (et non en proportion) : ainsi, l'œil est automatiquement attiré vers les réductions les plus grandes et, par le fait même, vers les revues les plus chères. Notons enfin que l'entreprise dont il est question n'a pas jugé bon de mélanger les réductions en dollars (« 25 $ de rabais sur 499 $ ») avec des réductions exprimées en proportions (« 25 % de rabais sur 2 $ ») comme c'est souvent le cas dans les dépliants publicitaires.

Les publicitaires poursuivent souvent un objectif diamétralement opposé à celui de ce manuel : ils utilisent les chiffres pour semer la confusion dans les esprits plutôt que pour susciter une réflexion méthodique.

Nous avons essayé d'estimer le prix moyen d'un abonnement annuel, à l'aide d'un petit échantillon de 12 éléments (voir le tableau 8.4). Étant donné que la population totale n'est pas très élevée (il y a 103 abonnements annuels offerts dans le dépliant), il aurait été facile de l'observer directement sans passer par un échantillon, avouons-le. Cependant, notre unique but est ici d'illustrer la façon d'estimer une moyenne avec un petit échantillon.

Tableau 8.4 - Les abonnements de *Publisher Clearing House*

		Calcul de l'écart type d'échantillon	
Magazine	Prix (en $)	Écarts²	
Canadian Living	23,98	0,33	= (23,98 - 24,56)²
Craft & Home	22,97	2,52	= etc.
Goldilocks & The Three Bears	20,50	16,46	
Golf Magazine	20,97	12,86	
Maclean's	19,96	21,13	
Playboy	34,50	98,87	
Soap Opera Digest	29,95	29,09	
Word Wize	22,97	2,52	
Islands	28,95	19,30	
Paints Works	28,97	19,48	
Reader's Digest	24,96	0,16	
Toronto Life	16,00	73,22	
Moyenne	**24,56**	**295,93**	somme des écarts²
Écart type	5,19	5,19	√[écarts²/(n-1)]
n (nombre d'éléments dans l'échantillon)	12		

Source : Publisher Clearing House, 1995.

Note : échantillon de 12 magazines dans une population de plus d'une centaine.

Le procédé d'estimation est le même que dans le cas précédent (la taille moyenne des anciens combattants), sauf que la table de la distribution normale (utilisée à l'étape 3) doit être remplacée par la table de la distribution de Student, mieux adaptée aux petits échantillons.

La table de distribution de Student (voir la figure 8.3) fonctionne un peu à l'envers de la table de distribution normale, puisqu'elle donne, pour une surface donnée, le seuil à atteindre (appelé valeur t de Student). Supposons que le niveau de confiance choisi soit de 0,95. On sélectionnera dans la table la colonne 0,025, ce qui éliminera une surface (et une probabilité) de 0, 025 (ou 2,5 %) sur chacune des deux extrémités de la courbe. Il nous reste maintenant à identifier la bonne ligne dans la table (degrés de liberté). C'est la taille de l'échantillon qui détermine le nombre de degrés de liberté, selon la formule suivante :

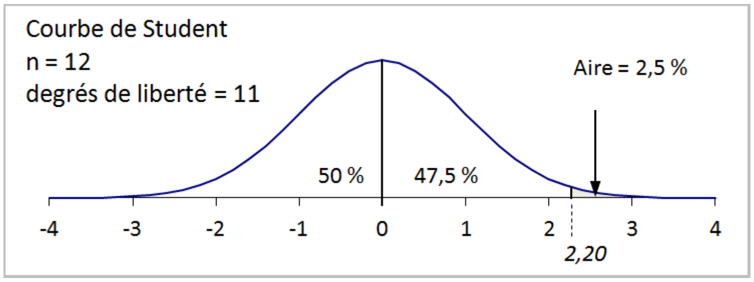
$$\text{Degrés de liberté (dans une estimation de moyenne)} = n - 1$$

où n est la taille de l'échantillon observé.

Dans notre exemple, la valeur t de Student pour une surface de 0,025 et 11 degrés de liberté est de 2,20.

FIGURE 8.3 - Distribution de Student

	Surface couverte à droite du nombre dans la table			
Degrés de liberté	0,05	**0,025**	0,01	0,005
1	6,31	12,71	31,82	63,66
2	2,92	4,30	6,96	9,92
3	2,35	3,18	4,54	5,84
4	2,13	2,78	3,75	4,60
5	2,02	2,57	3,36	4,03
6	1,94	2,45	3,14	3,71
7	1,89	2,36	3,00	3,50
8	1,86	2,31	2,90	3,36
9	1,83	2,26	2,82	3,25
10	1,81	2,23	2,76	3,17
11	1,80	**2,20**	2,72	3,11
12	1,78	2,18	2,68	3,05
13	1,77	2,16	2,65	3,01
14	1,76	2,14	2,62	2,98
15	1,75	2,13	2,60	2,95
16	1,75	2,12	2,58	2,92
17	1,74	2,11	2,57	2,90
18	1,73	2,10	2,55	2,88
19	1,73	2,09	2,54	2,86
20	1,72	2,09	2,53	2,85
21	1,72	2,08	2,52	2,83
22	1,72	2,07	2,51	2,82
23	1,71	2,07	2,50	2,81
24	1,71	2,06	2,49	2,80
25	1,71	2,06	2,49	2,79
26	1,71	2,06	2,48	2,78
27	1,70	2,05	2,47	2,77
28	1,70	2,05	2,47	2,76
29	1,70	2,05	2,46	2,76
30	1,70	2,04	2,46	2,75
35	1,69	**2,03**	2,44	2,72
40	1,68	2,02	2,42	2,70
45	1,68	2,01	2,41	2,69
60	1,67	2,00	2,39	2,66
120	1,66	1,98	2,36	2,62



Étapes de l'estimation : échantillon contenant moins de 30 éléments tirés d'une population normale

- 1. Choisir un niveau de confiance.
- 2. Calculer le nombre de degrés de liberté.
- 3. Trouver la valeur correspondante à ce niveau de confiance et au nombre de degrés de liberté dans la table de distribution de Student.
- 4. Calculer l'erreur type de l'échantillon.
- 5. Calculer la marge d'erreur.
- 6. Calculer les deux bornes de l'intervalle de confiance.

Avant de suivre ces étapes, il nous faut déterminer nous-mêmes la moyenne et l'écart type (ces deux chiffres nous étaient déjà fournis dans l'exemple précédent). Deux méthodes s'offrent à nous. Nous pouvons faire le calcul complet ou utiliser les fonctions d'un chiffrier électronique. Le tableau 8.4 illustre ces deux méthodes et donne une moyenne de 24,56 $ et un écart type de 5,19 $ pour notre échantillon de 12 éléments.

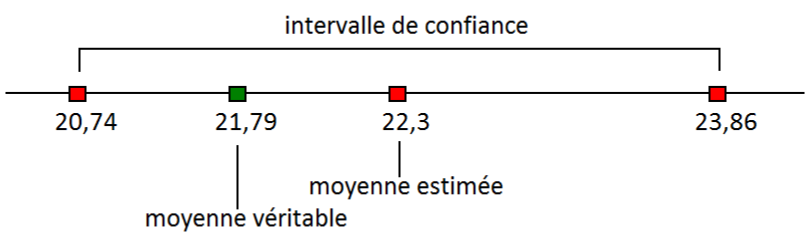
- Niveau de confiance : nous choisissons à nouveau 0,95, c'est-à-dire que nous rejetons 0,025 de chaque côté.
- Nombre de degrés de liberté : 12 – 1 = 11.
- Valeur t de Student correspondant à une surface de 0,025 et 11 degrés de liberté (voir la figure 8.3) : 2,20.
- Erreur type : 5,19 $/√(12) = 1,50 $.
- Marge d'erreur : Valeur t de Student x Erreur type = 2,20 x 1,50 $ = 3,30 $.
- Bornes de l'intervalle de confiance : entre [24,56 – 3,30] et [24,56 + 3,30], soit entre 21,26 $ et 27,30 $. Cela signifie qu'il y a 95 % de chances que le véritable prix moyen de l'abonnement soit compris entre 21,26 $ et 27,30 $. C'est plutôt vague, mais c'est mieux que rien!

Comme on peut le constater, il est difficile de faire une estimation précise dans ce cas-ci, car deux facteurs jouent simultanément contre nous : l'échantillon est très petit et l'écart type est relativement grand. Nous pourrions contourner la difficulté en augmentant la taille de l'échantillon. C'est ce que nous avons fait en sélectionnant 36 éléments au hasard, ce qui nous donne une moyenne de 22,30 $ et un écart type de 4,63 $.

- Erreur type : 4,63 $/√(36) = 0,77 $.
- Niveau de confiance : toujours 0,95.
- Nombre de degrés de liberté : 36 – 1 = 35.
- Valeur dans la table de Student (colonne 0,025 et ligne 35) : 2,03. Notons que lorsque la taille de l'échantillon dépasse 30, la distribution de Student commence à ressembler sérieusement à la distribution normale : nous aurions obtenu une valeur de 1,96 (très proche de 2,03) si nous avions utilisé la table normale pour un niveau de confiance de 0,95. C'est la raison pour laquelle on peut généralement se contenter d'utiliser la distribution normale lorsque la taille de l'échantillon atteint ou dépasse 30 éléments.
- Marge d'erreur : 2,03 x 0,77 $ = 1,56 $.
- Intervalle de confiance : entre [22,30 $ – 1,56 $] et [22,30 $ + 1,56 $], soit entre 20,74 $ et 23,86 $.

À titre d'information, notez que la véritable moyenne (que nous avons calculée à partir de l'ensemble de la population) se situe à 21,79 $ par abonnement. La figure 8.4 illustre la portée de notre dernière estimation. La moyenne véritable se trouve bien à l'intérieur de notre intervalle de confiance (il y avait seulement une chance sur 20 pour que cet intervalle ne contienne pas la moyenne), tout en étant quelque peu éloignée de la moyenne estimée à partir de notre échantillon (un écart de 0,51 $ les sépare).

FIGURE 8.4 - Moyenne estimée et moyenne véritable



EXERCICES 2

1. Les petits actuels sont-ils des ex-grands?

Il paraît que l'on rapetisse avec l'âge. D'autre part, on a observé que la taille moyenne des soldats américains augmente d'une génération à l'autre : 171 cm pendant la Première Guerre mondiale, 174 cm pendant la Deuxième, 175 cm pendant la Guerre du Vietnam et 176 cm pendant la Guerre du Golfe.

a) En vous servant des données du tableau 8.3, calculez l'intervalle de confiance pour les hommes âgés de 18 à 24 ans et pour les hommes âgés de 65 à 74 ans, avec un niveau de confiance de 0,95.

b) Même question avec un niveau de confiance de 0,99.

c) Commentez les écarts de taille entre les deux groupes d'âge.

2. Les abonnés absents

Nous avons tiré au hasard un nouvel échantillon de 12 abonnements de la société Publisher Clearing house (voir le tableau 8.4). Les prix de l'abonnement annuel sont respectivement les suivants : 29,97 $; 11,25 $; 21,97 $; 23,97 $; 21,95 $; 18,97 $; 22,97 $; 21,95 $; 19,67 $; 20,98 $; 24,97 $ et 27,97 $.

a) Calculez la moyenne du prix de l'abonnement.

b) Calculez l'écart type (d'échantillon).

c) Calculez l'erreur type.

d) Calculez l'intervalle de confiance, pour un niveau de confiance de 0,95.

3. VÉRIFIER UNE HYPOTHÈSE À L'AIDE D'UN ÉCHANTILLON

Dans la section précédente, nous avons montré dans quelle mesure un échantillon peut nous aider à mieux connaître la population dont il est issu. Nous avons utilisé la courbe normale pour estimer, avec une marge d'erreur plus ou moins grande, un paramètre d'une population comme, par exemple, la taille moyenne des Américains. Lorsque l'échantillon était plus petit, comme dans le cas du prix moyen des abonnements, nous avons remplacé la courbe normale par la courbe de Student.

Ici, la démarche est un peu différente, quoiqu'elle fasse appel aux mêmes outils. L'échantillon n'est plus utilisé pour estimer un paramètre, mais plutôt pour vérifier une hypothèse concernant la population. Nous nous demanderons par exemple comment un échantillon peut nous permettre d'affirmer avec un certain degré de certitude, que le QI des Asiatiques est plus élevé que la moyenne, ou que le volume crânien des criminels ne diffère pas de celui de la population en général.

Plus l'échantillon est petit, plus le hasard peut nous jouer des tours, mais il ne tient qu'à nous de garder le contrôle de la situation en évaluant le degré de risque encouru lorsque nous posons une hypothèse.

3.1. Un test unilatéral : le QI des Asiatiques

Le QI (quotient intellectuel) est un indicateur controversé. On ne sait pas exactement ce qu'il mesure : l'inné ou l'acquis? L'intelligence en général ou certaines aptitudes particulières choisies arbitrairement? Chose certaine, il est abusif d'évaluer l'intelligence d'un individu en se basant sur le QI du groupe auquel il appartient : ce n'est pas parce que le QI moyen de votre classe est faible cette année que vous êtes moins intelligent que d'habitude. Cette mise en garde étant faite, cherchons maintenant à satisfaire notre curiosité : les Asiatiques ont-ils, en moyenne, un QI supérieur à celui que l'on retrouve dans la population en général?

Une enquête qui porte sur la jeunesse américaine à travers les décennies (National Longitudinal Survey of Youth) nous fournit quelques chiffres. Un échantillon de 42 jeunes Américains originaires d'Extrême-Orient montre que ces derniers ont un QI moyen de 106 (contre 100 pour l'ensemble de la population, en 1990). La même enquête indique d'autre part que le QI est distribué de façon normale avec un écart type de 15. L'écart est-il suffisamment grand, compte tenu de la faible taille de l'échantillon, pour qu'on ne puisse l'attribuer au hasard?

Étapes du test d'hypothèse sur une moyenne

L'hypothèse selon laquelle la moyenne tirée de l'échantillon est la même que celle de la population s'appelle l'hypothèse nulle.

1. Formuler l'hypothèse. Généralement, l'hypothèse est formulée de façon positive : le QI moyen des Asiatiques (les gens de l'échantillon) est le même que le QI moyen de la population. Il faut aussi formuler une solution de rechange au cas où notre hypothèse serait rejetée : le QI moyen des Asiatiques est supérieur à celui de la population. L'hypothèse à vérifier est appelée hypothèse nulle. La solution de rechange, qui aurait pu prendre une autre forme, est appelée hypothèse alternative.

Le seuil de signification représente le risque de rejeter à tort l'hypothèse nulle.

2. Quel est le seuil de signification désiré? Autrement dit, quel est le risque de rejeter à tort une hypothèse vraie? Ici, nous pouvons par exemple décider de prendre le risque maximum de nous tromper 1 fois sur 20 (ou 0,05).

La valeur critique (il peut y en avoir deux) délimite la zone de rejet de l'hypothèse.

3. Trouver, à partir de la table de la distribution normale, le seuil à ne pas dépasser pour accepter l'hypothèse. Ici, le seuil, ou valeur critique est de +1,64 erreur type (voir la figure 8.5 ci-après). Dans la table normale, on cherchera la valeur correspondant à 0,45, ce qui nous permettra d'englober 0,05 dans la zone de rejet.

L'écart réduit représente le nombre d'erreurs types qui séparent la moyenne de l'échantillon de celle de la population.

4. Calculer l'écart réduit entre la moyenne estimée et la moyenne de la population. Pour cela, il faut diviser l'écart entre les deux moyennes (celle de l'échantillon moins celle de la population) par l'erreur type (dont la formule a été vue dans la section précédente).

Ici, Erreur type = Écart type/√(n) = 15/√(42) = 2,3 unités de QI. L'écart réduit est donc de (106 – 100)/2,3 = 2,61.

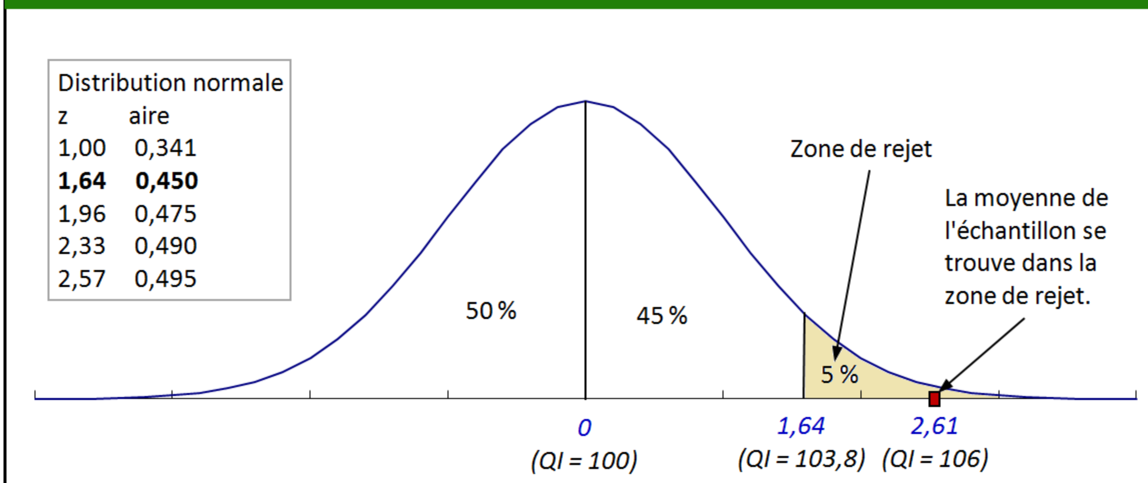
$$\textbf{Écart réduit (entre l'échantillon et la population)} = \frac{\text{Moyenne de l'échantillon} - \text{Moyenne de la population}}{\text{Erreur type}}$$

$$\textbf{Écart réduit (entre l'échantillon et la population)} = \frac{\text{Moyenne de l'échantillon} - \text{Moyenne de la population}}{\text{Écart type}/\sqrt{n}}$$

5. Accepter ou rejeter l'hypothèse après avoir comparé l'écart réduit (calculé à l'étape 4) à la valeur critique (calculée à l'étape 3). Ici, nous sommes dans la zone de rejet, puisque l'écart réduit (2,61) dépasse la valeur critique (1,64). Cet écart est assez grand pour que nous puissions rejeter l'hypothèse nulle (« le QI moyen des Asiatiques est le même que celui de la population ») et accepter l'hypothèse de rechange (« le QI moyen des Asiatiques est supérieur à celui de la population »).

En acceptant l'hypothèse alternative, nous courons encore le risque de nous tromper. Cependant, ce risque est inférieur aux limites que nous nous sommes fixées à l'étape 2 (soit 0,05 ou 5 %). La figure 8.5 illustre ce résultat.

FIGURE 8.5 - Test unilatéral



Lorsque la taille de l'échantillon est inférieure à 30, il est toujours possible d'effectuer un test d'hypothèse, à condition toutefois que la population soit distribuée de façon normale. On remplacera alors, dans l'étape 3, la table de distribution normale par la table de distribution de Student, comme nous l'avons fait dans la section précédente pour l'estimation de moyenne.

3.2. Un test bilatéral : les crânes de Lombroso

Encore un sujet délicat...

Non, il ne s'agit pas d'une nouvelle marque de pâtes. À la fin du XIX[e] siècle, Cesare Lombroso, dans le cadre d'une vaste étude, a mesuré la capacité crânienne de 121 criminels pour voir si elle différait de ce qu'on pouvait observer dans la population en général. Même si certaines de ses conclusions ont été remises en question par la suite, Lombroso peut être considéré comme un des fondateurs de la criminologie moderne.

La capacité crânienne des 121 têtes criminelles était en moyenne de 1450 cm^3, tandis que celle de la population était de 1475 cm^3 avec un écart type de 110 cm^3.

Lors d'un test d'hypothèse bilatéral, il y a deux zones de rejets. Cela signifie que la moyenne de l'échantillon peut aussi bien être inférieure que supérieure à la moyenne de la population.

1. Formuler l'hypothèse nulle : « la capacité crânienne moyenne des criminels est la même que celle de la population ». Hypothèse alternative : « la capacité crânienne moyenne des criminels est différente de celle de la population ». Comparez cette hypothèse alternative avec celle que nous avions formulée dans l'exemple précédent. Nous voulions savoir si le QI moyen des Asiatiques était *supérieur* (ou bien égal) à celui de la population, car rien ne nous laissait croire que ce QI puisse être inférieur. Cette fois, nous n'avons pas d'idée préconçue sur le sujet. Si jamais la capacité crânienne de criminels devait être différente de celle de la population, rien ne nous laisse présupposer qu'elle soit supérieure ou inférieure. Il faudra alors établir deux valeurs critiques (une valeur minimum et une valeur maximum), d'où le nom de test bilatéral.

2. Seuil de signification : nous décidons de prendre moins de risque que dans l'exemple précédent, étant donnée la gravité de la question, aussi nous choisissons un seuil de signification de 0,01

3. Valeurs critiques : puisque le test est bilatéral, il y a deux zones de rejets (une à gauche du minimum et une à droite du maximum) totalisant en tout 0,005 + 0,005 = 0,01 des possibilités. Les valeurs critiques correspondantes sont de –2,57 et +2,57 erreurs types (voir la figure 8.6 ci-après).

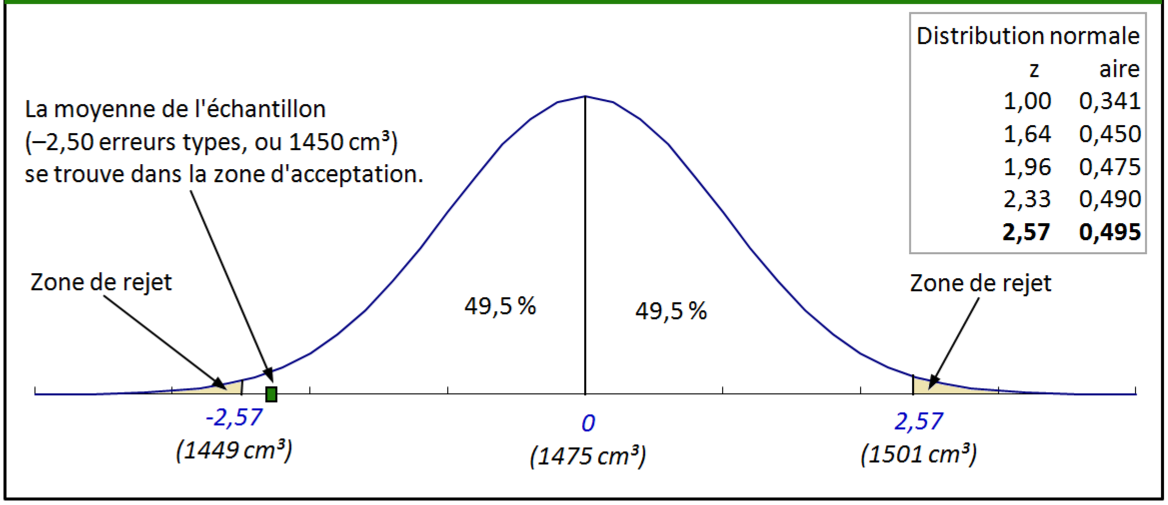
4. Erreur type de l'échantillon = 110/√(121) = 110/11 = 10 cm^3.

Écart réduit = (1450 – 1475)/10 = –25/10 = –2,5.

On pourrait aussi convertir les valeurs critiques dans l'unité de mesure de la variable, on obtient respectivement de 1475 – (2,57 x 10) et 1475 + (2,57 x 10) : la zone d'acceptation est comprise entre 1449 et 1501 cm^3. Dans ce dernier calcul, 1475 représente la moyenne, 2,57 la valeur critique trouvée dans la table normale et 10 l'erreur type.

5. Décision : l'écart réduit se trouve dans la zone de non-rejet. En d'autres mots, la capacité crânienne des criminels se trouve à l'intérieur des valeurs critiques, quoique de justesse. Jusqu'à nouvel ordre, et tant que nous n'aurons pas de données plus convaincantes, nous acceptons l'hypothèse nulle (« la capacité crânienne moyenne des criminels est la même que celle de la population »).

FIGURE 8.6 - Test bilatéral



EXERCICES 3

1. Encore le QI

Selon l'enquête longitudinale sur la jeunesse américaine de 1990 dont il a été question dans cette section, le QI moyen des Américains de confession juive serait de 114,5. Ce chiffre a été obtenu à partir d'un échantillon de 99 personnes. On sait d'autre part que l'écart type de ce QI est de 15 pour l'ensemble de la population américaine.

Formulez une hypothèse et testez-la.

2. Encore des crânes

Une enquête d'Ernest Hooton basée sur un très large échantillon (1672 individus) indique que la circonférence de la tête des Bostoniens blancs est en moyenne de 563 mm avec un écart type de 10 mm (Ernest Hooton, The American Criminal, 1939). Un petit échantillon de 25 Bostoniens membres d'une profession libérale révèlent que ces derniers ont une circonférence moyenne de tête de 569,9 mm.

a) L'hypothèse nulle est la suivante : « la circonférence de la tête des membres de l'échantillon observé ne diffère pas de façon significative de celle de la population bostonienne en général ». Compte tenu de cette formulation, testez l'hypothèse nulle avec un seuil de signification de 0,05.

b) Même question avec un échantillon de 194 commerçants dont la circonférence moyenne de tête est de 565,7.

4. TEST D'HYPOTHÈSE SUR DEUX MOYENNES

Dans la section précédente, nous voulions déterminer si un groupe d'individus particulier avait des caractéristiques différentes de celles de l'ensemble de la population. C'est pourquoi nous avons comparé la moyenne obtenue grâce à un échantillon à la moyenne de la population. Cette fois, nous allons comparer directement deux échantillons issus de deux populations apparentées. Nous vérifierons par exemple si les Américains de race noire sont plus grands (ou moins grands) que leurs compatriotes de race blanche, ou si l'estime de soi est plus élevée au Québec chez les garçons que chez les filles.

Nous utiliserons encore une fois des outils familiers : la courbe normale (lorsque chacun des deux échantillons contient au moins 30 éléments) et la courbe de Student (dans la situation contraire). Dans ce dernier cas, il faut aussi que la population d'où sont extraits les échantillons soit distribuée sur le modèle de la courbe normale.

4.1. Qui est le plus grand?

Selon l'enquête menée aux États-Unis dont nous avons parlé un peu plus haut (voir tableau 8.3), les hommes de race blanche de ce pays mesurent en moyenne 175,8 cm contre 175,5 cm pour les hommes de race noire. L'écart entre ces deux moyennes est-il significatif? Même si la notion de race est plus ou moins arbitraire*, ce genre de données peut avoir une réelle utilité pour un concepteur de vêtement ou pour un chercheur qui étudie les conséquences de certaines habitudes alimentaires. En sciences humaines, il ne faut jamais avoir peur de mesurer. Ce qui compte, c'est de traiter les chiffres avec rigueur et honnêteté.

Avant 1980, un enfant n'était classifié officiellement comme Blanc par le gouvernement américain que si ses deux parents étaient Blancs. Lorsque l'un des parents était Noir, l'enfant était considéré comme Noir (sauf si l'autre parent était Hawaiien!). Aujourd'hui, l'enfant est classifié en fonction de la race de la mère. Pour compliquer les choses, les résidents originaires de l'Amérique latine, qui n'ont pas cette obsession maladive de voir le monde en noir ou en blanc sont regroupés dans une catégorie à part.

Où l'on constate qu'il est facile de dire des âneries.

Il ne nous a pas été difficile de vérifier, à partir d'échantillons, que la taille moyenne des Scandinaves dépasse celle des Philippins, que celle des hommes dépasse celle des femmes et que celle des jeunes gens de la génération actuelle dépasse celle de leur parent. Ces faits établis, il est facile de les interpréter de façon erronée comme dans les exemples suivants : « les Scandinaves sont supérieurs aux Philippins », « tu es une femme, je suis un homme, donc je suis plus grand que toi ». On ne peut se baser sur les caractéristiques *moyennes* d'une population pour déduire les caractéristiques d'un *individu* en particulier. Ce serait faire preuve de bêtise encore plus que de racisme. D'autre part, il ne faut pas confondre *différent* avec *inférieur*. La première notion est vérifiable alors que la seconde relève d'un jugement de valeur et non de la science. Enfin, si tout le monde s'entend sur le concept de taille et sur la longueur des centimètres, il en va autrement lorsqu'on mesure l'estime de soi, le sentiment d'échec ou le quotient intellectuel. Dans ces derniers cas, il faut alors redoubler de prudence, même si les enquêtes peuvent fournir des résultats intéressants.

Étapes du test d’hypothèse sur deux moyennes

1. Formuler l’hypothèse de la façon suivante : la moyenne du premier échantillon est égale à la moyenne du second échantillon (c’est l’*hypothèse nulle*). Formuler aussi une *hypothèse alternative* qui sera retenue si on rejette l’hypothèse nulle. L’hypothèse alternative peut être formulée de trois façons selon les circonstances (c’est au chercheur de choisir) :
– Moyenne 1 ≠ Moyenne 2 (test bilatéral)
– Moyenne 1 > Moyenne 2 (test unilatéral)
– Moyenne 1 < Moyenne 2 (test unilatéral).

2. Choisir un seuil de signification. Le hasard peut faire en sorte que même si deux populations possèdent la même caractéristique moyenne, les deux échantillons tirés soient si différents qu’ils nous font rejeter l’hypothèse nulle. Quel est le risque, que nous sommes prêts à prendre, de rejeter ainsi à tort l’hypothèse nulle?

3. Trouver, à partir de la table de distribution normale, la valeur critique (ou les valeurs critiques)qui nous permettra d’accepter ou de rejeter l’hypothèse nulle. Pour cela, il faut tenir compte de l’hypothèse alternative : le test est-il unilatéral ou bilatéral? (On note que ces trois premières étapes sont les mêmes que pour le test d’hypothèse sur une moyenne vu dans la section précédente.)

4. Calculer l’écart entre les deux moyennes (ou *écart réduit*) de la façon suivante :

$$\textbf{Écart réduit (entre deux échantillons)} = \frac{\text{Moyenne 1} - \text{Moyenne 2}}{\sqrt{\left(\frac{[\text{Écart type 1}]^2}{n1}\right) + \left(\frac{[\text{Écart type 2}]^2}{n2}\right)}}$$

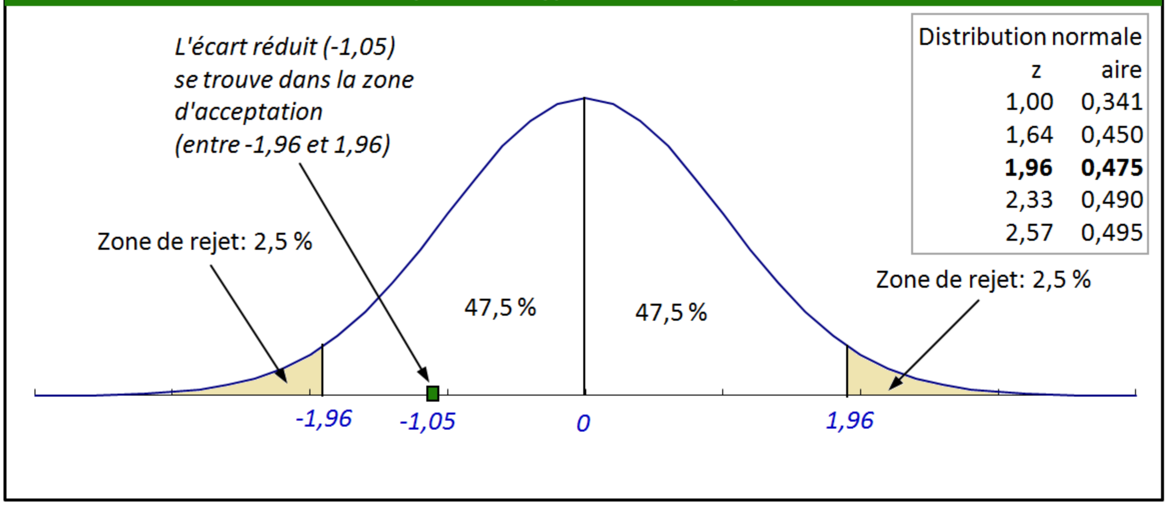
5. Décision : accepter l’hypothèse nulle si l’écart réduit que l’on vient de calculer ne déborde pas de la valeur critique (ou des valeurs critiques) définie à l’étape 3.

Application de la démarche à notre exemple

- 1. Hypothèse nulle : « La taille moyenne des Noirs est la même que celle des Blancs ». Hypothèse alternative : « La taille moyenne des Noirs est *différente* de celle des Blancs. »
- 2. Nous choisissons un seuil de signification de 0,05.
- 3. Il s’agit d’un test bilatéral. Les valeurs critiques pour un seuil de signification de 0,05 sont de –1,96 et +1,96. L’hypothèse nulle sera rejetée si l’écart réduit n’est pas compris à l’intérieur de ces limites.
- 4. Écart réduit :
  nous reproduisons dans le tableau 8.5 ci-après les valeurs pertinentes du tableau 8.3 accompagnées de quelques données dérivées.
  Numérateur de l’écart réduit : Moyenne1 – Moyenne2 = 175,5 – 175,8 = –0,3
  Dénominateur de l’écart réduit :
  $\sqrt{[(\text{Écart type1}^2/n1) + (\text{Écart type2}^2/n2)]} = \sqrt{[0{,}072 + 0{,}010]} = \sqrt{[0{,}082]} = 0{,}286$
  Écart réduit = –0,3/0,286 = –1,05
- 5. Décision : l’écart entre les deux moyennes est trop faible pour que nous rejetions l’hypothèse nulle. Jusqu’à preuve du contraire, nous considérons que la taille moyenne des Noirs est égale à la taille moyenne des Blancs (voir la figure 8.7).

Tableau 8.5 - Taille des Américains (données pertinentes au calcul de l'écart réduit)						
		n	**Moyenne**	**Écart type**	**Écart type²**	**Écart type²/n**
Variable 1	Taille des Noirs	649	175,51	6,86	47,03	0,072
Variable 2	Taille des Blancs	5148	175,77	7,11	50,58	0,010

FIGURE 8.7 - Test sur deux moyennes : hypothèse non rejetée



4.2. L'estime de soi chez les garçons et les filles

Le corps humain se transforme rapidement lors de la puberté. La taille augmente de 25 centimètres environ et le poids de 11 kilogrammes. On change de voix, on change de visage, on se demande parfois si on n'est pas dans le corps de quelqu'un d'autre. Les garçons gagnent surtout en muscles et en os, tandis que les filles gagnent plutôt en graisses. Ce changement d'image peut parfois diminuer ou accroître l'estime de soi, selon qu'il répond ou non aux stéréotypes véhiculés par la société. D'après une enquête menée au Québec*, il semble, à première vue, que ce soit les filles qui sont le plus désavantagées à cet égard (voir le tableau 8.6).

Tableau 8.6 - L'enquête sur les adolescents au Québec

Score d'estime de soi (échelle de 10 à 40)

	Taille de l'échantillon (n)	**Score moyen**	**Écart type**	**Écart type²**	**Écart type²/n**
Garçons	153	32,77	4,29	18,40	0,120
Filles	156	30,49	4,93	24,30	0,156

Note : Un score élevé indique une forte estime de soi.

Score de croyances irrationnelles (échelle de 50 à 300)

	Taille de l'échantillon (n)	**Score moyen**	**Écart type**
Garçons	153	150,89	25,36
Filles	156	148,76	27,84

Note : Un score élevé indique un faible niveau de croyances irrationnelles.

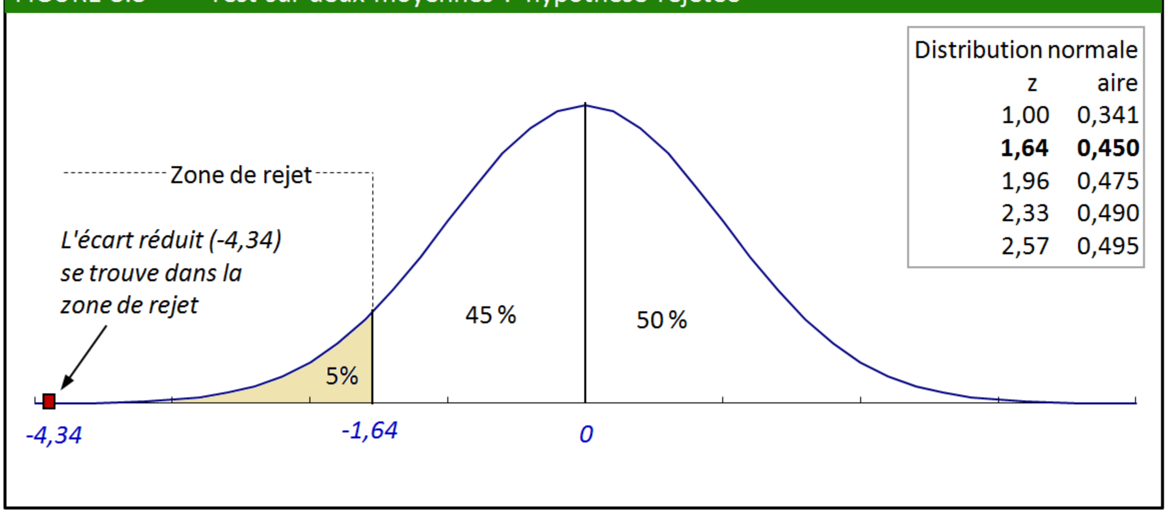
Source : Diane Marcotte, Revue québécoise de psychologie, vol. 16, nº 3, 1995.

L'estime de soi a été mesurée à partir d'un questionnaire contenant 10 questions. Les 309 adolescents choisis, tous âgés de 11 à 18 ans, devaient s'autoévaluer sur une échelle ordinale de 1 à 4. Les scores pouvaient donc varier de 10 (faible estime de soi) à 40 (forte estime de soi). En prime, nous avons ajouté sur le tableau le score du test de croyances irrationnelles contenant 50 questions basées sur une échelle ordinale de 1 à 6 (plus le score est bas, plus les croyances irrationnelles sont grandes).

Le niveau d'estime de soi est-il le même chez les deux sexes? Essayons de déterminer si l'écart entre le score moyen des garçons et des filles est suffisant, compte tenu de la taille de l'échantillon, pour nous permettre de rejeter cette hypothèse.

- 1. Hypothèse nulle : « Le niveau d'estime de soi est le même chez les deux sexes ». Hypothèse alternative : « le niveau d'estime de soi est plus faible chez les filles ».
- 2. Nous choisissons un seuil de signification de 0,95.
- 3. Nous avons affaire à un test unilatéral. La valeur critique, pour un seuil de signification de 0,45, est de 1,64 dans la table normale. Nous rejetterons donc l'hypothèse nulle si l'écart réduit que nous calculerons à l'étape 4 est inférieur à –1,64 (voir la figure 8.8).
- 4. Écart réduit = (30,49 – 32,77)/√[(4,93)^2/156 + (4,29)^2/153] = –2,28/0,53 = –4,34.
- 5. Décision : l'écart réduit (–4,34) est en deçà de la valeur critique (–1,64). Nous rejetons l'hypothèse nulle.

FIGURE 8.8 - Test sur deux moyennes : hypothèse rejetée



EXERCICES 4

1. La pensée cartésienne? Moi, je n'y crois pas du tout.

En vous basant sur les données du tableau 8.6, testez l'hypothèse suivante : le niveau de croyances irrationnelles est le même chez les filles que chez les garçons.

EXERCICES SUPPLÉMENTAIRES

1. Un peu de vocabulaire

a) Identifiez dans la phrase suivante : le niveau de confiance, l'intervalle de confiance et la marge d'erreur.

« Dans 95 % des cas, la proportion véritable des gens qui trouvent les tribunaux trop peu sévères se situe entre [79 % – 2,5 %] et [79 % + 2,5 %] (ou entre 76,5 % et 81,5 %). Pour faire plus journaliste, on dira aussi que, 19 fois sur 20, le résultat peut varier au maximum de plus ou moins 2,5 %. »

b) Identifiez, dans la phrase suivante : le seuil de signification, les valeurs critiques. Formulez l'hypothèse nulle.

« Étant donné que l'écart réduit entre les deux moyennes estimées dépasse 1,96 nous acceptons, avec une probabilité de nous tromper inférieure à 5 %, l'hypothèse alternative qui veut que le QI moyen des Quadrifluviens de l'Est soit différent de celui des Quadrifluviens de l'Ouest. »

2. Des réfugiées politiques

Selon une étude publiée dans *Santé mentale au Québec* (vol. 21, no 1, printemps 1996, par Marta Yong), un échantillon de 94 Somaliennes habitant la région d'Ottawa possède les caractéristiques suivantes :

- L'âge des membres de l'échantillon varie de 18 à 50 ans avec une moyenne de 32,72 ans et un écart type de 8,27 ans;
- Il y a 59 membres de l'échantillon qui sont mariées et 15 qui sont célibataires;
- Il y a 44 % des membres de l'échantillon qui ont terminé leurs études secondaires et 29 % qui ont terminé des études universitaires (apparemment, il s'agit de deux catégories exclusives);
- Il y a 49 % des membres de l'échantillon qui sont des résidentes permanentes et 25 % qui sont des citoyennes canadiennes;
- Il y a 82 % des membres de l'échantillon qui ont immigré pour des raisons politiques et 13 % pour des raisons familiales.

(On suppose que l'échantillon est aléatoire.)

a) Identifiez les cinq variables dont il vient d'être question, indiquez leur échelle de mesure, et indiquez leurs catégories s'il s'agit de variables qualitatives.

b) Faites une estimation de l'âge moyen, pour un niveau de confiance de 0,95.

c) Calculez la marge d'erreur des proportions, pour un niveau de confiance de 0,95.

3. Test controversé sur des crânes

Le tableau 8.7 ci-après reproduit les données d'Ernest Hooton dont il a été question dans ce chapitre.

a) Testez l'hypothèse selon laquelle les manœuvres ont la même circonférence moyenne de tête que les membres de la population en général. (Utilisez un seuil de signification de 0,05.)

b) Testez l'hypothèse selon laquelle la circonférence moyenne de la tête des cadres est la même que celle des manœuvres. (Utilisez un seuil de signification de 0,05.)

Tableau 8.7 - Les crânes des Bostoniens selon Ernest Hooton

	Nombre (n)	**Circonférence de la tête**	**Écart type** (en mm)	**Erreur type**
Professions libérales	25	569,9	9,5	1,9
Cadres	61	566,5	11,7	1,5
Employés	107	566,2	11,4	1,1
Commerçants	194	565,7	11,1	0,8
Fonctionnaires	25	564,1	12,5	2,5
Ouvriers qualifiés	351	562,9	11,2	0,6
Gens de maison	262	562,7	11,3	0,7
Manœuvres	647	560,7	7,6	0,3
Total	1672			
Ensemble de la population		**562,8**	**10**	

Source : Ernest HOOTON, The Americain Criminal, Harvard University Press, 1939, dans Stephen Jay GOULD, La mal-mesure de l'homme, Livre de Poche, 1986.

Note : on trouvera dans l'ouvrage de Gould une bonne critique de ce genre de mesure.

4. Laboratoire

a) À l'aide d'un chiffrier électronique, construisez un tableau vous permettant de calculer automatiquement l'erreur type et le coefficient de variation à partir des données pertinentes (vous pouvez vous inspirer du tableau 8.2b).

5. Recherche : une taxe bien visible

Au Canada, les taxes de vente sont rarement incluses dans les prix affichés. Les porte-parole gouvernementaux prétendent que cette manière d'agir favorise la transparence en permettant aux consommateurs-contribuables de « voir » les taxes. Vous devez faire une mini-enquête pour vérifier, dans la mesure de vos moyens, si ce point correspond à la réalité.

Sélectionnez un échantillon de 30 à 50 personnes, en faisant en sorte, si possible, qu'il soit relativement représentatif. Posez à ces personnes des questions pour vérifier:

1) s'ils connaissent le taux de la taxe de vente fédérale (TPS) et de la taxe provinciale pour divers produits (vêtements, aliments, bière, livres, etc.);

2) s'ils ont tendance, de façon *spontanée*, à sous-estimer la taxe lorsque les prix sont hauts et à la surestimer lorsque les prix sont bas (ou l'inverse);

c) s'ils trouvent que la façon d'afficher les prix sans la taxe favorise la transparence.

Identifiez les variables correspondant à chaque question et les échelles de valeurs qui leur sont associées. Compilez les résultats. Calculez la marge d'erreur des proportions (en faisant la supposition que votre échantillon est aléatoire). Si vous le pouvez, faites des hypothèses et vérifiez-les. Commentez les résultats.

6. Recherche : la taille des gens de votre entourage

Dans cette mini-enquête, vous allez essayer d'estimer la taille moyenne des gens de votre entourage afin de la comparer à celle des Canadiens en général (176,2 cm pour les hommes avec un écart type de 7 cm, et 161,9 cm pour les femmes avec un écart type de 6 cm).

a) Définissez la population dont vous voulez estimer la taille : lieu (quartier, région, institution), sexe, âge (adultes, enfants de 12 ans), etc.

b) Sélectionnez un échantillon d'au moins 30 personnes appartenant à la population que vous avez définie, en faisant en sorte, dans la mesure du possible, que l'échantillon soit représentatif.

c) Faites une estimation de la taille moyenne de votre population et calculez l'intervalle de confiance correspondant, pour un niveau de confiance de 0,95.

d) Testez l'hypothèse suivante : « la taille moyenne de la population observée est la même que celle de la population canadienne en général ».

DOSSIER 8 DES ÉLECTIONS HISTORIQUES

Le 26 octobre 1993, le Bloc québécois, parti voué à l'indépendance du Québec, fait une percée spectaculaire aux élections fédérales canadiennes. Bien qu'il ne soit implanté qu'au Québec, le Bloc réussit même à devenir l'opposition officielle à Ottawa. Le Parti conservateur, premier aux élections précédentes, subit alors la déroute au Québec : il n'obtient que 12,9 % des voix et ne fait élire qu'un seul député.

Les résultats

Les tableaux ci-dessous montrent comment le vote s'est distribué au Québec. On y constate notamment que 3 720 312 citoyens et citoyennes ont exercé leur droit de vote (tableau D8.1a), que 49,2 % d'entre eux ont choisi le Bloc québécois (tableau D8.1b) et que 25,9 % d'entre eux proviennent de l'île de Montréal (tableau D8.1c).

Tableau D8.1 - Élections fédérales du 26 octobre 1993 au Québec

Tableau D8.1a - Résultats bruts (nombre de votes)

	Bloc québécois	Libéral	Autres	Total
Île de Montréal	347 977	484 399	132 961	**965 337**
Reste du Québec	1 484 090	767 992	502 893	**2 754 975**
Total	**1 832 067**	**1 252 391**	**635 854**	**3 720 312**

Tableau D8.1b - Pour quel parti la région a-t-elle voté? (en % de la région)

	Bloc québécois	Libéral	Autres	Total
Île de Montréal	36,0	50,2	13,8	**100**
Reste du Québec	53,9	27,9	18,3	**100**
Total	**49**	**34**	**17**	

Tableau D8.1ac - De quelle région proviennent les votes obtenus? (en % du parti)

	Bloc québécois	Libéral	Autres	Total
Île de Montréal	19,0	38,7	20,9	**26**
Reste du Québec	81,0	61,3	79,1	**74**
Total	**100**	**100**	**100**	

Source : Le Devoir, mercredi 27 octobre 1993.

Tableau D8.1a (données brutes)

Le premier tableau contient les résultats bruts pour l'ensemble du Québec. S'il donne les grandes lignes, on ne peut en tirer de résultats bien précis.

Tableau D8.1b (pourcentages en lignes)

Dans l'île de Montréal, le Parti libéral obtient la moitié des votes (50,2 %) et devance largement le Bloc québécois. En province, par contre, le Bloc obtient environ deux fois plus de votes que le Parti libéral (53,9 % contre 27,9 %). Au total (dans l'ensemble du Québec), le Bloc obtient 49,2 % de toutes les voix.

Tableau D8.1c (pourcentages en colonnes)

Le Parti libéral obtient une partie appréciable de ses voix dans l'île de Montréal : 38,7 % de tous les électeurs libéraux proviennent de la métropole (qui ne compte pourtant que 25,9 % des électeurs du Québec). Inversement, le Bloc va chercher des appuis relativement plus solides en région : 81 % de ses voix en proviennent, ce qui est nettement supérieur à la proportion totale d'électeurs qui résident en région (74,1 %).

On constate également que, pour la plupart des observateurs, c'est le tableau D8.2 qui fournit l'information la plus intéressante. Revenons donc quelques instants sur une donnée de ce tableau : le pourcentage de 49,2 % obtenu par le Bloc québécois dans l'ensemble du Québec. Ce pourcentage représente la moyenne entre les résultats à Montréal (36,0 %) et les résultats en région (53,9 %). Il s'agit bien sûr d'une moyenne *pondérée*, en ce sens que le poids de Montréal n'est pas identique au poids du reste du Québec. La moyenne pondérée se calcule de la manière suivante :

$$[36 \times 25{,}9\ \%] + [53{,}9 \times 74{,}1\ \%] = 49{,}3\ \%$$

[Score de Montréal x Poids de Montréal] + [Score des régions x poids des régions] = Score national

Le même chiffre peut être obtenu directement à partir des chiffres bruts du tableau D8.1a (comme tous les chiffres des tableaux D8.1b et D8.1c, d'ailleurs) :

$$1\ 832\ 067/3\ 720\ 312 = 0{,}492 = 49{,}2\ \%.$$

Note : nos chiffres étant arrondis, on constate parfois un écart d'un dixième.

L'*île de Montréal* regroupe les 10 comtés de Montréal-Est et les 10 comtés de Montréal-Ouest. Les 55 comtés restants sont groupés sous la catégorie *reste du Québec*. Le tableau D8.2 illustre de façon plus détaillée les divergences régionales.

Tableau D8.2 - Élections fédérales du 26 octobre 1993 au Québec Résultats par région

	(Nombre de votes)						(Proportions en %)				
	BQ	Libéral	PC	NPD	Autres	**Total**	BQ	Libéral	PC	NPD	Autres
Québec	134 001	58 648	30 868	4 274	6 761	**234 552**	57,1	25,0	13,2	1,8	2,9
Québec-Métro	157 955	62 789	55 964	4 667	25 507	**306 882**	51,5	20,5	18,2	1,5	8,3
Outaouais	77 173	98 853	28 260	4 005	7 760	**216 051**	35,7	45,8	13,1	1,9	3,6
Estrie	149 356	100 411	47 805	3 476	6 064	**307 112**	48,6	32,7	15,6	1,1	2,0
Mauricie	102 198	61 776	30 280	1 741	1 979	**197 974**	51,6	31,2	15,3	0,9	1,0
Rive-Sud	325 975	159 784	61 919	8 589	5 017	**561 284**	58,1	28,5	11,0	1,5	0,9
Laurentides-Lanaudières	185 699	62 513	37 154	3 350	3 825	**292 541**	63,5	21,4	12,7	1,1	1,3
Nord-Ouest	41 431	15 912	22 457	951	830	**81 581**	50,8	19,5	27,5	1,2	1,0
Bas-Saint-Laurent-Gaspé-Côte-Nord	120 784	62 520	39 368	2 612	2 405	**227 689**	53,0	27,5	17,3	1,1	1,1
Saguenay-Lac-Saint-Jean	100 531	21 183	26 583	1 880	435	**150 612**	66,7	14,1	17,6	1,2	0,3
Montréal-Est	222 604	179 797	35 566	8 857	20 456	**467 280**	47,6	38,5	7,6	1,9	4,4
Monréal-Ouest	125 373	304 602	43 743	11 123	13 216	**498 057**	25,2	61,2	8,8	2,2	2,7
Laval	88 987	63 603	19 792	1 954	4 361	**178 697**	49,8	35,6	11,1	1,1	2,4
Ensemble du Québec	**1 832 067**	**1 252 391**	**479 759**	**57 479**	**98 616**	**3 720 312**	**49,2**	**33,7**	**12,9**	**1,5**	**2,7**

Source : Le Devoir, mercredi 27 octobre 1993.

Que prédisaient les sondages?

Nous reproduisons ici les trois derniers sondages parus dans la presse québécoise avant les élections. Les deux premiers (tableau D8.3) sont des sondages pancanadiens, aussi les marges d'erreur sont-elles assez élevées au niveau régional. C'est pourquoi nous nous attarderons plus spécialement sur le sondage SOM-La Presse-TVA (tableau D8.4) dont l'échantillon est le plus grand à l'échelle du Québec.

Tableau D8.3 - sondages avant les élections fédérales du 26 octobre 1993

Angus-Reid (18-20 octobre)

		Canada	**Québec**
			(en %)
Bloc Québécois	BQ	14	52
Libéral	PLC	43	30
Conservateur	PC	17	13
Néo-démocrate	NPD	8	2
Réformiste	RP	17	
Autres		1	3
		100	100
Marge d'erreur 19 fois sur 20 (en %)		2,5	4
Taille de l'échantillon		3 329	597

Source : Le Devoir, 23 octobre 1993.

Gallup (17-20 octobre)

		Canada	**Québec**
		(en %)	
Bloc Québécois	BQ	12	50
Libéral	PLC	44	31
Conservateur	PC	16	14
Néo-démocrate	NPD	7	
Réformiste	RP	19	
Autres		2	5
		100	100
Marge d'erreur 19 fois sur 20 (en %)		3,1	
Taille de l'échantillon		1 011	

Source : La Presse, 22 octobre 1993.

Tableau D8.4 - Répartition des indécis dans les sondages (au Québec)

Sondage SOM - La Presse - TVA (17-20 octobre 1993)						*26 octobre*
(en %, sauf exception)		**Résultats bruts**	**Répartition proportionnelle**	**Répartition corrigée SOM**	**Répartition corrigée 33/66**	***Élections***
		[1]	[2]	[3]	[4]	[5]
Bloc Québécois	BQ	36	56	45	48	*49,2*
Libéral	PLC	18	28	29	33	*33,7*
Conservateur	PC	8	12	20	15	*12,9*
Néo-démocrate	NPD	1	2	3	2	*1,5*
Autres	RP	1	2	3	2	*2,7*
Indécis		36				
Total		**100**	**100**	**100**	**100**	***100***
Ne voteront pas	5					
Refus de répondre	5					
Marge d'erreur 19 fois sur 20 (en %)			3,15			
Taille de l'échantillon			1 023			

Source : La Presse, 22 octobre 1993.

La marge d'erreur dépend de la taille de l'échantillon.

Tout d'abord, vérifions la marge d'erreur publiée. On a rejoint 1023 personnes dont 5 % (soit 1023 x 0,05 = 51) ont refusé de répondre. En tout, 972 personnes (1023 – 51) ont donc répondu. En principe, il existe une marge d'erreur pour chacune des proportions, mais nous la calculerons, comme à l'accoutumée pour une proportion type de 50 %.

$$\text{Erreur type} = \sqrt{[0{,}5 \times 0{,}5/\sqrt{972}]} = 0{,}5/31{,}18 = 0{,}016 = 1{,}6\ \%.$$

$$\text{Marge d'erreur (pour un niveau de confiance de 95 \%)} = 1{,}96 \times 1{,}6\ \% = 3{,}14\ \%$$

Les indécis, bêtes noires des sondeurs.

On constate dans le tableau D8.4 que plus du tiers (36 %) des électeurs sont encore indécis, quelques jours avant le vote. Étant donné que les indécis sont, par définition, des gens différents des autres, leur présence nuit à la représentativité de l'échantillon.

Dans ce sens, la méthode qui consiste à répartir les indécis de façon proportionnelle, comme on le fait dans la colonne 2, est sans doute la moins appropriée. Voici tout de même comment s'est faite cette répartition. Étant donné que 64 % seulement des répondants se sont prononcés, il reste 36 % d'indécis à distribuer. Le Parti libéral, par exemple, ayant obtenu 18 % des intentions de vote sur les 64 % de gens qui se sont affichés, on lui alloue 18/64 des indécis, soit 18/64 x 36 = 10 points de pourcentage. Le Parti libéral passe de 18 % (colonne 1) à 28 % (colonne 2).

Les sondeurs de SOM ont préféré utiliser une autre méthode (colonne 3). Pour mieux répartir les indécis, ils demandaient à ces derniers d'évaluer, sur une échelle de 1 à 10, la probabilité d'aller voter pour chaque parti. Cette méthode a cependant ses limites : si le sondé ne sait pas lui-même ce qu'il veut, toutes les astuces pour lui tirer les vers du nez deviennent inutiles.

La troisième méthode consiste à utiliser l'expérience des élections précédentes (colonne 4). Comme il s'agit de choisir entre deux options fondamentales et non entre le parti Coke et le parti Pepsi, on peut sûrement tirer une leçon des votes passés où l'enjeu était similaire. On estime alors, pour l'avoir déjà observé, que les indécis sont plus craintifs ou conservateurs que la moyenne. En conséquence, on distribue les indécis pour deux tiers dans les partis du statu quo (soit 2/3 x 36 % = 24 %) et pour un tiers dans le parti du changement (soit 1/3 x 36 % = 12 %).

En fin de compte, le sondage demeure un bon outil pour faire des pronostics, à condition de tenir compte des faits suivants :

- 1) Le sondage constitue une photographie de l'opinion une semaine avant le vote : l'opinion peut avoir évolué après le sondage.
- 2) L'existence d'un nombre élevé d'indécis biaise les résultats en minant la représentativité du sondage.
- 3) Il y a toujours un risque (modeste et mesurable) de tomber, par hasard, sur un échantillon peu représentatif.

QUESTIONS

1. Les perdants

a) Sachant que les conservateurs ont obtenu 8,2 % des voix dans l'île de Montréal et 14,5 % dans le reste du Québec, quel est le pourcentage de votes obtenus par les conservateurs dans l'ensemble du Québec?

b) Quel est le nombre de voix obtenues par les conservateurs dans l'ensemble du Québec?

2. Les mensonges

Corrigez les fausses affirmations ci-dessous.

a) 38,7 % des Montréalais ont voté pour le Parti libéral (tableau D8.1c).

b) 50,2 % des électeurs libéraux proviennent de l'île de Montréal (tableau D8.1b)

3. Les vainqueurs

a) Identifiez les régions du Québec où le Parti libéral est arrivé en tête (tableau D8.2)

b) Identifiez les régions du Québec où le Bloc québécois obtient une majorité de voix (tableau D8.2).

c) Quelle est la région où chacun des deux principaux partis obtient son meilleur score (tableau D8.2)?

d) Vérifiez, à partir des résultats bruts (tableau D8.1a), que les libéraux ont bien obtenu 50,2 % des votes de l'île de Montréal (tableau D8.1b) et que 81 % des votes du Bloc proviennent du reste des régions (tableau D8.1c).

4. Les marges d'erreur

Vérifiez si les marges d'erreur publiées avec les sondages Angus-Reid et Gallup s'avèrent correctes (tableaux D8.3a et b).

Note : utilisez un niveau de confiance de 95 % et une proportion type de 50 %.

CHAPITRE 9 LES RELATIONS ENTRE VARIABLES

TABLE DES MATIÈRES

1. Schéma de relations
2. Relation entre deux variables qualitatives
3. Relation entre deux variables quantitatives
4. Les accidents de la route

• Exercices supplémentaires

• Dossier

Lorsque deux variables « varient » de concert, on peut affirmer qu'elles sont reliées. Mais quelle est la nature réelle de leur relation? Est-ce une relation de cause à effet? Autrement dit, est-ce que l'une des variables (la *variable dépendante*) dépend de l'autre (la *variable indépendante*) et si oui, laquelle, pourquoi et comment? Ou bien sommes-nous en présence d'une simple corrélation? Dans ce cas, se pourrait-il que les deux variables en jeu dépendent d'une troisième variable?

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

- Comment, en examinant les chiffres, peut-on déceler s'il existe une relation entre deux variables?
- Comment, lorsque les chiffres proviennent d'une enquête, peut-on déterminer si la relation observée peut être attribuable au hasard de l'échantillonnage?
- Quelle est la nature de la relation? Est-ce une relation de cause à effet ou est-ce une simple corrélation?
- Quelle est la force de la relation? Peut-on utiliser les résultats observés pour faire des prédictions intéressantes?

1. SCHÉMA DE RELATIONS

Qui influence quoi?

Avant de faire de savants calculs sur la force éventuelle d'une relation, il est indispensable d'établir comment les variables sont reliées les unes aux autres. Pour cela, rien ne vaut un bon vieux dessin. Nous vous proposons ici deux procédés (schéma et courbe) permettant de clarifier les relations entre les variables et de faire ressortir certains aspects intéressants de ces relations.

1.1. Sagesse populaire

En Italie, une croyance populaire très ancienne veut que le vin soit un fortifiant idéal pour un préadolescent un peu anémique. Un verre quotidien, de *rouge* évidemment, et le sang retrouve ses couleurs. Au pays de Descartes, cette coutume fait sourire. Au pays de Victoria, elle scandalise. Enfin, au pays de Ti-Poil, on se contente de taxer le vin sans juger personne. Et si la sagesse populaire italienne avait raison envers et contre tous?

Qu'est-ce qui se cache derrière cette corrélation?

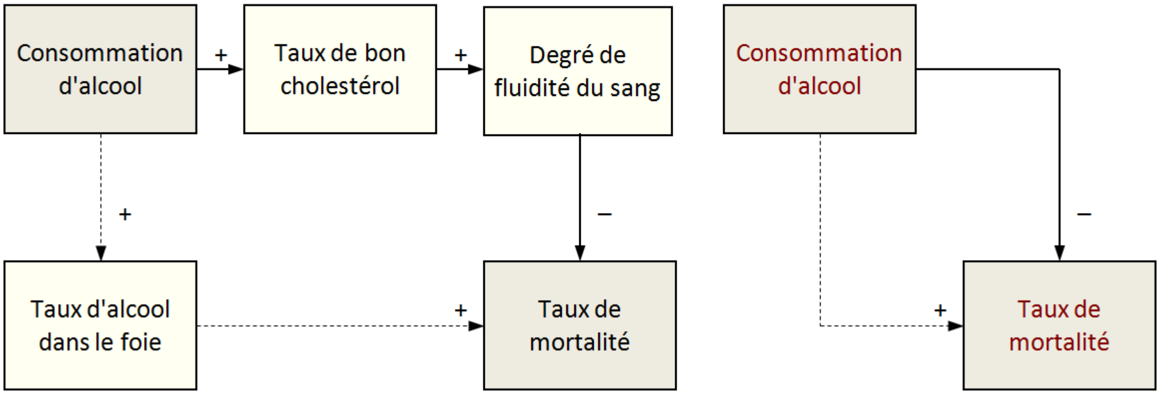
Selon une recherche citée par le magazine *Sciences et Avenir* de février 1995, on a observé une *corrélation négative* entre la consommation d'alcool et le taux de mortalité : *corrélation* parce que les deux variables varient en même temps, et *négative* parce qu'elles varient en sens inverse.

Il existe théoriquement trois manières de relier ces variables :

- 1) La consommation d'alcool (variable indépendante) exerce une influence sur le taux de mortalité (variable dépendante).
- 2) Le taux de mortalité (variable indépendante) exerce une influence sur la consommation d'alcool (variable dépendante).
- 3) La consommation d'alcool et le taux de mortalité sont toutes deux influencées par une troisième variable à découvrir.

La deuxième hypothèse doit être rejetée d'emblée parce qu'elle est absurde. Par ailleurs, c'est grâce à la découverte des chaînons manquants reliant les variables que nous pourrons finalement écarter la troisième hypothèse et retenir la première. C'est ce que nous avons voulu illustrer dans la figure 9.1.

FIGURE 9.1 – L'alcool, gage de santé jusqu'à un certain point



Comment construire un schéma de variables.

Le schéma de variables, dont nous avons déjà vu quelques exemples au chapitre 6, illustre les relations de cause à effet (représentées par une flèche) entre les variables (représentées par des cases). Les signes qui accompagnent les flèches représentent le sens de la relation : un signe positif indique que les variables augmentent en même temps et diminuent en même temps (relation *directe*); un signe négatif indique que les variables varient en sens *inverse*. Évidemment, le signe ne s'applique pas pour des variables purement nominales dans lesquelles il n'y a ni ordre ni direction.

Dans le schéma, on constate qu'il y a une relation *directe* entre la consommation d'alcool et le taux de bon cholestérol : les deux variables varient dans le même sens, ce que nous indiquons par un signe positif. On note aussi qu'il existe une relation *inverse* entre le degré de fluidité du sang et le taux de mortalité : lorsque le premier augmente, le second diminue, ce que nous indiquons par un signe négatif.

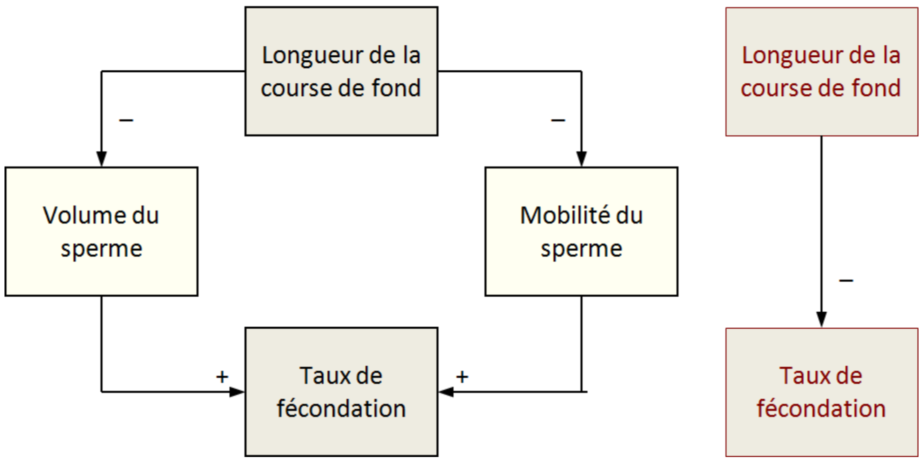
Dans la partie droite de la figure 9.1, nous avons fait la synthèse de la relation en éliminant toutes les variables intermédiaires : la cause (consommation d'alcool) est reliée à l'effet (taux de mortalité) par une flèche surmontée d'un signe négatif. Ce signe est obtenu en combinant tous les signes successifs du schéma original, en prenant pour principe que la relation change de sens chaque fois que l'on rencontre un chiffre négatif.

Alors, vite! Une bouteille et ça presse? Minute! Il est clair que nous nous sommes contentés d'étudier un seul aspect de la question. Passé un certain seuil, par exemple, la consommation d'alcool est reliée de façon *inverse* au taux de mortalité par l'intermédiaire de ce cher vieux foie que nous avons tous.

1.2. Tarzan au lit

Rien de tel qu'épouser une personne bardée de prix Nobel ou de médailles olympiques pour avoir des enfants beaux et forts, n'est-ce pas? Cependant, avant de faire votre demande en mariage, prenez note du fait suivant : on a observé, chez les hommes qui s'adonnent à la course de fond, que le sperme était moins volumineux et les spermatozoïdes moins mobiles que d'ordinaire. De quoi diminuer les chances de fécondation? On serait tenté de prétendre que le faible volume de sperme réduit la mobilité des spermatozoïdes, un peu à la manière des billots qui se traînent misérablement sur une rivière l'été. Il n'en est rien cependant. Les phénomènes sont seulement corrélés, car ils dépendent d'une variable commune : la longueur de la course de fond, et l'état d'épuisement qui en résulte. C'est très sérieux! Nous l'avons également appris dans la revue *Sciences et Avenir* (voir le schéma de la figure 9.2).

FIGURE 9.2 - Rien ne sert de courir si on veut se reproduire



1.3. L'alcool au volant

Cette fois, l'alcool est plutôt nuisible.

Vous avez un ami qui a déjà conduit en état d'ivresse sans avoir d'accident. Cet ami en déduit que l'alcool et le volant font bon ménage, d'autant plus que son oncle, antialcoolique notoire, a déjà provoqué un carambolage monstre sur l'autoroute métropolitaine après avoir bu un verre d'eau.

Il n'en demeure pas moins que les conducteurs qui ont une prédilection pour la dive bouteille ont beaucoup plus de *chance* d'être impliqués dans un accident de la route que les non-buveurs. C'est ce qu'indique l'étude menée par Statistique Canada (voir la figure 9.3). Si un dessin vaut mille mots, une courbe vaut parfois mille calculs : nous avons tracé, sous la figure 9.3, des courbes* qui mettent en évidence la relation entre consommation d'alcool (sur l'axe horizontal) et taux d'accident (sur l'axe vertical).

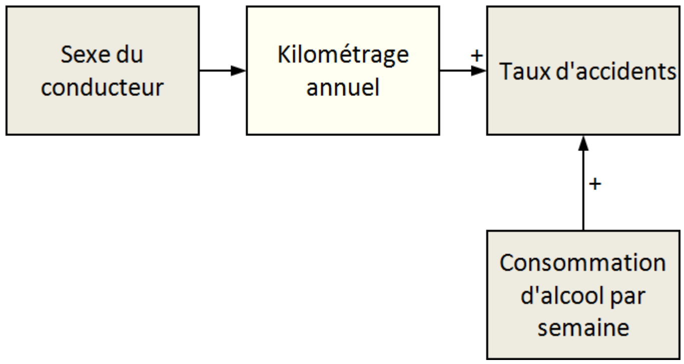
Ce genre de courbe n'est possible que si les échelles des variables sont, au minimum, des échelles ordinales.

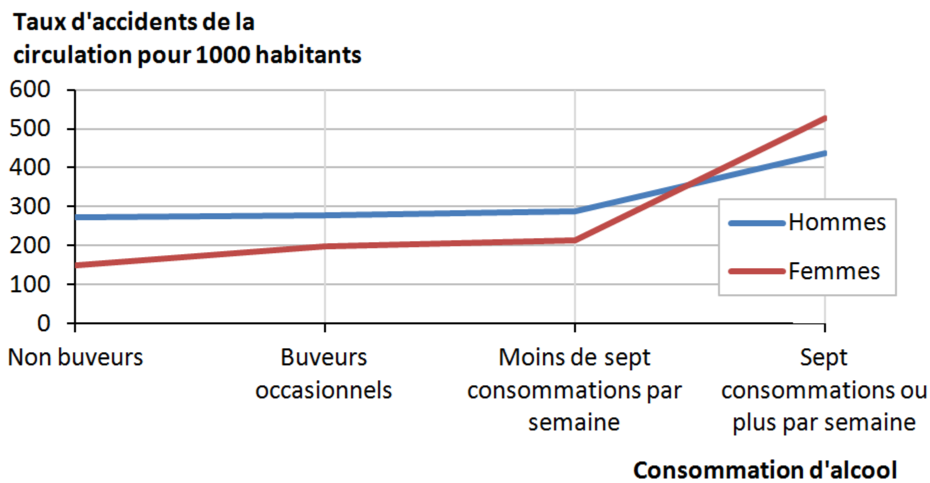
FIGURE 9.3 - Accidents de la circulation et consommation d'alcool

Taux d'accidents de la circulation pour 1000 habitants en 1988

	Hommes	Femmes
Non buveurs	274	149
Buveurs occasionnels	278	198
Moins de sept consommations par semaine	288	214
Sept consommations ou plus par semaine	437	526

Source : Statistique Canada, Les accidents au Canada, 1991. Données de 1988.





Comment interpréter les différences entre hommes et femmes? Une fausse explication serait de dire que les femmes boivent moins que les hommes. Que cela soit vrai ou non, cela n'a pas d'influence ici puisqu'on nous donne des taux d'accidents par catégorie de buveurs, et non des fréquences brutes. Il se pourrait également que les femmes conduisent mieux que les hommes — sauf celles qui boivent le plus. Cette hypothèse n'est pas à rejeter, quoiqu'il soit facile de proposer une autre explication : il se peut que les femmes aient moins d'accidents parce qu'elles conduisent moins. Il ne reste plus qu'à trouver les chiffres qui permettent de confirmer ou de rejeter toutes ces hypothèses.

1.4. La télévision

Si la durée d'écoute hebdomadaire de la télévision est très élevée au Québec, elle commence néanmoins à décliner. Les méchantes langues diront que c'est parce que de nouvelles formes d'abrutissement ont été mises au point. Au fait, savez-vous qui regarde le plus la télévision : les hommes ou les femmes? Les « vrais » Québécois ou les « faux »? Encore une fois, des chiffres bien choisis vont pouvoir détruire (ou renforcer) quelques préjugés.

Dans la figure 9.4, nous avons choisi de représenter par des courbes la relation entre *heures d'écoute* (sur l'axe vertical, la position préférée de la variable dépendante) et *groupe d'âge* (sur l'axe horizontal*. Les courbes mettent bien en évidence le fait que l'écoute de la télévision augmente avec l'âge. Cela dit, la véritable explication vient peut-être du fait que les vieux ont plus de temps libre que les jeunes (voir le schéma de variables au centre de la figure 9.4). Pour vérifier cette hypothèse, on pourrait consulter une des nombreuses études sur le temps libre*. Cette influence de l'âge sur les heures d'écoute se retrouve systématiquement chez les deux sexes et auprès des deux groupes ethniques.

FIGURE 9.4 - Qui écoute la télévision?

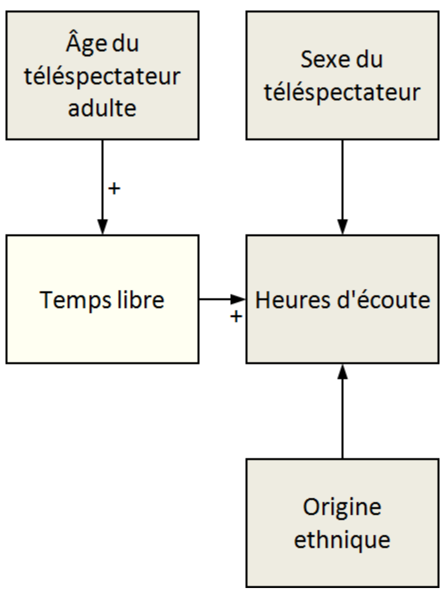
Heures d'écoute hebdomadaire moyenne de la télévision au Québec

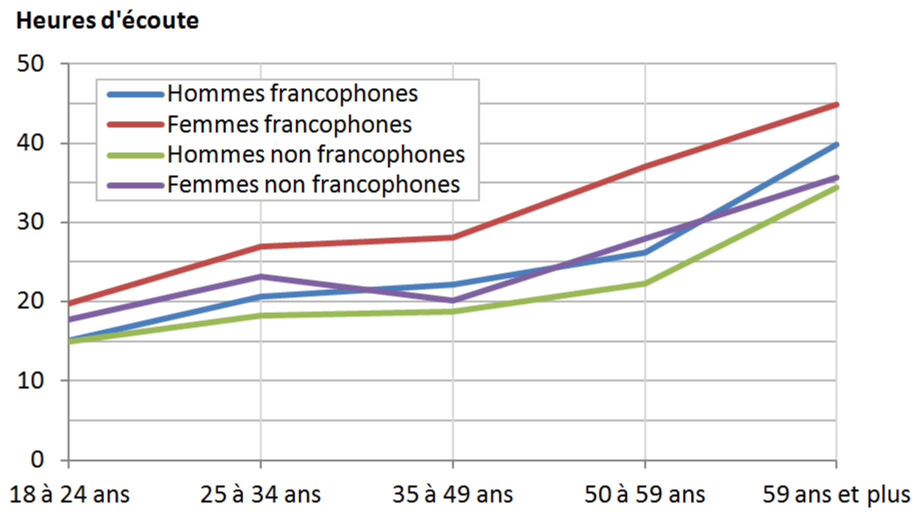
	Francophones		Non francophones	
	Hommes	**Femmes**	**Hommes**	**Femmes**
18 à 24 ans	15,1	19,8	15,0	17,7
25 à 34 ans	20,6	27,0	18,3	23,2
35 à 49 ans	22,1	28,1	18,7	20,1
50 à 59 ans	26,2	37,1	22,3	27,9
59 ans et plus	39,8	44,8	34,4	35,7
Total adultes (18 ans et plus)	**24,3**	**31,8**	**22,3**	**25,8**

Sources : Statistique Canada, L'écoute de la télévision, Automne 1994. Sondages BBM.

Il aurait été impossible de construire de telles courbes avec les autres variables (sexe et origine ethnique) qui appartiennent à des échelles nominales.

Par exemple : L'emploi du temps, Statistique Canada, 1991, SC 11-612F, no 4.



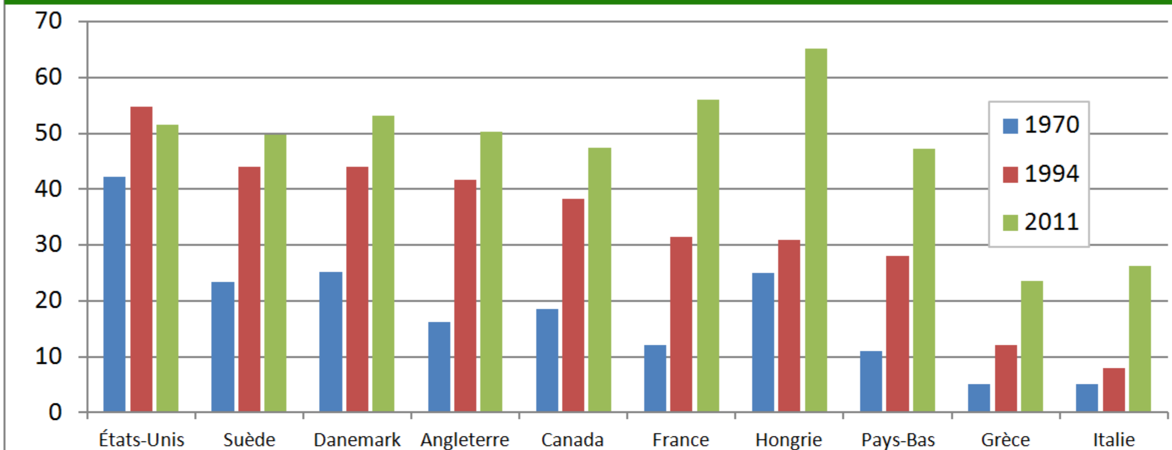


Encore une fois, l'utilisation de courbes et d'un schéma de variables a permis de mettre en évidence la relation et d'enrichir la description du phénomène.

1.5. Un cas plus complexe : les deux variables s’influencent mutuellement

Le taux de divorce, c’est-à-dire le nombre de divorces par rapport au nombre de mariages, est en hausse dans tous les pays industrialisés. En observant la figure 9.5, on se demande si le divorce et le développement économique ne sont pas corrélés.

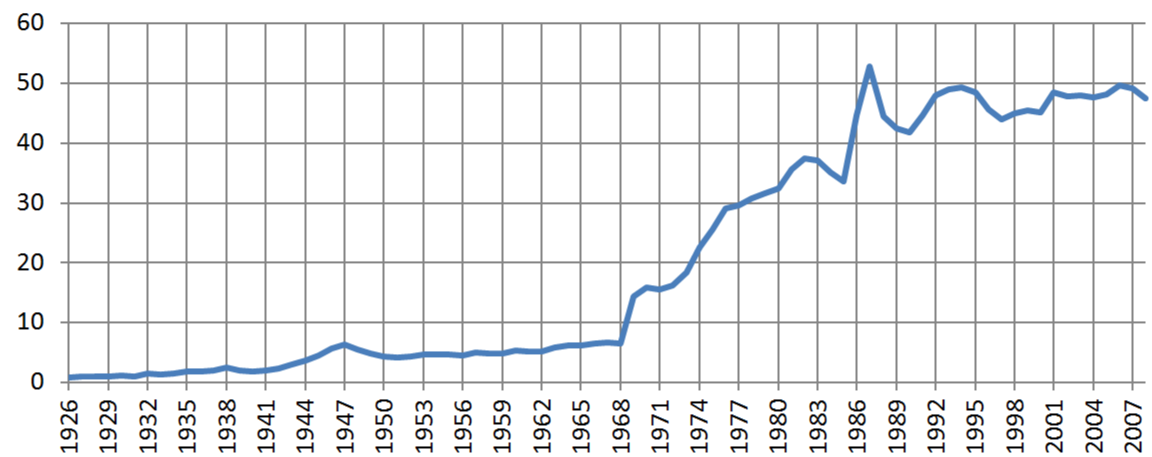
FIGURE 9.5a - Nombre de divorces pour 100 mariages dans divers pays



Note : Données de 2008 au lieu de 2011 pour le Canada et les États-Unis; données de 2010 pour la Grèce.

Sources : Courrier international, 29 juin 1995. Eurostat, Statistique Canada, US Census Bureau.

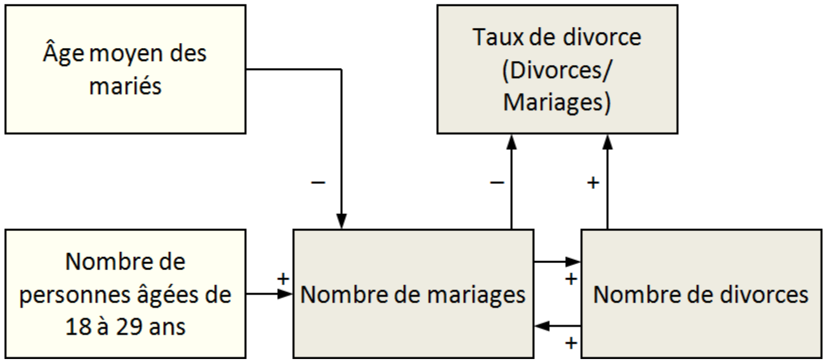
FIGURE 9.5b - Évolution du nombre de divorces pour 100 mariages au Canada



Source : Statistique Canada, Enquêtes 3232 et 3235.

Nous utiliserons ce dernier exemple pour montrer qu’il n’est pas toujours facile de distinguer la variable dépendante de la variable indépendante dans une relation de cause à effet. Le nombre de divorces dépend, entre autres, du nombre de mariages : seuls les gens mariés peuvent divorcer. Mais le nombre de divorces peut, lui aussi, influencer le nombre de mariages : chaque divorcé est un futur marié potentiel remis « sur le marché ». Nous avons illustré cette influence réciproque à la figure 9.6, que nous avons par ailleurs enrichie en tenant compte de certains facteurs démographiques : lorsqu’une vague de jeunes arrive à l’âge de convoler, le nombre de mariages augmente naturellement, suivi, quelques années plus tard par une hausse du nombre de divorces.

FIGURE 9.6 - Interrelation entre divorce et mariage

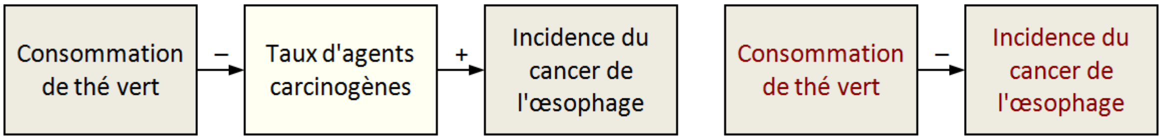


EXERCICES 1

1. Le thé, autre gage de santé

Répondez aux questions en vous référant au schéma de variable de la figure 9.7.

FIGURE 9.7 - Le thé vert est bon pour la santé



a) Quelle est la nature de la relation entre la consommation de thé vert et l'incidence du cancer de l'œsophage (simple corrélation ou relation de cause à effet)?

b) Si la relation en est une de cause à effet, quelle est la variable associée à la cause, quelle est la variable associée à l'effet et quel est le sens de la relation (direct ou inverse)?

2. Quand on n'en a pas on l'aime, quand on en a on s'ennuie

a) Pour mettre en évidence la relation entre l'âge et l'organisation du temps des individus, tracez des courbes, à partir des données du tableau 9.1.

TABLEAU 9.1 - La répartition du temps des Canadiens

(nombre d'heures)	Soins personnels	Activités productrices	Activités de loisir	Non précisé	**Total**
15 à 24 ans	10,9	7,6	5,4	0,0	**24,0**
25 à 34 ans	10,6	8,5	4,8	0,0	**24,0**
35 à 44 ans	10,7	8,7	4,6	0,1	**24,0**
45 à 54 ans	10,7	8,0	5,2	0,1	**24,0**
55 à 64 ans	11,4	6,3	6,2	0,1	**24,0**
65 ans et plus	12,4	3,8	7,7	0,1	**24,0**

Source : Statistique Canada, L'emploi du temps, 1991. Données de 1986.

Note : Soins personnels : dormir, se nourrir, se laver, se vêtir.

Activités productrices : travail rémunéré, activités éducatives, travaux ménagers, soin des enfants, etc.

Activités de loisir : télévision, lecture, divertissement, sport, passe-temps, bénévolat, etc.

b) À propos de la figure 9.4, nous avions émis une hypothèse selon laquelle les personnes plus âgées ont de bonnes raisons d'écouter la télévision, étant donné qu'elles disposent de plus de temps libre que les jeunes. Le tableau 9.1 confirme-t-il cette hypothèse?

2. RELATION ENTRE DEUX VARIABLES QUALITATIVES

Après avoir constaté la présence d'une relation entre deux variables, il reste à évaluer la force de cette relation. Par ailleurs, si les chiffres obtenus proviennent d'un simple échantillon, il faut prendre quelques précautions supplémentaires en s'assurant que les résultats observés ne sont pas l'effet du hasard.

2.1. Le tableau croisé

Chaque colonne du tableau croisé correspond à une catégorie d'une des deux variables et chaque ligne correspond à une catégorie de l'autre variable.

Lorsque les variables sont qualitatives, on a souvent recours à un croisement entre les catégories de chaque variable*. Dans le cas du tabagisme chez les jeunes, par exemple, on pourrait croiser la variable *sexe* (en colonnes) avec la variable *consommation de cigarettes* (en lignes) pour déterminer si le sexe a une influence sur le comportement (l'inverse est peu probable!). Chaque case du tableau croisé contiendra une fréquence, c'est-à-dire le nombre d'individus correspondant simultanément à la caractéristique de la colonne et à celle de la ligne.

Nous examinerons le cas des variables quantitatives dans la section suivante.

Dans chaque case du tableau croisé, on inscrit les fréquences associées à la catégorie de la colonne et à celle de la ligne correspondantes.

Les données du tableau 9.2 et du tableau 9.3 proviennent de deux enquêtes effectuées dans les années 1990 sur les adolescents au Québec. Nous les avons extraites d'un article publié dans la *Revue québécoise de psychologie* dans lequel l'auteur cherche à montrer que la perception des gens à l'égard des adolescents est plus influencée par les préjugés que par la réalité. Notre propos est ici beaucoup plus modeste : nous utiliserons deux exemples de cette recherche pour étudier la présence et la force d'une relation entre deux variables qualitatives. Le premier exemple traite d'un comportement (« fumes-tu la cigarette? ») et le second d'une perception des choses (« mes parents se chicanent souvent entre eux »). Nous essaierons de déterminer si le sexe a une influence sur le comportement (dans le premier cas) ou sur les perceptions (dans le second).

Commençons par observer les données brutes de l'enquête sur le tabagisme. Dans le tableau 9.2a, on constate que 5580 élèves du secondaire ont été interrogés, dont 2650 garçons et 2930 filles. Sur les 2930 filles interrogées, 322 déclarent fumer régulièrement, 355 à l'occasion et 2253 jamais. Toutes proportions gardées, puisque l'échantillon contient davantage de filles que de garçons, ces dernières semblent plus enclines à fumer que leurs confrères.

TABLEAU 9.2 - Le tabagisme chez les élèves du secondaire au Québec : « Fumes-tu la cigarette? »

Tableau 9.2a - Fréquences observées

	Garçons	Filles	Total
Régulièrement	196	322	**518**
À l'occasion	217	355	**572**
Jamais	2237	2253	**4490**
Total	**2650**	**2930**	**5580**

Tableau 9.2b - Proportion de chaque case (en %)

	Garçons	Filles	Total
Régulièrement	3,5	5,8	**9,3**
À l'occasion	3,9	6,4	**10,3**
Jamais	40,1	40,4	**80,5**
Total	**47,5**	**52,5**	**100,0**

Tableau 9.2c - Fréquences théoriques

	Garçons	Filles	Total
Régulièrement	246	272	**518**
À l'occasion	272	300	**572**
Jamais	2132	2358	**4490**
Total	**2650**	**2930**	**5580**

Tableau 9.2d - Écarts au carré relatifs

	Garçons	Filles	Total
Régulièrement	10,16	9,19	**19,36**
À l'occasion	10,99	9,94	**20,94**
Jamais	5,14	4,65	**9,78**
Total	**26,29**	**23,78**	**50,08**

Source : données brutes tirées de l'article de Richard CLOUTIER, « L'image des adolescents rongée par les mythes », Revue québécoise de psychologie, vol. 16, nº 3, 1995.

Note : Données de 1991. Enquête menée au Québec auprès des élèves du secondaire.

Pour en avoir le cœur net, observons les proportions calculées dans le tableau 9.2b. Étant donné que les filles représentent 52,5 % de l'échantillon, on peut s'attendre à retrouver dans la colonne « filles » des proportions légèrement supérieures à celles de la colonne « garçons ». Or, on constate que sur les 9,3 % d'élèves qui fument régulièrement, 5,8 % sont des filles et 3,5 % sont des garçons : on ne peut plus parler de « légères » différences.

Étant donné que 9,3 % des élèves fument régulièrement et que 47,5 % d'entre eux sont des garçons, on pourrait s'attendre à ce que 9,3 % x 47,5 % des fumeurs réguliers soient des garçons. Sur les 5580 élèves interrogés, il devrait donc y avoir « théoriquement » 9,3 % x 47,5 % x 5580 = 246 garçons classés comme fumeurs réguliers. C'est ce qu'on appelle la fréquence théorique (tableau 9.2c), par opposition à la fréquence observée (tableau 9.2a).

Fréquence théorique d'une case = Proportion de la ligne x Proportion de la colonne x Fréquence totale

Fréquence théorique de filles qui ne fument jamais = 80,5 % x 52,5 % x 5580 = 2358
(Ou encore : 0,805 x 0,525 x 5580 = 2358)

Si le sexe n'exerçait aucune influence sur le tabagisme, la répartition des élèves entre les diverses catégories (tableau 9.2a) devrait ressembler à celle obtenue dans le tableau des fréquences théoriques (tableau 9.2c). Lorsque l'on compare ces fréquences théoriques aux fréquences observées dans l'enquête, on constate toutefois certains écarts, qui ne sont ni énormes ni négligeables. L'utilisation d'un outil assez répandu, le Khi carré, va nous permettre d'évaluer l'importance de cet écart et de l'interpréter.

Examinons la première case du tableau 9.2a : 196 individus y sont recensés alors qu'on s'attendait à en retrouver 246 (voir tableau 9.2c). On est donc en déficit de 50 individus (–50) sur un total de 246. Comme c'est l'écart absolu qui est important, nous nous débarrassons du signe en mettant cet écart au carré avant de le diviser par le total. Nous obtenons ainsi l'écart au carré relatif.

Écart^2 relatif = (Fréquence observée – Fréquence théorique)^2/Fréquence théorique

Écart^2 relatif pour les garçons qui fument régulièrement = (196 – 246)^2/246 = 2500/246 = 10,16

On retrouve ce dernier chiffre dans le tableau 9.2d. Le Khi carré n'est autre que la somme de tous ces écarts : il est ici de 50,08. Si les fréquences observées correspondaient exactement aux fréquences théoriques, le Khi carré aurait une valeur de 0. Il nous faut maintenant interpréter la valeur de 50,08 que nous avons obtenue.

2.2. L'écart peut-il être attribué au hasard?

Les données que nous venons d'utiliser proviennent d'une enquête. Faute d'information précise sur la population des élèves du secondaire dans son ensemble, nous devons nous contenter d'un échantillon. Mais quelles sont les chances que nous soyons tombés, par malheur, sur un échantillon non représentatif? Comme pour le test d'hypothèse (vu au chapitre précédent), nous allons nous donner un seuil de signification, mettons 0,05 (ou 5 %, ou 1/20). Cela signifie que nous ne voulons courir le risque de nous tromper qu'une fois sur 20 si jamais nous émettons l'hypothèse selon laquelle les écarts sont suffisamment grands pour ne pas être le simple effet du hasard*.

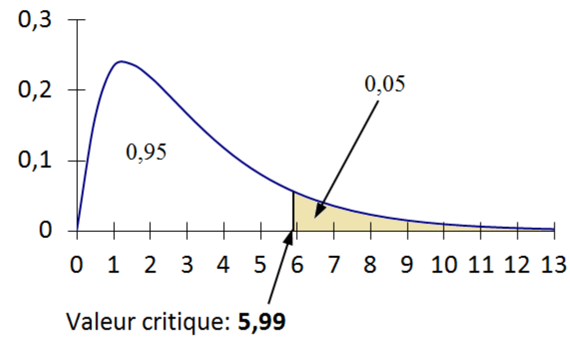
Nous aurions pu prendre un seuil de signification plus grand ou plus petit : le tout dépend du degré du risque que nous sommes prêts à accepter dans le contexte.

La figure 9.8 nous indique la valeur critique, c'est-à-dire la valeur minimale que doit atteindre notre Khi carré pour que nous puissions accepter notre hypothèse avec un risque de nous tromper inférieur à 5 %. Mais comme le Khi carré est influencé par la taille du tableau (plus le tableau est grand, plus les écarts s'accumulent), il nous faut tenir compte du nombre de colonnes et lignes du tableau. C'est ce qu'on appelle le nombre de degrés de liberté.

FIGURE 9.8 - Distribution du Khi carré

Valeur critique selon le seuil de signification et le nombre de degrés de liberté

Degrés	Seuil de signification		
de liberté	0,1	0,05	0,01
1	2,71	3,84	6,64
2	4,61	**5,99**	9,21
3	6,25	7,82	11,34
4	7,78	9,49	13,28
5	9,24	11,07	15,09
6	10,64	12,59	16,81
7	12,02	14,07	18,48
8	13,36	15,51	20,09
9	14,68	16,92	21,67



Degrés de liberté = (Nombre de colonnes – 1) x (Nombre de lignes – 1)

Degrés de liberté = (2 – 1) x (3 – 1) = 1 x 2 = 2

Dans la table de distribution du Khi carré, la valeur correspondant à un seuil de signification de 0,05 et à 2 degrés de liberté est égale à 5,99. Or, l'écart que nous avons calculé (50,08) dépasse largement cette *valeur critique*. En fait, nous sommes si loin au-dessus de la valeur critique qu'il est pratiquement *impossible* que les écarts constatés soient imputables à un hasard d'échantillonnage. Cette certitude n'est pas étrangère au fait que nous disposons d'un échantillon très élevé : la loi des grands nombres joue encore en notre faveur. Autrement dit, le Khi carré a d'autant plus de chances d'être élevé que l'échantillon est grand.

Le nombre de degrés de liberté peut s'interpréter de la façon suivante : nous avons sous la main un tableau dans lequel on croise les deux catégories d'une première variable (le sexe) avec les trois catégories d'une seconde variable (le tabagisme). Il va de soi que les totaux rajoutés aux lignes et aux rangées du tableau sont déterminés d'avance et ne relèvent en rien du hasard. Or, une fois que l'on connaît les fréquences associées aux garçons, par exemple, on peut en déduire automatiquement les fréquences des filles. De la même façon, une fois connues les fréquences des deux premières catégories concernant le tabagisme, nous connaissons également la fréquence de la troisième catégorie. En somme, la dernière colonne et la dernière ligne du tableau (totaux exclus) sont prédéterminées par les autres colonnes et lignes. C'est pourquoi on les exclut dans le calcul du nombre de degrés de liberté.

2.3. La relation est-elle forte?

Si nous venons d'éviter une première embûche, il nous faut maintenant évaluer dans quelle mesure la relation observée est forte. Nous pouvons affirmer que les filles sont plus portées à fumer que les garçons. Mais cela est-il suffisant, par exemple, pour nous permettre de prédire facilement, à partir de son sexe, si un élève fume?

Si nous avions croisé le sexe des élèves avec une autre variable, comme le port du soutien-gorge ou de la coquille protectrice dans les arts martiaux, nous aurions sûrement eu des écarts encore plus tranchés. Notre Khi carré aurait été suffisamment élevé pour que nous puissions déduire à partir de ses habitudes vestimentaires, et sans grand risque de nous tromper, si un *budōka* est une fille ou un garçon.

Un des instruments les plus courants pour évaluer la force de la relation entre deux variables qualitatives est le *V de Cramer*. On comprendra que le Khi carré ne peut tout dire à lui tout seul, puisqu'il dépend en partie de la taille de l'échantillon. Le V de Cramer tient compte de cet aspect. Notez, en observant la formule ci-après, que vous auriez pu facilement inventer vous-même un coefficient correspondant et passer ainsi à la postérité.

$$\textbf{V de Cramer} = \sqrt{\frac{\text{Khi}^2}{n \times (K - 1)}}$$

où *n* représente la taille de l'échantillon et *K* le nombre minimal de rangées et de colonnes.

On compte 3 rangées et 2 colonnes dans le tableau 9.2. C'est le plus petit de ces deux chiffres que l'on retient pour calculer le V de Cramer.

$$\text{V de Cramer} = \sqrt{[50{,}08/5580 \times (2-1)]} = 0{,}095$$

S'il n'y avait aucune relation entre les deux variables, le V de Cramer serait égal à 0, tout comme le Khi carré. Si la relation était parfaitement tranchée (par exemple si toutes les filles fumaient et si tous les garçons ne fumaient pas), le V de Cramer serait égal à 1. Dans le cas du port du soutien-gorge ou de la coquille protectrice, on aurait peut-être un V de Cramer égal à 0,99<$F>. En général, on considère que l'association entre les variables commence à être intéressante à partir de 0,10, forte à partir de 0,40 et robuste à partir de 0,70, mais cela dépend du contexte.

Si vous décidez d'entreprendre une recherche sur le sujet, nous vous prions de nous communiquer les résultats.

Ici, nous avons un coefficient relativement faible (0,095). Il y a certes une différence de comportement entre les garçons et les filles, mais cette différence est trop faible pour qu'on puisse prédire, à partir de son sexe, si un élève fume. Dans les paragraphes qui suivent, nous proposerons deux exemples de conclusions que l'on pourrait tirer à la lecture de ce chiffre : la première est acceptable et la seconde, abusive.

Une conclusion honnête

Il est clair que les écolières fument un peu plus que les garçons. Pour mieux comprendre la situation, nous pourrions nous poser diverses questions : « Pourquoi telle fille fume-t-elle? Pourquoi telle autre ne fume-t-elle pas? Qu'est-ce qui pousse un jeune à fumer? Qu'est-ce qui pousse un garçon à fumer? », etc.

Une conclusion biaisée

« Puisque les filles fument plus que les garçons, je refuse d'embaucher des filles dans mon usine de dynamite. » Cette affirmation serait aussi ridicule que les suivantes : « Puisque les faux-monnayeurs sont plus souvent des étrangers (V de Cramer à l'appui), un commerçant ne devrait pas accepter les billets de ses clients italiens » ou encore « Puisqu'une proportion relativement grande d'Asiatiques ont un QI supérieur à 100, les universités de devraient pas engager de professeurs originaires d'Europe. »

2.4. Mes parents se chicanent, un peu, beaucoup

Les données du tableau 9.3 proviennent d'une enquête effectuée auprès d'un échantillon de 3180 élèves du secondaire du Québec âgés de 11 à 19 ans. L'échantillon initial comptait 6121 élèves choisis au hasard, mais 2916 questionnaires ne furent pas remplis et 25 questionnaires furent rejetés. Certaines des questions portaient sur la violence verbale et physique au sein de la famille. On demandait notamment à l'élève d'évaluer si ses parents se chicanaient souvent entre eux, à partir de l'échelle ordinale suivante : 1. Correspond tout à fait à ce que je vis; 2. Correspond un peu à ce que je vis; 3. Ne correspond pas vraiment à ce que je vis; 4. Ne correspond pas du tout à ce que je vis.

TABLEAU 9.3 - Les perceptions des adolescents sur leurs parents
Mes parents se chicanent souvent…

Tableau 9.3a - Fréquences observées

	Garçons	Filles	Total
Tout à fait	58	125	**183**
Un peu	179	269	**448**
Pas vraiment	395	524	**919**
Pas du tout	738	842	**1580**
Total	**1370**	**1760**	**3130**

Tableau 9.3b - Proportion de chaque case (en %)

	Garçons	Filles	Total
Tout à fait	1,9	4,0	**5,8**
Un peu	5,7	8,6	**14,3**
Pas vraiment	12,6	16,7	**29,4**
Pas du tout	23,6	26,9	**50,5**
Total	**43,8**	**56,2**	**100,0**

Tableau 9.3c - Fréquences théoriques

	Garçons	Filles	Total
Tout à fait	80	103	**183**
Un peu	196	252	**448**
Pas vraiment	402	517	**919**
Pas du tout	692	888	**1580**
Total	**1370**	**1760**	**3130**

Tableau 9.3d - Écarts au carré relatifs

	Garçons	Filles	Total
Tout à fait	6,10	4,75	**10,84**
Un peu	1,49	1,16	**2,65**
Pas vraiment	0,13	0,10	**0,23**
Pas du tout	3,12	2,43	**5,54**
Total	**10,83**	**8,43**	**19,27**

Source : données brutes tirées de l'article de Richard CLOUTIER, « L'image des adolescents rongée par les mythes », Revue québécoise de psychologie, vol. 16, nº 3, 1995.

Note : Données de 1994. Enquête menée au Québec auprès des élèves du secondaire.

Sur les 3130 élèves qui ont été en mesure de répondre à cette question précise, les filles semblent relativement plus nombreuses à estimer que leurs parents se chicanent. Étant donné que les garçons et les filles partagent généralement les mêmes parents, on peut considérer que les réponses reflètent non seulement la réalité, mais également la *perception* de cette réalité. Par ailleurs, il faut noter que la plupart des parents sont bel et bien des adeptes de la coexistence pacifique.

Des calculs similaires à ceux effectués pour le tableau 9.2 nous montrent que le Khi carré est égal à 19,27. Ce tableau contient cependant plus de cases que le précédent : on y compte 2 colonnes et 4 lignes, soit (2 – 1) × (4 – 1) = 1 × 3 = 3 degrés de liberté. Si l'on prend ici encore un seuil de signification de 0,05, la valeur critique est de 7,82 (revoir la figure 9.8). Notre Khi carré est donc suffisamment grand pour qu'on ne puisse pas mettre les différences observées entre garçons et filles sur le dos du hasard.

Encore une fois, le V de Cramer est relativement petit : $\sqrt{[19{,}27/3130 \times (2 - 1)]} = 0{,}08$. L'association entre le sexe et la perception de la réalité est donc plutôt faible. Même si la relation existe, il serait présomptueux de formuler des généralisations.

EXERCICES 2

1. Des jeunes drogués

a) Vérifiez dans le tableau 9.2 la proportion, la fréquence théorique et l'écart pour les filles qui fument régulièrement.

b) Quelle aurait été la valeur critique du Khi carré si nous avions choisi un seuil de signification de 0,01 pour le tableau 9.2? Commentez.

2. Une jeunesse qui s'envole en fumée

Le tableau 9.4 contient des informations similaires à celles du tableau 9.2. Les données sont cependant tirées d'une enquête différente et portent sur une période plus récente.

TABLEAU 9.4 - Le tabagisme chez les élèves du secondaire au Québec quelques années plus tard : « Fumes-tu la cigarette? »

Fréquences observées

	Garçons	Filles	Total
Régulièrement	174	341	**515**
À l'occasion	153	353	**506**
Jamais	1054	1100	**2154**
Total	**1381**	**1794**	**3175**

Source : données brutes tirées de l'article de Richard CLOUTIER, « L'image des adolescents rongée par les mythes », Revue québécoise de psychologie, vol. 16, nº 3, 1995.

Note : Données de 1994. Enquête menée au Québec auprès des élèves du secondaire.

a) Construisez un tableau contenant les proportions pour chaque case.

b) Construisez un tableau contenant les fréquences théoriques pour chaque case.

c) Construisez un tableau contenant les écarts au carré relatifs.

d) Calculez le Khi carré. Comparez à la valeur critique dans le tableau de distribution du Khi carré pour un seuil de signification de votre choix.

e) Calculez le V de Cramer.

f) Comparez les résultats obtenus à ceux du tableau 9.2. Commentez.

3. RELATION ENTRE DEUX VARIABLES QUANTITATIVES

Comme nous venons de le voir, le tableau croisé est un outil privilégié pour observer une relation entre deux variables *qualitatives*. Lorsque les deux variables sont *quantitatives*, on doit avoir recours à un autre procédé : la corrélation. La situation suivante expliquera de quoi il en retourne.

3.1. La corrélation

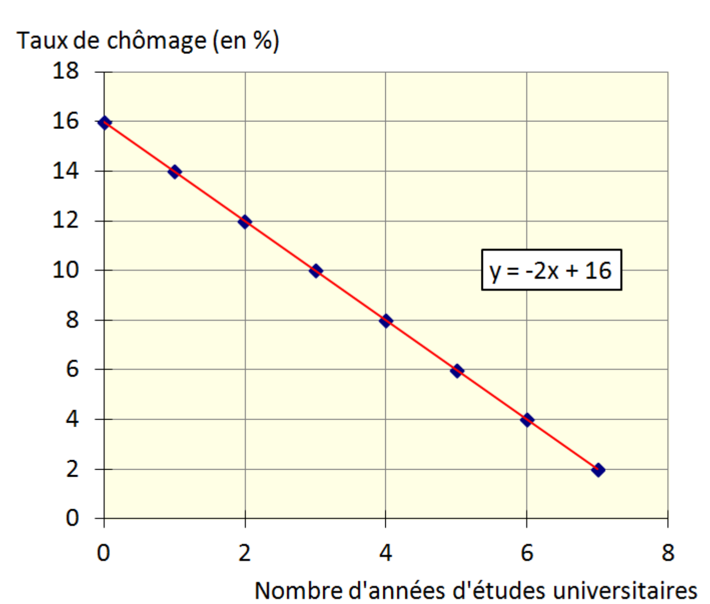
On dit souvent qu'il y a des chômeurs instruits. Même si cela est vrai, il s'agit néanmoins d'une espèce relativement rare si on la compare à celle des chômeurs « ignorants ». Supposons que le taux de chômage soit en moyenne de 2 % pour une personne qui détient un doctorat (7 ans d'études universitaires), de 6 % pour une personne qui détient une maîtrise (5 ans d'études universitaires) et de 10 % pour une personne qui détient un baccalauréat (3 ans d'études universitaires). Si la tendance se maintient, comme on dit, il est probable que le taux de chômage soit de 16 % pour une personne qui a interrompu ses études juste avant de rentrer à l'université. On pourrait même construire une formule qui permette de prédire le taux de chômage associé à un nombre x d'années d'études universitaires : Puisqu'il semble que chaque année d'étude fasse baisser le chômage de 2 points de pourcentage, la formule du taux de chômage serait donc la suivante :

$$y = 16 - 2x$$

Lorsque des points, représentant la valeur d'une variable par rapport à une autre variable, sont plus ou moins alignés, la droite de régression est celle qui s'éloigne le moins possible de l'ensemble des points.

Dans cette équation, y représente (en points de pourcentage) le taux de chômage, et x, le nombre d'années d'études universitaires. Après 4 années d'études universitaires, le taux de chômage serait, selon notre formule, de $16 - 2 \times 4 = 8$ points de pourcentage. L'équation peut également être représentée sous forme de courbe (ou droite de régression) comme dans la figure 9.9.

FIGURE 9.9 - Les chômeurs instruits, et les autres



Il y a corrélation entre deux variables lorsque ces deux variables se suivent de façon plus ou moins systématique, que ce soit dans le même sens ou en sens inverse.

Dans l'exemple que nous venons de présenter, on peut dire qu'il existe une corrélation parfaite entre les deux variables : le niveau universitaire atteint et le taux de chômage. Évidemment, la réalité humaine n'est pas aussi simple, et de toute façon, on ne peut pas tirer de grandes conclusions d'un échantillon aussi petit. Il existe cependant de nombreuses situations reliées aux sciences humaines dans lesquelles il est possible de tracer une droite de régression mettant en relation deux variables x et y. Il s'agit alors de déterminer la valeur des paramètres qui caractérisent l'équation de cette droite. Ces deux paramètres sont ici les nombres 16 et –2. On cherchera alors à évaluer dans quelle mesure les faits observés coïncident avec la droite tracée.

Le coefficient de corrélation mesure la force de la corrélation entre deux variables.

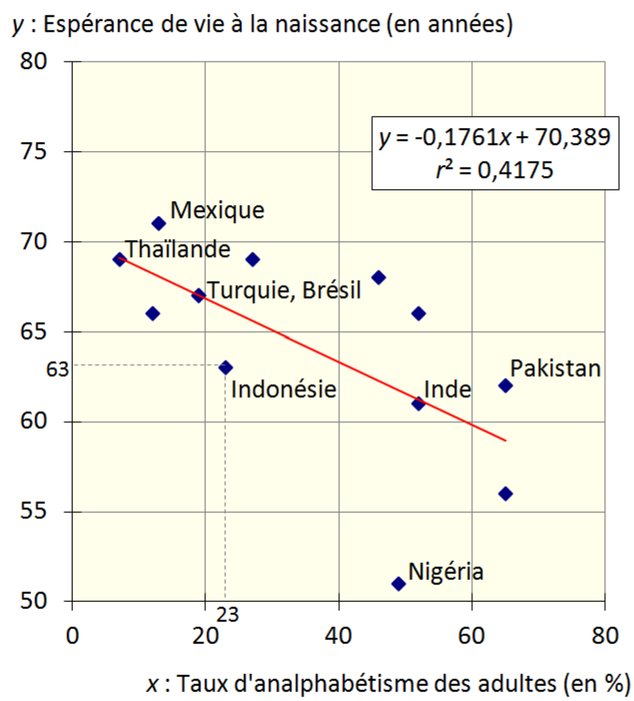
Plus généralement, on écrit l'équation de la droite sous la forme : $y = a + bx$, dans laquelle a est l'ordonnée à l'origine et b la pente. Une fois qu'on a déterminé la valeur de a et de b, on devrait être capable, connaissant x, de trouver y avec un certain degré de fiabilité (mesuré par le coefficient de corrélation). C'est ce que nous allons faire avec des données internationales.

3.2. Qui s'instruit… vit plus longtemps?

Nous avons choisi d'étudier la corrélation éventuelle entre l'éducation et la santé des gens. Plus précisément, nous avons retenu deux variables facilement observables à travers le monde : le taux d'analphabétisme des adultes et l'espérance de vie à la naissance. Pour ne pas encombrer le tableau, nous n'avons sélectionné que les gros pays (peuplés de 50 millions et plus) pour lesquels les données existent. Nous avons exclu de l'échantillon les pays industrialisés pour lesquels les taux officiels d'analphabétismes sont égaux à zéro. Grâce à ces simplifications, peut-être excessives, il nous sera plus facile de montrer comment construire et interpréter la droite de régression. En fin de compte, nous avons retenu 13 pays et nous avons représenté les données sous forme d'un tableau accompagné d'un graphique (voir figure 9.10).

FIGURE 9.10 - Analphabétisme et espérance de vie dansle monde

	Taux d'analphabétisme des adultes *(%)* 1990	Espérance de vie à la naissance *(années)* 1993
Vietnam	12	66
Bangladesh	65	56
Inde	52	61
Nigéria	49	51
Pakistan	65	62
Chine	27	69
Égypte	52	66
Indonésie	**23**	**63**
Thailande	7	69
Turquie	19	67
Iran	46	68
Brésil	19	67
Mexique	13	71



Source : Banque mondiale, *Rapport sur le développement dans le monde* 1995.
Note : Nous avons sélectionné tous les pays à revenu faible ou intermédiaire dont la population dépassait 50 millions d'habitants et pour lesquels les données étaient disponibles.

Chacun des points du graphique représente un pays. Le point correspondant à l'Indonésie, par exemple, a une abscisse de 23 (par rapport à l'axe horizontal) et une ordonnée de 63 (par rapport à l'axe vertical). On ne voit que 12 points sur le graphique, parce que deux pays, le Brésil et la Turquie, possèdent exactement les mêmes valeurs. Étant donné que le graphique a été tracé à l'aide d'un chiffrier électronique, il n'a pas été difficile, en sélectionnant la bonne option, d'y rajouter la droite de régression qui correspond à l'ensemble des points. Le chiffrier a même eu la bonté de nous fournir l'équation de la droite de régression (sous la forme y = bx + a), ainsi qu'un coefficient que nous interprèterons un peu plus loin. Certes, les points du graphique sont loin d'être parfaitement alignés, mais la droite donne quand même une tendance générale. On peut affirmer, d'emblée, qu'il existe une certaine corrélation entre l'analphabétisme et l'espérance de vie, et que ces deux variables évoluent en sens inverse.

3.3. Le coefficient de corrélation

Comme nous l'avons indiqué un peu plus haut, le coefficient de corrélation mesure la force de la relation entre les deux variables. Mais avant d'interpréter ce coefficient, il faut le calculer. Pour ce faire, il existe deux méthodes : la méthode facile (en utilisant les fonctions intégrées d'un chiffrier électronique) et la méthode à papa (en se tapant une série de calculs, simples mais laborieux). Nous avons déjà fait connaissance avec la méthode facile, puisque nous l'avons utilisée pour tracer la droite de régression et obtenir gratuitement l'équation de la droite et un certain coefficient (r^2) qui n'est autre que le carré du coefficient de corrélation (revoir le graphe de la figure 9.10).

La méthode à papa.

Dans le tableau 9.5, nous indiquons toutes les étapes du calcul du coefficient de corrélation et des paramètres de la droite de régression et nous reproduisons les formules correspondantes. Si vous disposez d'un chiffrier électronique (c'est presque indispensable dès qu'on utilise des chiffres en sciences humaines), vous pouvez vous dispenser de cette étape fastidieuse.

Tableau 9.5 - Calcul des paramètres de la droite de régression

	Calculs	Vérifications avec les fonctions du chiffrier	
b	-0,176	*-0,176*	*Droitereg(plage y; plage x)*
a	70,389	*70,389*	*Ordonnee.Origine(plage y; plage x)*
r	-0,646	*-0,646*	*Coefficient.Correlation(plage y; plage x)*

	x	*y*	*xy*	*x²*	*y²*
n (taille de l'échantillon)	13				
Moyenne	34,5	64,3			
Écart type	20,8	5,7			
Somme	449	836	27957	20717	54148
Vietnam	12	66	792	144	4356
Bangladesh	65	56	3640	4225	3136
Inde	52	61	3172	2704	3721
Nigéria	49	51	2499	2401	2601
Pakistan	65	62	4030	4225	3844
Chine	27	69	1863	729	4761
Égypte	52	66	3432	2704	4356
Indonésie	**23**	**63**	1449	529	3969
Thailande	7	69	483	49	4761
Turquie	19	67	1273	361	4489
Iran	46	68	3128	2116	4624
Brésil	19	67	1273	361	4489
Mexique	13	71	923	169	5041

Pente de la droite : *b* = [(*n* × Somme des *xy*) – (Somme des *x* × Somme des *y*)] / [(*n* × Somme des x^2) – (Somme des x)^2]
Ordonnée à l'origine : *a* = Moyenne des *y* – (*b* × Moyenne des *x*)
Coefficient de corrélation : *r* = *b* × (Écart type de *x*) / (Écart type de *y*)

Plus le coefficient de corrélation est proche de 0, moins la corrélation est forte.

Le coefficient de corrélation est construit de telle sorte qu'il est égal à +1 ou –1 lorsque les points sont parfaitement alignés. Dans notre exemple, le coefficient de corrélation est égal à –0,646. Le signe négatif indique que les deux variables évoluent en sens inverse. La valeur absolue du coefficient (0,646) semble relativement élevée, mais pour l'interpréter il est nécessaire de tenir compte de la taille de l'échantillon. On pourra alors, moyennant certains calculs supplémentaires, faire une hypothèse sur l'existence d'une association entre les deux variables et la tester avec la table de distribution de Student. Si vous tenez vraiment à savoir comment, il vous faudra consulter un ouvrage spécialisé.

EXERCICES 3

1. Échalote, Bouboule, Brummel et les autres

Vous devez évaluer s'il existe une corrélation entre le *poids* et la *taille* à l'aide d'un échantillon d'au moins 30 individus. (Pour les besoins de la cause, il est acceptable que l'échantillon ne soit pas tiré au hasard.) Tracez le nuage de points et calculez le coefficient de corrélation (à la main ou en utilisant le chiffrier électronique, calcul fourni)

2. Étranges corrélations

Les corrélations présentées ci-après paraissent pour le moins étranges. À vous de leur donner une explication logique et identifiant convenablement les variables impliquées et les relations qui les unissent.

a) On a observé une corrélation entre homicides et pointures des souliers. Plus précisément, il semble que les homicides soient plus fréquemment commis par des individus possédant des pieds plus grands que la moyenne.

b) On a déjà constaté, en Californie, une corrélation entre les ventes de bière et le taux de mortalité chez les personnes âgées et les bambins.

4. LES ACCIDENTS DE LA ROUTE

Les jeunes sont-ils susceptibles de provoquer plus d'accidents que leurs aînés? Les femmes conduisent-elles mieux que les hommes? Est-il plus dangereux de rouler en Chine qu'aux États-Unis? Nous essaierons de répondre, partiellement, à toutes ces questions en nous servant de tous les outils vus dans ce chapitre.

4.1. Un bref tour du monde

Examiner quelques données brutes pour découvrir le schéma de variables.

Commençons par déblayer le terrain en examinant quelques données brutes publiées par le gouvernement japonais dans les années 1990. Nous reproduisons dans le tableau 9.6 le nombre d'accidents de la route et le nombre de victimes de ces accidents pour six pays. Si les États-Unis se classent premiers, devant le Japon, au chapitre des accidents et du nombre de blessés, c'est la Chine qui détient le triste record du nombre de tués.

TABLEAU 9.6 - Les accidents de la route dans le monde

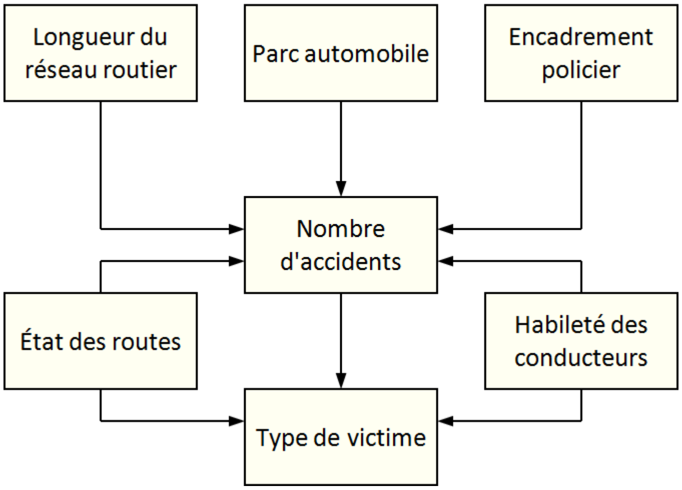
	Accidents	Blessés	Tués	(année)
Chine	228 326	144 021	58 723	**1992**
États-Unis	2 540 946	3 600 307	44 529	**1990**
Japon	662 388	810 245	11 105	**1991**
France	148 890	205 968	9 617	**1991**
Mexique	18 337	43 000	5 700	**1991**
Canada	182 294	261 139	3 957	**1990**
Québec			2 209	**1973**
Québec	304 160	45 155	845	**1995**
Québec	104 070	39 105	436	**2012**

Source : Japan international statistical survey, 1994. SAAQ, Bilan routier 2012.

Le tableau 9.6 contient également quelques données concernant le Québec et couvrant quatre décennies. On y constate que le nombre d'accidents et de victimes tend à diminuer considérablement avec le temps, malgré l'augmentation du nombre d'automobiles sur les routes.

Ce premier contact chiffré avec la situation étudiée nous permet de prendre conscience de la complexité du problème. Il est clair que l'habileté des conducteurs ou l'état des routes ne sont pas les seules variables qui peuvent exercer une influence sur le nombre d'accidents. Aux États-Unis, il y a plus de kilomètres de routes qu'au Japon (pays plus petit) et plus de véhicules qu'au Canada (pays moins peuplé) et qu'au Mexique (pays moins riche). La figure 9.11 illustre une façon de relier ces différentes variables ainsi que d'autres.

FIGURE 9.11 - Les accidents de la route



Il est plus difficile de comparer des situations hétérogènes.

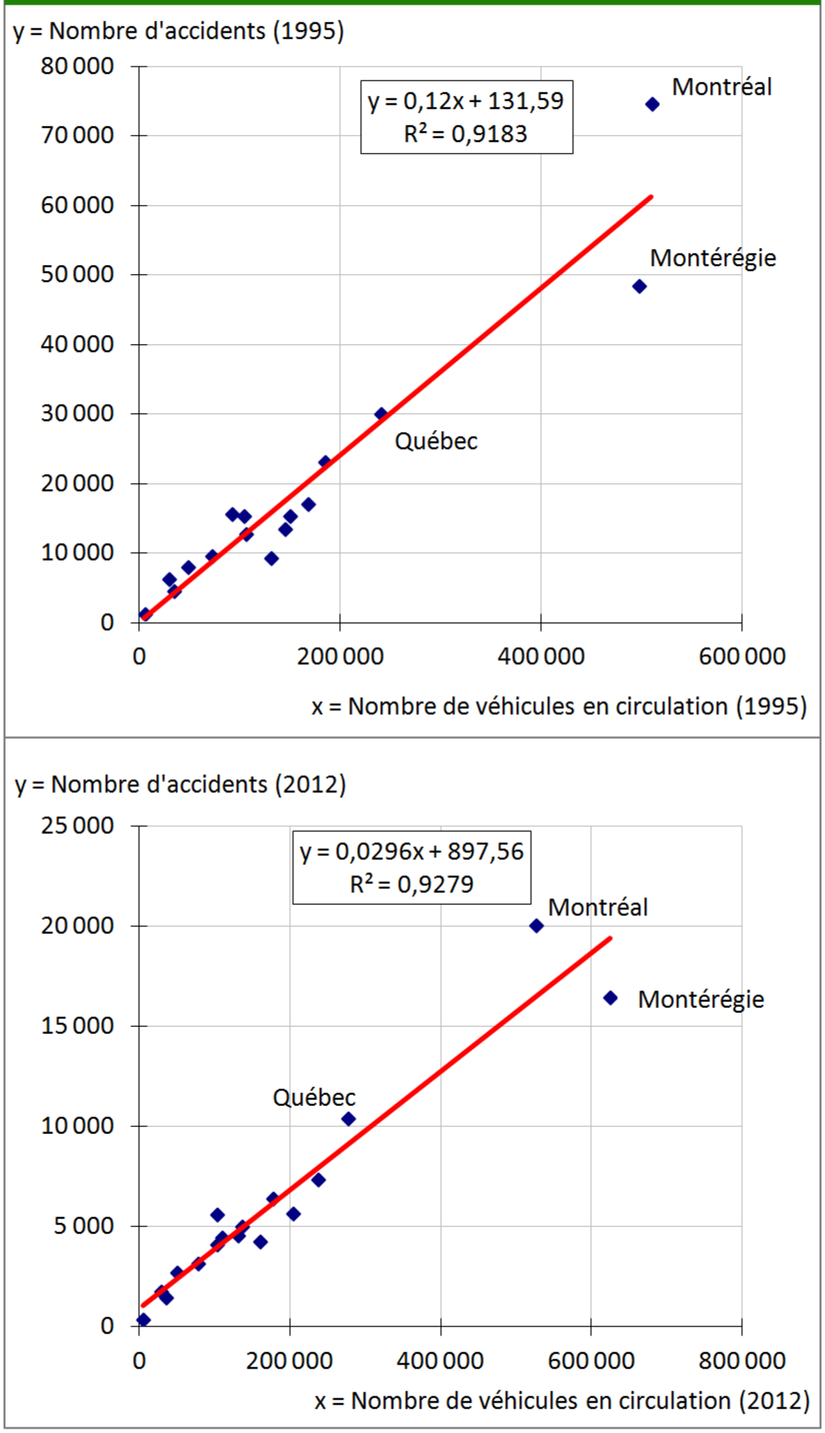
Dans le tableau 9.6, avez-vous remarqué la situation anormale de la Chine qui compte 25 fois moins de blessés mais plus de tués que les États-Unis (dans ce dernier cas, s'agit-il de piétons?). Cela pourrait bien être attribuable en bonne partie à la façon dont chaque pays définit chaque variable : si tout le monde s'entend sur ce qu'est un tué, il en va autrement pour les notions d'accident et de blessé. Nous éviterons ce genre d'écueil en restreignant maintenant notre étude au Québec.

4.2. La situation dans les régions du Québec

Dans la figure 9.12, nous avons choisi de relier la variable *nombre de véhicules* à la variable *nombre d'accidents* en examinant la situation dans les diverses régions du Québec. La figure est accompagnée d'une droite de régression, entièrement construite avec un chiffrier électronique et sans aucun calcul. La relation entre les deux variables saute aux yeux. Le coefficient de corrélation (*r*) est très proche de 1, ce qui signifie que la corrélation est très forte. Il reste néanmoins que le nombre de véhicules n'explique pas entièrement le nombre d'accidents. Il vaudrait la peine de chercher à isoler les autres variables.

Il existe une explication relativement évidente à la situation de Montréal et de la Montérégie, dont les points s'écartent étrangement — et symétriquement — de la droite de régression. De nombreux conducteurs de la Montérégie se rendent chaque jour à Montréal. En même temps que les automobilistes, les risques d'accident se déplacent alors d'une région à l'autre. Le même phénomène entre les villes-centres et leur grande banlieue se remarque, à moins grande échelle, dans la région de Québec en 2012.

FIGURE 9.12 - Accidents et parc automobile au Québec



(Données de la figure 9.12 - Accidents et parc automobile au Québec)

	Nombre de véhicules de promenade en circulation		Nombre total d'accidents		
	1995	**2012**	**1995**	**2012**	
Gaspésie-Îles de la Madeleine	33 875	36 424	4 454	1 494	Gaspésie-Îles de la Madeleine
Bas-Saint-Laurent	72 234	78 641	9 542	3 195	Bas-Saint-Laurent
Saguenay-Lac-Saint-Jean	92 473	103 640	15 608	5 618	Saguenay-Lac-Saint-Jean
Québec	240 644	277 558	29 957	10 413	Capitale-Nationale
Chaudière-Appalaches	149 742	178 067	15 343	6 421	Chaudière-Appalaches
Mauricie-Bois-Francs	184 809	110 631	23 091	4 451	Mauricie
		103 647		4 112	Centre-du-Québec
Estrie	106 090	131 180	12 759	4 561	Estrie
Montérégie	497 115	624 682	48 351	16 451	Montérégie
Montréal	509 606	526 548	74 647	20 090	Montréal
Laval	130 761	160 481	9 250	4 277	Laval
Lanaudières	145 225	205 169	13 473	5 658	Lanaudières
Laurentides	167 435	237 580	17 088	7 371	Laurentides
Outaouais	104 441	137 078	15 313	5 049	Outaouais
Abitibi-Témiscamingue	48 260	50 272	7 886	2 725	Abitibi-Témiscamingue
Côte-Nord	29 082	29 339	6 162	1 791	Côte-Nord
Nord du Québec	4 995	4 912	1 236	393	Nord du Québec
Non précisé	928	2 641			Non précisé
Total	**2 517 715**	**2 998 490**	**304 160**	**104 070**	**Total**

Source des données brutes : Société de l'Assurance automobile du Québec, Bilan 1995, Bilan statistique 2012.

Dans les *Cahiers de géographie du Québec* d'avril 1996, des chercheurs ont utilisé des variables plus raffinées pour étudier la situation : la densité d'accidents (nombre d'accidents au km²), le taux de motorisation (nombre de véhicules de promenade par habitant), la densité de la population (nombre d'habitants par km²) et l'encadrement policier (nombre de policiers pour 1000 habitants). On remarque que ces quatre variables sont en réalité des rapports, ce qui permet de comparer des régions dont les caractéristiques sont différentes. Les chercheurs ont alors calculé la valeur de ces variables pour chacune des municipalités régionales de comté du Québec et ont ensuite cherché à établir des corrélations.

Cette étude montre qu'au Québec, la densité d'accidents est directement reliée à la densité de la population. L'équation de régression qui relie ces deux variables est la suivante :

$$\text{Densité d'accidents} = 0{,}0084 + 0{,}0053 \text{ Habitants/km}^2$$

Selon cette équation, une région qui compterait 10 habitants/km² aurait une densité annuelle d'accidents de 0,0084 + (10 x 0,0053) = 0,0084 + 0,053 = 0,0614 accident/km². Le coefficient de corrélation est de 0,99, ce qui indique que les deux variables se suivent de très près. Lorsque l'on croise les variables *densité d'accident* et *encadrement policier*, la corrélation est beaucoup plus faible

(r = –0,37) et les variables évoluent en sens inverse. Même si la présence policière semble calmer les chauffards, son influence semble relativement faible par rapport à d'autres variables. Notons enfin que des calculs similaires effectués dans les 52 districts de gendarmerie de la Belgique donnent des résultats très semblables. Ce genre d'étude peut être particulièrement utile lorsqu'il s'agit de mettre en place une politique de sécurité routière.

4.3. Qui conduit le mieux?

Nous nous demandions un peu plus haut si le sexe ou l'âge peuvent exercer une influence sur le nombre d'accidents. Le tableau 9.7 montre qu'il y a *relativement* moins de femmes que d'hommes qui se trouvent impliqués dans des accidents au Québec. Mais ici encore, il faut être prudent avant de conclure que les femmes conduisent mieux (ou se soûlent moins) que les hommes. On devrait se poser les questions suivantes : lequel des deux sexes fait le plus de kilomètres dans l'année? Lequel roule le plus la nuit? Lequel fréquente le plus souvent des routes de campagne? Etc.?

Tableau 9.7 - Les accidents selon le sexe du conducteur au Québec

(1995)	**Accident**	**Pas d'accident**	**Total**
Femmes	77 502	1 897 904	1 975 406
Hommes	180 136	2 162 022	2 342 158
Total	**257 638**	**4 059 926**	**4 317 564**
Khi²	27 110		
V de Cramer	0,079		

Source des données brutes : Société de l'Assurance automobile du Québec, Bilan 1995.

(2012)	**Accident**	**Pas d'accident**	**Total**
Femmes	45 351	2 441 185	2 486 536
Hommes	72 003	2 636 221	2 708 224
Total	**117 354**	**5 077 406**	**5 194 760**
Khi²	4 092		
V de Cramer	0,028		

Source des données brutes : Société de l'Assurance automobile du Québec, Bilan 2012.

Dans le tableau 9.7, on constate à nouveau que le nombre d'accidents a tendance à baisser fortement au Québec sur le long terme, encore plus pour les hommes que pour les femmes. Cela dit, la très grande majorité des conducteurs n'ont pas eu d'accident au cours de l'année considérée.

Dans le tableau 9.8, on constate que les jeunes conducteurs ont *relativement* plus d'accidents que leurs aînés. Mais là encore, le phénomène peut être influencé par plusieurs variables : l'expérience du conducteur, son attitude, l'état des routes qu'il fréquente particulièrement et l'âge du véhicule.

Tableau 9.8 - Les accidents selon l'âge du conducteur au Québec

(1995)	**Accident**	**Pas d'accident**	**Total**
16-19 ans	23 346	150 221	173 567
20-24 ans	32 903	312 172	345 075
25-35 ans	64 626	903 727	968 353
Total	**120 875**	**1 366 120**	**1 486 995**
Khi²	10 241		
V de Cramer	0,083		

Source des données brutes : Société de l'Assurance automobile du Québec, Bilan 1995.

(2012)	**Accident**	**Pas d'accident**	**Total**
16-19 ans	12 207	140 845	153 052
20-24 ans	18 826	344 110	362 936
25-34 ans	27 067	810 923	837 990
Total	**58 100**	**1 295 878**	**1 353 978**
Khi²	8 066		
V de Cramer	0,077		

Source des données brutes : Société de l'Assurance automobile du Québec, Bilan 2012.

Pour les deux tableaux précédents, nous avons calculé le Khi carré et le V de Cramer. Étant donné que l'échantillon est très grand (il s'agit en réalité de la population au complet), le Khi carré est très élevé. Nos données sont, par définition, représentatives, aussi il est inutile de faire un test d'hypothèse. Le Khi carré nous sert uniquement à calculer la force de l'association grâce au V de Cramer. On constate, dans le tableau 9.8 que le V de Cramer est de 0,083 en 1995. Cela signifie que si l'âge explique une partie du problème, il est très loin de l'expliquer dans son entier. Bien que faible, le coefficient de Cramer s'avère intéressant si on cherche à comparer la situation de 1995 à celle de 1991 ou de 2012, par exemple. Le V de Cramer était alors de 0,077, ce qui prouve qu'il existe une certaine stabilité dans la relation.

4.4. Des calculs plus simples en disent parfois plus long

Pour tirer les choses au clair, mettons de côté les coefficients compliqués et servons-nous de simples rapports. Selon une enquête concernant la région de l'Outaouais (1995), les hommes au volant sont impliqués 3,5 fois plus souvent que les femmes dans des accidents mortels. Ce rapport baisse à 1,8 pour les accidents avec blessures légères. Par ailleurs, les conducteurs parcourent deux fois plus de distance que les conductrices. Cela démontre qu'à distance égale parcourue, les conducteurs masculins ont plus d'accidents que les femmes, du moins en ce qui concerne les accidents graves.

Pour mieux cerner la relation entre l'âge et la fréquence des accidents, nous avons calculé le nombre d'accidents par rapport au nombre de titulaires de permis de conduire pour différents groupes d'âge. Les chiffres, que nous reproduisons dans le tableau 9.9, parlent d'eux-mêmes (troisième colonne du tableau).

Tableau 9.9 - Rapport entre le nombre d'accidents et le nombre de titulaires de permis selon l'âge

(1995)	Nombre de conducteurs impliqués dans un accident	Nombre de détenteurs de permis de conduire	Rapport Accidents/ Détenteurs de permis
16-19 ans	23 346	173 567	0,135
20-24 ans	32 903	345 075	0,095
25-35 ans	64 626	968 353	0,067
35-44 ans	58 534	1 088 868	0,054
45-54 ans	40 853	830 264	0,049
55-64 ans	21 591	499 919	0,043
65 ans et plus	17 750	403 351	0,044

Source des données brutes : Société de l'Assurance automobile du Québec, Bilan 1995.

(2012)	Nombre de conducteurs impliqués dans un accident	Nombre de détenteurs de permis de conduire	Rapport Accidents/ Détenteurs de permis
16-19 ans	12 207	153 052	0,080
20-24 ans	18 826	362 936	0,052
25-34 ans	27 067	837 990	0,032
35-44 ans	23 709	901 271	0,026
45-54 ans	24 532	1 108 089	0,022
55-64 ans	17 856	960 971	0,019
65 ans et plus	15 948	863 530	0,018

Source des données brutes : Société de l'Assurance automobile du Québec, Bilan 2012.

EXERCICES 4

1. Mon camion c'est ma maison

Commentez le tableau 9.10.

Tableau 9.10 - Types de véhicules dans la région métropolitaine

(1995)	Automobiles	Camions légers	Total
Montérégie	497 115	107 671	**604 786**
Montréal	509 606	71 287	**580 893**
Total	**1 006 721**	**178 958**	**1 185 679**

Source : Société de l'Assurance automobile du Québec, Bilan routier 1995.

Khi²:	7074
V de Cramer:	0,077

(2012)	Automobiles	Camions légers	Total
Montérégie	624 682	255 111	**879 793**
Montréal	532 392	184 394	**716 786**
Total	**1 157 074**	**439 505**	**1 596 579**

Source : Société de l'Assurance automobile du Québec, Bilan routier 2012.

2. Jeune et fou

a) Complétez le tableau 9.11.

Tableau 9.11 - Rapport entre le nombre d'accident et le nombre de titulaires de permis selon l'âge

(1991)	Nombre de conducteurs impliqués dans un accident	Nombre de détenteurs de permis de conduire	Rapport Accidents/ Détenteurs de permis
16-19 ans	25 246	185 762	
20-24 ans	38 884	371 183	
25-34 ans	78 367	1 073 556	
35-44 ans	59 296	1 001 300	
45-54 ans	36 738	681 342	
55-64 ans	21 272	456 939	
65 ans et plus	15 052	314 768	

Source des données brutes : Société de l'Assurance automobile du Québec, Bilan 1995.

b) Tracez sur un graphe la courbe du rapport Accidents/Détenteurs de permis en fonction de l'âge. Tracez une deuxième courbe sur le même graphe en utilisant les données de 1995 (voir tableau 9.9).

c) Comparez la situation de 1991 à celle de 1995 et commentez.

3. Recherche

Mettez à jour les tableaux 9.7, 9.8 et 9.9, et commentez l'évolution des données.

EXERCICES SUPPLÉMENTAIRES

1. Des relations à vérifier

La pauvreté est en hausse aux États-Unis dans les années 1990. Une idée très répandue dans les milieux conservateurs veut que l'augmentation des dépenses de l'aide sociale soit causée par la prolifération des mères adolescentes et des mères célibataires. Dans d'autres milieux, soi-disant progressistes, on affirme souvent que l'instruction n'est plus un moyen efficace pour éviter la pauvreté. Avant de commenter ces affirmations, répondez aux questions *a* et *b* ci-après.

a) Représentez graphiquement les données suivantes : le salaire mensuel moyen est de 508 $ pour ceux qui n'ont pas terminé leurs études secondaires, de 1080 $ pour ceux qui détiennent un diplôme du secondaire, de 1303 $ pour ceux qui ont entamé des études postsecondaires et de 2339 $ pour ceux qui ont terminé des études universitaires.

b) Représentez graphiquement les données suivantes : l'âge médian des mères qui bénéficient de l'aide sociale est de 27,4 ans, 24,5 % d'entre elles ont moins de 21 ans et 7,1 % d'entre elles ont moins de 18 ans.

(Source des données : Scientific American, octobre 1996, données de 1993 et 1995.)

2. Le suicide selon Durkheim

Dans un ouvrage classique intitulé *Le suicide*, Émile Durkheim constate que, si chaque suicide est un phénomène individuel, le taux de suicide d'une population s'avère facilement prévisible à court terme. Entre 1871 et 1875, le taux de suicide (nombre annuel de suicides pour 100 000 habitants) était de 25,5 au Danemark et de 3,5 en Italie. En Suisse, en 1876, il variait de 8,3 à 8,7 chez les catholiques et de 29,3 à 45,6 chez les protestants.

En France, entre 1835 et 1843, ce taux était de 30,6 en été, de 28,3 au printemps, de 21,0 en automne et de 20,1 en hiver. Entre 1848 et 1857, il s'élevait à 4,6 pour les gens de 16 à 21 ans, à 9,8 pour ceux de 21 à 30 ans, à 11,5 pour ceux de 31 à 40 ans et à 16,4 pour ceux de 41 à 50 ans. Entre 1889 et 1891, ce taux de suicide chez les hommes mariés âgés de 26 à 35 ans était de 10,6 contre 25,7 chez les hommes célibataires. Chez les femmes mariées du même groupe d'âge, ces taux étaient respectivement de 2,8 et de 6,1 chez les célibataires. Durkheim note d'autre part que le taux est plus élevé dans les grandes villes que dans les petites, plus élevé le jour que la nuit, plus élevé au début de la semaine qu'à la fin.

Par ailleurs, des données relevées au XX^{e} siècle indiquent que le taux de suicide était au plus bas en 1917 (14,0) et en 1943 (11,3) alors qu'il atteignait un sommet en 1912 (22,9) et en 1934 (21,7).

Au Québec, le taux de suicide passait de 4,2 en 1950-54 à 13,8 en 1975-79. En 1995, le taux de suicide était de 58 en Hongrie, de 30,0 en France, de 34,0 en Suisse, de 12,0 en Espagne et de 20,4 aux États-Unis. Du côté des personnes de 75 ans et plus, le taux de suicide se révèle 12 fois plus élevé pour les hommes que pour les femmes pour atteindre 186,2 en Hongrie et 114,0 en France.

Voici pour terminer quelques comparaisons entre pays pour la période 2001-2010. Le taux de suicide du Canada était de 17,3 contre 17,7 aux États-Unis. Celui du Japon était de 36,2 contre 13,0 en Inde. Celui de Cuba était de 19,0 contre 3,9 en République dominicaine, celui de la Russie était de 53,9 contre 37,8 en Ukraine.

(Sources : Émile Durkheim, Le suicide, 1897, réédition 1960 : Presses universitaires de France. Organisation mondiale de la Santé. Marie-France Charron, Le suicide au Québec, 1983. PNUD, Rapport sur le développement humain 2014.)

a) Construisez un schéma contenant toutes les variables qui influent sur le taux de suicide. Indiquez par des flèches le sens des relations. Lorsque cela est pertinent, accompagnez la flèche d'un signe positif ou négatif pour indiquer si les variables évoluent dans le même sens ou en sens inverse.

b) Recherche. Obtenez d'autres données dans le but de confirmer, de rectifier ou de compléter le schéma de variables que vous venez de tracer.

3. Ex-maltraité, future brute?

Joan Kaufman et Edward Zigler, de l'université de Yale, ont essayé d'établir si la violence parentale envers les enfants se transmettait de génération en génération. Pour ce faire, ils ont utilisé les résultats d'une enquête menée auprès de 282 parents d'enfants admis dans un service de soins intensifs pour enfants. Ils constatèrent que 49 de ces parents avaient été maltraités durant leur enfance et que 10 de leurs enfants furent maltraités dans l'année qui suivit leur visite à l'hôpital. Parmi ces 10 enfants, 9 étaient issus de parents qui avaient été maltraités eux-mêmes. (Source : Sciences humaines, no 65, octobre 1996.)

a) Construisez, à partir des données ci-dessus, un tableau croisé de deux colonnes et de deux lignes.

b) Faites un test pour vérifier l'hypothèse qu'il existe une relation entre la façon dont les parents ont été traités dans leur enfance et la façon dont ils traitent leurs enfants.

c) Commentez l'affirmation suivante : « 90 % des enfants maltraités sont issus de parents eux-mêmes maltraités dans leur enfance, mais seulement 18 % des ex-enfants maltraités deviennent des parents maltraitants ».

4. Deux variables qui se suivent de près

a) Représentez les points du tableau 9.12 sur un graphe.

Tableau 9.12 - Accidents et parc automobile au Québec

(1991)	Nombre de véhicules de promenade en circulation	Nombre total d'accidents
Gaspésie-Îles de la Madeleine	31 865	4 276
Bas-Saint-Laurent	69 853	9 022
Saguenay-Lac-Saint-Jean	91 751	15 647
Québec	237 672	31 446
Chaudière-Appalaches	144 177	14 377
Mauricie-Bois-Francs	182 066	22 567
Estrie	102 461	12 767
Montérégie	488 436	48 390
Montréal	534 938	77 850
Laval	130 294	8 938
Lanaudières	136 142	13 401
Laurentides	155 532	16 661
Outaouais	100 622	16 438
Abitibi-Témiscamingue	46 797	7 555
Côte-Nord	28 348	6 110
Nord du Québec	5 282	1 274
Non précisé	902	
Total	**2 487 138**	**306 719**

Source des données brutes: Société de l'Assurance automobile du Québec, bilan 1995.

b) Commentez le graphe.

c) Si vous avez un chiffrier électronique, essayez de tracer une droite de régression sur le même modèle que la figure 9.12.

5. Un sport dangereux?

Quelque 223 hockeyeurs qui ont été admis à l'urgence de l'hôpital de l'Enfant Jésus de Québec, entre le 1er octobre 1991 et le 30 avril 1992. Le tableau 9.13 indique comment se répartissaient les blessures selon le type de jeu pratiqué et la partie du corps touchée.

Tableau 9.13 - Les blessures au hockey

	Type de blessure				
Type de jeu	**Tête**	**Membre supérieur**	**Membre inférieur**	**Tronc**	**Total**
Hockey organisé (ligue, club) avec patins	20	42	48	20	**130**
Hockey non organisé avec patins	11	10	8	6	**35**
Hockey organisé (ligue, club) sans patins	11	5	14	3	**33**
Hockey non organisé sans patins	11	5	6	3	**25**
Total	**53**	**62**	**76**	**32**	**223**

Source : Revue canadienne de Santé publique, volume 87, nº 4, juillet-août 1996.

Essayez d'établir s'il existe une relation entre le type de hockey pratiqué et le type de blessure reçue. Si nécessaire, réduisez le tableau à deux lignes, en regroupant certaines catégories.

6. Recherche : un tableau croisé

Faites une mini enquête qui vous permettra de construire un tableau croisé. Choisissez deux variables qualitatives comportant chacune de deux à quatre catégories. Constituez un échantillon (aléatoire dans la mesure du possible) et compilez vos résultats dans un tableau croisé. Si nécessaire, augmentez la taille de votre échantillon pour éviter que certaines cases du tableau ne contiennent des fréquences inférieures à 5. Calculez le Khi carré et le V de Cramer et commentez les résultats.

7. Recherche : l'héritabilité du poids

Dans cette mini enquête, vous chercherez à mesurer la corrélation entre le poids (ou la taille) des pères et de leurs fils, ou des mères et de leurs filles.

Prenez une balance (ou un mètre) et promenez-vous dans votre quartier. Pesez les paires (parent-enfant) tout en ménageant leur susceptibilité. Notez leur âge et vérifiez si vous êtes en présence d'un parent biologique ou d'un parent adoptif. Pour éliminer l'influence de la variable *âge*, prenez uniquement des enfants adultes ou, à la rigueur, des enfants d'un âge précis.

Tracez un nuage de points. Tracez la courbe de tendance. Calculez le coefficient de corrélation.

Que proposez-vous pour améliorer l'échantillon choisi?

DOSSIER 9 UN TABOU UNIVERSEL

L'inceste, ou union de deux êtres du même sang, est un des derniers tabous de notre société « post-moderne ». C'est aussi un tabou universel, puisqu'on le retrouve pratiquement dans toutes les sociétés depuis la nuit des temps. Dans ce domaine, comme c'est souvent le cas en sciences humaines, les partisans de l'*inné* s'opposent aux partisans de l'*acquis*. Pour les premiers, l'horreur de l'inceste est inscrite dans les gènes des humains, comme dans celui d'autres animaux. Pour les seconds, au contraire, l'obligation du mariage à l'extérieur du clan vient du besoin de maintenir la paix (en créant des liens familiaux avec des ennemis potentiels) et de préserver la santé de l'espèce. Quoi qu'il en soit, la nature, dans sa bienveillance, ou la société, dans sa sagesse, auraient, en créant ce tabou, évité à la race humaine de dégénérer. Car il est largement accepté que la consanguinité est source de déficiences physiques ou mentales.

L'étude de Seemenova, effectuée en ex-Tchécoslovaquie auprès de 256 enfants tend à confirmer cette croyance, du moins à première vue. Mais si on prend la peine d'approfondir un peu la question, on se rend compte que la relation de cause à effet entre consanguinité et dégénérescence est loin d'être évidente.

Selon l'étude de Seemenova, la *proportion* d'enfants ayant des troubles physiques ou mentaux est considérablement plus élevée chez les enfants nés d'un inceste (32/161) que chez les autres enfants (2/95). La même remarque s'applique à la mortalité infantile (voir le tableau D9.1).

Tableau D9.1 - Proximité génétique et déficiences physiques ou mentales

Échantillon	**Enfants nés de l'inceste**	**Groupe témoin**
Femme avec son père	88	
Femme avec son frère	72	
Femme avec son fils	1	
Total	**161**	**95**

Pathologie	**Nombre d'enfants ayant des troubles physiques ou mentaux**	**Nombre d'enfants morts dans leur première année**
Parents incestueux	32	15
Parents du groupe témoin	**2**	**5**

Source: Recherche effectuée en Tchécoslovaquie par Seemenova, dans in Jacques-Dominique de LANNOY et Pierre FEYEERIESEN, L'inceste, PUF, 1992.

Il y a donc une relation apparente entre la consanguinité des parents et les déficiences des enfants. Mais quelles sont les chances que les différences observées entre les deux groupes d'enfants soient dues au hasard de l'échantillonnage? C'est ce que nous allons vérifier à l'aide du calcul du Khi carré.

Calcul des fréquences théoriques

Compte tenu de l'échantillon choisi, le tableau D9.2 montre que les différences entre les fréquences observées et les fréquences théoriques sont relativement faibles (voir tableaux D9.2a et D9.2b). Nous avons calculé les fréquences théoriques de la façon suivante : s'il n'y avait aucune association entre les variables, 20/256 des 161 enfants nés de parents incestueux devraient être morts dans leur première année et 236/256 de ces 161 enfants devraient avoir survécu. Le chiffre figurant dans la case supérieure gauche du tableau D9.2b est donc de 20/256 x 161 = 12,6.

TABLEAU D9.2 - Consanguinité et déficiences

Tableau D9.2a - Fréquences observées

	Enfants morts dans leur première année	Autres enfants	Total
Parents incestueux	15	146	**161**
Autres parents	5	90	**95**
Total	**20**	**236**	**256**

Tableau D9.2b - Fréquences théoriques

	Enfants morts dans leur première année	Autres enfants	Total
Parents incestueux	12,6	148,4	**161,0**
Autres parents	7,4	87,6	**95,0**
Total	**20,0**	**236,0**	**256,0**

Tableau D9.2c - Écarts au carré relatifs

	Enfants morts dans leur première année	Autres enfants	Total
Parents incestueux	0,47	0,04	**0,51**
Autres parents	0,79	0,07	**0,86**
Total	**1,26**	**0,11**	**1,36**

Tableau D9.2d - Proportion observées (en %)

	Enfants morts dans leur première année	Autres enfants	Total
Parents incestueux	5,9	57,0	**62,9**
Autres parents	2,0	35,2	**37,1**
Total	**7,8**	**92,2**	**100,0**

Tableau D9.2e - Proportion théoriques (en %)

	Enfants morts dans leur première année	Autres enfants	Total
Parents incestueux	4,9	58,0	**62,9**
Autres parents	2,9	34,2	**37,1**
Total	**7,8**	**92,2**	**100,0**

1	Degrés de liberté
0,05	Seuil de signification
3,841	Valeur critique du Khi²
0,243	Probabilité de dépasser le Khi² calculé

Note : Les chiffres affichés sont arrondis, mais les calculs sont effectués avec un degré de précision maximum.

Calcul de l'écart global : le Khi carré

Nous avons ensuite calculé les écarts au carré relatifs entre les fréquences théoriques et observées. Pour la case supérieure gauche, on obtient $(12{,}6 - 15)^2/12{,}6 = 2{,}42/12{,}6 = 5{,}76/12{,}6 = 0{,}47$. Pour l'ensemble du tableau, ces écarts atteignent un total de 1,36 : il s'agit du Khi carré calculé.

Interprétation du Khi carré calculé

S'il n'existait aucune association entre les variables étudiées (Consanguinité et Déficiences), les fréquences observées devraient normalement correspondre aux fréquences théoriques. Mais comme nous travaillons avec un échantillon, il se peut que celui-ci s'écarte plus ou moins de sa distribution théorique par le simple jeu du hasard. Si l'écart global mesuré (ici le Khi carré) est trop faible, on peut alors juger qu'il n'est pas significatif. Par contre, si cet écart est suffisamment fort, on ne peut plus accuser le hasard : la présence d'un écart fort montre qu'il y a une association entre les variables. Il nous reste maintenant à chiffrer précisément ce que nous entendons par écart *faible* et *fort*.

En prenant le risque de nous tromper 5 fois sur 100, nous pourrons affirmer qu'il y a une association entre les variables si le Khi carré calculé dépasse 3,84.

Lorsque l'on consulte une table de distribution du Khi carré, on constate que la *valeur critique* du Khi carré pour un *seuil de signification* de 0,05 et avec 1 *degré de liberté* est de 3,84*. Voici ce que ce jargon signifie : parmi tous les échantillons que nous aurions pu recueillir, seulement 5 % (ou 0,05) d'entre eux ont un écart global (ou Khi carré calculé) supérieur à 3,84.

Rappelons que le nombre de degrés de liberté correspond au nombre de cases à l'intérieur du tableau à l'exclusion des cases de la dernière colonne et de la dernière ligne. Puisque notre tableau comprend ici 2 lignes et 2 colonnes, il possède 1 degré de liberté.

Est-ce que l'écart que nous avons observé correspond à nos attentes?

L'écart global de notre tableau (le Khi carré calculé, qui valait 1,36) est plus faible que le minimum attendu (la valeur trouvée dans la table de Khi carré, qui était de 3,84). La possibilité que l'écart observé soit le fruit du hasard est trop forte (ou n'est pas assez faible) pour que nous puissions affirmer qu'il existe une association significative entre les variables *Consanguinité* et *Mortalité infantile*.

Il existe une association significative entre *consanguinité* et *troubles physiques et mentaux*.

Si on répète le même exercice pour tester l'hypothèse d'une association entre *consanguinité* et *troubles physiques et mentaux*, on constate que la valeur du Khi carré calculé est nettement plus significative : elle est de 16,38. La probabilité de tomber sur un tel échantillon uniquement par le fait du hasard est à toutes fins pratiques nulle (cette probabilité est de 0,005 % plus précisément). Nous avons obtenu ces derniers résultats en quelques secondes : le tableau D9.2 ayant été créé avec un chiffrier électronique, il nous a suffi de modifier les 4 fréquences observées pour obtenir instantanément le nouveau Khi carré calculé. Dans le même ordre d'idée, la valeur critique du Khi carré a été obtenue par une simple fonction du chiffrier : il est rare aujourd'hui que des chercheurs utilisent encore des outils aussi primitifs que des tables de distribution.

Maintenant que nous savons que l'écart obtenu n'est pas dû au hasard, il reste à en évaluer l'ampleur en tenant compte de la taille de l'échantillon. Pour cela, nous calculons le V de Cramer, selon la formule donnée dans le chapitre.

$$\text{V de Cramer} = \sqrt{\frac{Khi^2}{n \times (K-1)}}$$

V de Cramer = √[16,37 / (256 x 1)] = √[0,064] = 0,253

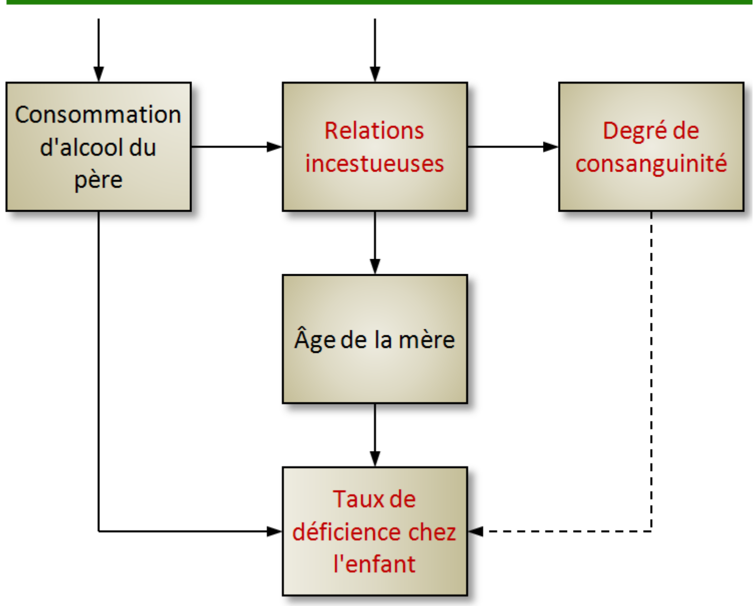
En fin de compte, l'association entre *consanguinité* et *troubles physiques et mentaux* est bien réelle, même si elle n'est pas extrêmement forte. Cela est d'autant plus significatif que nous n'avons pas été en mesure de prouver l'existence d'une association entre *consanguinité* et *mortalité infantile*.

L'association entre *consanguinité des parents* et *déficiences des enfants* semble donc démontrée et l'affaire aurait pu être classée si on n'avait pas eu la curiosité de mieux connaître les parents incestueux. On constata alors deux faits troublants : comparativement aux parents du groupe témoin, les pères incestueux étaient plus facilement alcooliques, et les mères incestueuses accouchaient à un âge significativement plus précoce (voir le tableau D9.3). Or, on sait que ces deux facteurs augmentent les risques de déficiences (quand les parents boivent, les enfants trinquent). Il pourrait alors n'y avoir qu'une simple corrélation entre *inceste* et *déficience*, ces deux variables dépendant (en partie) de variables communes (Consommation d'alcool du père, Âge précoce de la mère). Le schéma de la figure D9.1 tente de mettre en évidence plusieurs aspects de cette corrélation.

Tableau D9.3 - Autres facteurs pouvant influencer les variables

Autres facteurs	**Alcooliques chroniques**	**Âge de la mère à l'accouchement**
Parents incestueux	13	Plus précoce
Autres parents	2	Moins précoce

FIGURE D9.1 - Schéma de variables



Si l'on tient compte des facteurs socio-économiques reliés à la situation, on s'aperçoit que la relation entre consanguinité et déficience est loin d'être prouvée. Les cas d'inceste provenaient souvent de familles défavorisées en ce qui concerne les conditions de logement et d'alimentation, l'état de santé de la mère et la connaissance des soins à apporter aux enfants. Or, tous ces facteurs peuvent avoir une influence directe sur la santé des enfants aussi bien que sur le comportement des parents.

L'inceste est un sujet difficile à traiter. Tout d'abord, l'inceste suscite un dégoût « naturel » chez la plupart des gens. D'autre part, il est souvent associé à l'abus sexuel d'un adulte envers un enfant. Enfin, un certain nombre de préjugés pseudo-scientifiques circulent à son sujet. Mais comme tout sujet relié à l'être humain, il mérite d'être étudié avec sérieux et rigueur.

QUESTIONS

1. Les pères incestueux sont-ils des alcooliques?

Pour répondre aux questions suivantes, utilisez les données des tableaux D9.3 et D9.4.

a) Quelle est la proportion d'alcooliques chroniques dans chacune des catégories de pères (pères incestueux et pères du groupe témoin)?

b) Est-il possible que les différences observées entre ces catégories soient uniquement attribuables au hasard? Comment peut-on évaluer la probabilité d'une telle éventualité?

c) Complétez le tableau D9.4.

d) Calculez le Khi carré et interprétez.

TABLEAU D9.4 - Inceste et alcoolisme

Tableau D9.4a - Fréquences observées

	Alcoolisme chronique	Non alcoolisme chronique	Total
Pères incestueux	13	148	
Autres pères	2	93	
Total			

Tableau D9.4b - Fréquences théoriques

	Alcoolisme chronique	Non alcoolisme chronique	Total
Pères incestueux	9		
Autres pères			
Total			

Tableau D9.4c - Écarts au carré relatifs

	Alcoolisme chronique	Non alcoolisme chronique	Total
Pères incestueux	1,35		
Autres pères			
Total	**3,63**		

Tableau D9.4d - Proportion observées (en %)

	Alcoolisme chronique	Non alcoolisme chronique	Total
Pères incestueux			
Autres pères			**37,1%**
Total			**100,0%**

Tableau D9.4e - Proportion théoriques (en %)

	Alcoolisme chronique	Non alcoolisme chronique	Total
Pères incestueux	3,7%		
Autres pères			**37,1%**
Total			**100,0%**

	Degrés de liberté
0,05	Seuil de signification
	Valeur critique du Khi²
	Probabilité de dépasser le Khi² calculé

2. Schéma de variables

Complétez le schéma proposé à la figure 1 en ajoutant quelques cases (variables) et quelques flèches (relations).

TABLES DE DISTRIBUTION

Distribution normale
Distribution de Student
Distribution du Khi carré
Indice des prix à la consommation (IPC)

TABLE DE DISTRIBUTION NORMALE

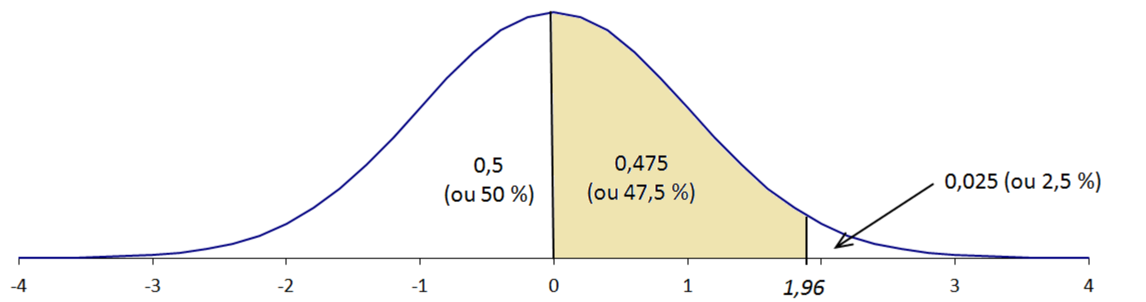


Table de distribution normale

Les valeurs en bleu dans la table représentent l'aire comprise entre 0 et la cote z.

Cote z	Aire	Cote z	Aire	Cote z	Aire	Cote z	Aire	Cote z	Aire	Cote z	Aire	Cote z	Aire	Cote z	Aire
0,00	0,0000	0,50	0,1915	1,00	0,3413	1,50	0,4332	2,00	0,4772	2,50	0,4938	3,00	0,4987	3,50	0,49977
0,01	0,0040	0,51	0,1950	1,01	0,3438	1,51	0,4345	2,01	0,4778	2,51	0,4940	3,01	0,4987	3,51	0,49978
0,02	0,0080	0,52	0,1985	1,02	0,3461	1,52	0,4357	2,02	0,4783	2,52	0,4941	3,02	0,4987	3,52	0,49978
0,03	0,0120	0,53	0,2019	1,03	0,3485	1,53	0,4370	2,03	0,4788	2,53	0,4943	3,03	0,4988	3,53	0,49979
0,04	0,0160	0,54	0,2054	1,04	0,3508	1,54	0,4382	2,04	0,4793	2,54	0,4945	3,04	0,4988	3,54	0,49980
0,05	0,0199	0,55	0,2088	1,05	0,3531	1,55	0,4394	2,05	0,4798	2,55	0,4946	3,05	0,4989	3,55	0,49981
0,06	0,0239	0,56	0,2123	1,06	0,3554	1,56	0,4406	2,06	0,4803	2,56	0,4948	3,06	0,4989	3,56	0,49981
0,07	0,0279	0,57	0,2157	1,07	0,3577	1,57	0,4418	2,07	0,4808	2,57	0,4949	3,07	0,4989	3,57	0,49982
0,08	0,0319	0,58	0,2190	1,08	0,3599	1,58	0,4429	2,08	0,4812	2,58	0,4951	3,08	0,4990	3,58	0,49983
0,09	0,0359	0,59	0,2224	1,09	0,3621	1,59	0,4441	2,09	0,4817	2,59	0,4952	3,09	0,4990	3,59	0,49983
0,10	0,0398	0,60	0,2257	1,10	0,3643	1,60	0,4452	2,10	0,4821	2,60	0,4953	3,10	0,4990	3,60	0,49984
0,11	0,0438	0,61	0,2291	1,11	0,3665	1,61	0,4463	2,11	0,4826	2,61	0,4955	3,11	0,4991	3,61	0,49985
0,12	0,0478	0,62	0,2324	1,12	0,3686	1,62	0,4474	2,12	0,4830	2,62	0,4956	3,12	0,4991	3,62	0,49985
0,13	0,0517	0,63	0,2357	1,13	0,3708	1,63	0,4484	2,13	0,4834	2,63	0,4957	3,13	0,4991	3,63	0,49986
0,14	0,0557	0,64	0,2389	1,14	0,3729	1,64	0,4495	2,14	0,4838	2,64	0,4959	3,14	0,4992	3,64	0,49986
0,15	0,0596	0,65	0,2422	1,15	0,3749	1,65	0,4505	2,15	0,4842	2,65	0,4960	3,15	0,4992	3,65	0,49987
0,16	0,0636	0,66	0,2454	1,16	0,3770	1,66	0,4515	2,16	0,4846	2,66	0,4961	3,16	0,4992	3,66	0,49987
0,17	0,0675	0,67	0,2486	1,17	0,3790	1,67	0,4525	2,17	0,4850	2,67	0,4962	3,17	0,4992	3,67	0,49988
0,18	0,0714	0,68	0,2517	1,18	0,3810	1,68	0,4535	2,18	0,4854	2,68	0,4963	3,18	0,4993	3,68	0,49988
0,19	0,0753	0,69	0,2549	1,19	0,3830	1,69	0,4545	2,19	0,4857	2,69	0,4964	3,19	0,4993	3,69	0,49989

0,20	0,0793	0,70	0,2580	1,20	0,3849	1,70	0,4554	2,20	0,4861	2,70	0,4965	3,20	0,4993	3,70	0,49989
0,21	0,0832	0,71	0,2611	1,21	0,3869	1,71	0,4564	2,21	0,4864	2,71	0,4966	3,21	0,4993	3,71	0,49990
0,22	0,0871	0,72	0,2642	1,22	0,3888	1,72	0,4573	2,22	0,4868	2,72	0,4967	3,22	0,4994	3,72	0,49990
0,23	0,0910	0,73	0,2673	1,23	0,3907	1,73	0,4582	2,23	0,4871	2,73	0,4968	3,23	0,4994	3,73	0,49990
0,24	0,0948	0,74	0,2704	1,24	0,3925	1,74	0,4591	2,24	0,4875	2,74	0,4969	3,24	0,4994	3,74	0,49991
0,25	0,0987	0,75	0,2734	1,25	0,3944	1,75	0,4599	2,25	0,4878	2,75	0,4970	3,25	0,4994	3,75	0,49991
0,26	0,1026	0,76	0,2764	1,26	0,3962	1,76	0,4608	2,26	0,4881	2,76	0,4971	3,26	0,4994	3,76	0,49992
0,27	0,1064	0,77	0,2794	1,27	0,3980	1,77	0,4616	2,27	0,4884	2,77	0,4972	3,27	0,4995	3,77	0,49992
0,28	0,1103	0,78	0,2823	1,28	0,3997	1,78	0,4625	2,28	0,4887	2,78	0,4973	3,28	0,4995	3,78	0,49992
0,29	0,1141	0,79	0,2852	1,29	0,4015	1,79	0,4633	2,29	0,4890	2,79	0,4974	3,29	0,4995	3,79	0,49992
0,30	0,1179	0,80	0,2881	1,30	0,4032	1,80	0,4641	2,30	0,4893	2,80	0,4974	3,30	0,4995	3,80	0,49993
0,31	0,1217	0,81	0,2910	1,31	0,4049	1,81	0,4649	2,31	0,4896	2,81	0,4975	3,31	0,4995	3,81	0,49993
0,32	0,1255	0,82	0,2939	1,32	0,4066	1,82	0,4656	2,32	0,4898	2,82	0,4976	3,32	0,4995	3,82	0,49993
0,33	0,1293	0,83	0,2967	1,33	0,4082	1,83	0,4664	2,33	0,4901	2,83	0,4977	3,33	0,4996	3,83	0,49994
0,34	0,1331	0,84	0,2995	1,34	0,4099	1,84	0,4671	2,34	0,4904	2,84	0,4977	3,34	0,4996	3,84	0,49994
0,35	0,1368	0,85	0,3023	1,35	0,4115	1,85	0,4678	2,35	0,4906	2,85	0,4978	3,35	0,4996	3,85	0,49994
0,36	0,1406	0,86	0,3051	1,36	0,4131	1,86	0,4686	2,36	0,4909	2,86	0,4979	3,36	0,4996	3,86	0,49994
0,37	0,1443	0,87	0,3078	1,37	0,4147	1,87	0,4693	2,37	0,4911	2,87	0,4979	3,37	0,4996	3,87	0,49995
0,38	0,1480	0,88	0,3106	1,38	0,4162	1,88	0,4699	2,38	0,4913	2,88	0,4980	3,38	0,4996	3,88	0,49995
0,39	0,1517	0,89	0,3133	1,39	0,4177	1,89	0,4706	2,39	0,4916	2,89	0,4981	3,39	0,4997	3,89	0,49995
0,40	0,1554	0,90	0,3159	1,40	0,4192	1,90	0,4713	2,40	0,4918	2,90	0,4981	3,40	0,4997	3,90	0,49995
0,41	0,1591	0,91	0,3186	1,41	0,4207	1,91	0,4719	2,41	0,4920	2,91	0,4982	3,41	0,4997	3,91	0,49995
0,42	0,1628	0,92	0,3212	1,42	0,4222	1,92	0,4726	2,42	0,4922	2,92	0,4982	3,42	0,4997	3,92	0,49996
0,43	0,1664	0,93	0,3238	1,43	0,4236	1,93	0,4732	2,43	0,4925	2,93	0,4983	3,43	0,4997	3,93	0,49996
0,44	0,1700	0,94	0,3264	1,44	0,4251	1,94	0,4738	2,44	0,4927	2,94	0,4984	3,44	0,4997	3,94	0,49996
0,45	0,1736	0,95	0,3289	1,45	0,4265	1,95	0,4744	2,45	0,4929	2,95	0,4984	3,45	0,4997	3,95	0,49996
0,46	0,1772	0,96	0,3315	1,46	0,4279	**1,96**	**0,4750**	2,46	0,4931	2,96	0,4985	3,46	0,4997	3,96	0,49996
0,47	0,1808	0,97	0,3340	1,47	0,4292	1,97	0,4756	2,47	0,4932	2,97	0,4985	3,47	0,4997	3,97	0,49996
0,48	0,1844	0,98	0,3365	1,48	0,4306	1,98	0,4761	2,48	0,4934	2,98	0,4986	3,48	0,4997	3,98	0,49997
0,49	0,1879	0,99	0,3389	1,49	0,4319	1,99	0,4767	2,49	0,4936	2,99	0,4986	3,49	0,4998	3,99	0,49997

TABLE DE DISTRIBUTION DE STUDENT

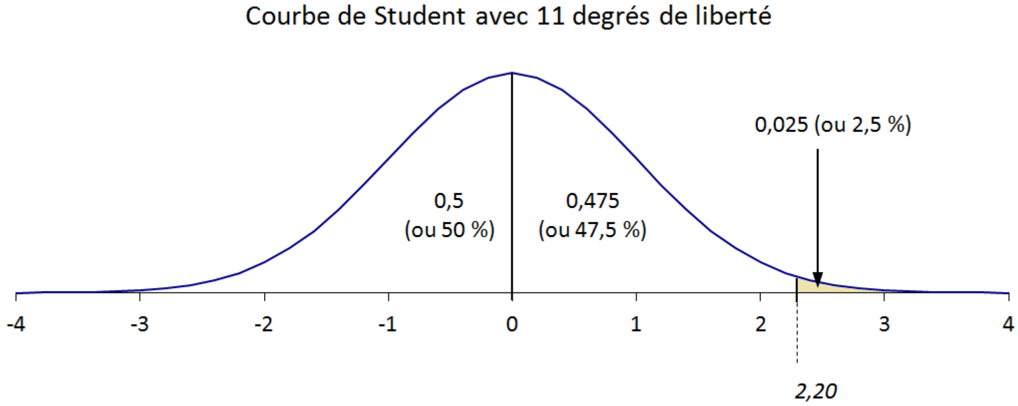


Table de distribution de Student					
Degrés de liberté	Surface couverte à droite du nombre figurant à l'intérieur de la table				
	0,1	0,05	**0,025**	0,01	0,005
1	3,078	6,314	12,71	31,82	63,66
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	**2,201**	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779

27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
31	1,309	1,696	2,040	2,453	2,744
32	1,309	1,694	2,037	2,449	2,738
33	1,308	1,692	2,035	2,445	2,733
34	1,307	1,691	2,032	2,441	2,728
35	1,306	1,690	2,030	2,438	2,724
36	1,306	1,688	2,028	2,434	2,719
37	1,305	1,687	2,026	2,431	2,715
38	1,304	1,686	2,024	2,429	2,712
39	1,304	1,685	2,023	2,426	2,708
40	1,303	1,684	2,021	2,423	2,704
45	1,301	1,679	2,014	2,412	2,690
60	1,296	1,671	2,000	2,390	2,660
120	1,289	1,658	1,980	2,358	2,617

TABLE DE DISTRIBUTION DU KHI CARRÉ

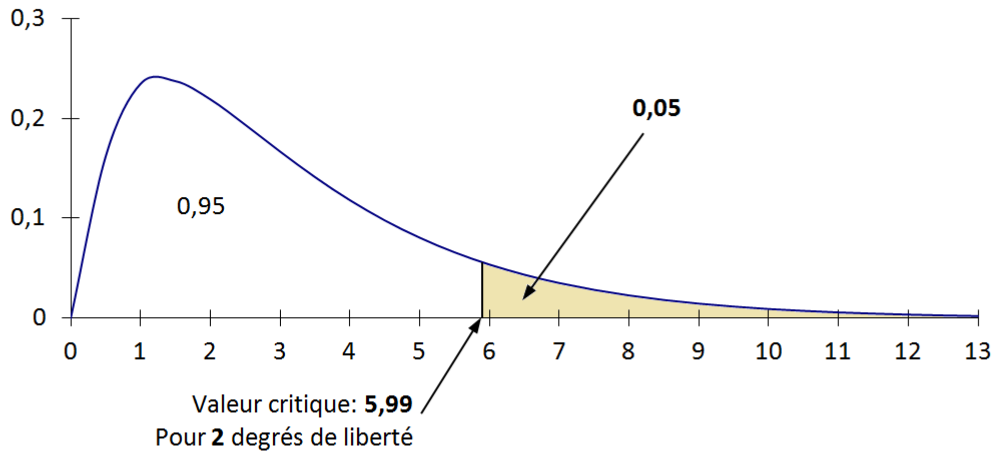
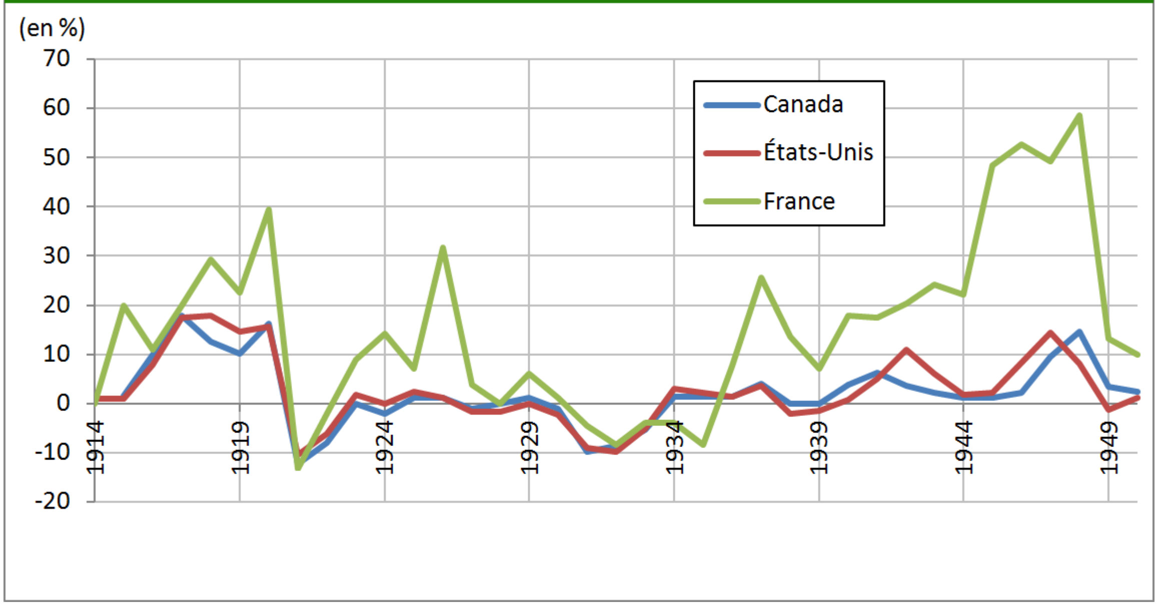


Table de distribution du Khi carré

Degrés de liberté	Seuil de signification 0,1	**0,05**	0,025	0,01	0,005	0,001
1	2,706	3,841	5,024	6,635	7,879	10,828
2	4,605	**5,991**	7,378	9,210	10,597	13,816
3	6,251	7,815	9,348	11,345	12,838	16,266
4	7,779	9,488	11,143	13,277	14,860	18,467
5	9,236	11,070	12,833	15,086	16,750	20,515
6	10,645	12,592	14,449	16,812	18,548	22,458
7	12,017	14,067	16,013	18,475	20,278	24,322
8	13,362	15,507	17,535	20,090	21,955	26,124
9	14,684	16,919	19,023	21,666	23,589	27,877
10	15,987	18,307	20,483	23,209	25,188	29,588
11	17,275	19,675	21,920	24,725	26,757	31,264
12	18,549	21,026	23,337	26,217	28,300	32,909
13	19,812	22,362	24,736	27,688	29,819	34,528
14	21,064	23,685	26,119	29,141	31,319	36,123
15	22,307	24,996	27,488	30,578	32,801	37,697
16	23,542	26,296	28,845	32,000	34,267	39,252
17	24,769	27,587	30,191	33,409	35,718	40,790
18	25,989	28,869	31,526	34,805	37,156	42,312
19	27,204	30,144	32,852	36,191	38,582	43,820
20	28,412	31,410	34,170	37,566	39,997	45,315

INDICE DES PRIX À LA CONSOMMATION

Évolution des taux d'inflation annuels entre les deux guerres



Évolution des prix à la consommation

	Indice des prix à la consommation			Taux annuel d'inflation		
	Canada	États-Unis	France	Canada	États-Unis	France
	2002=100	1982-84=100	1998=100			
1913		9,9	0,058			
1914	6,0	10,0	0,058		1,0	0,0
1915	6,1	10,1	0,070	1,7	1,0	19,9
1916	6,7	10,9	0,077	9,8	7,9	11,0
1917	7,9	12,8	0,093	17,9	17,4	20,0
1918	8,9	15,1	0,120	12,7	18,0	29,2
1919	9,8	17,3	0,147	10,1	14,6	22,5
1920	11,4	20,0	0,205	16,3	15,6	39,5
1921	10,0	17,9	0,178	-12,3	-10,5	-13,2
1922	9,2	16,8	0,174	-8,0	-6,1	-2,1
1923	9,2	17,1	0,190	0,0	1,8	8,9
1924	9,0	17,1	0,217	-2,2	0,0	14,3
1925	9,1	17,5	0,232	1,1	2,3	7,1
1926	9,2	17,7	0,306	1,1	1,1	31,7
1927	9,1	17,4	0,317	-1,1	-1,7	3,8
1928	9,1	17,1	0,317	0,0	-1,7	0,0
1929	9,2	17,1	0,337	1,1	0,0	6,1
1930	9,1	16,7	0,341	-1,1	-2,3	1,2
1931	8,2	15,2	0,325	-9,9	-9,0	-4,5
1932	7,5	13,7	0,298	-8,5	-9,9	-8,4
1933	7,1	13,0	0,286	-5,3	-5,1	-3,9

1934	7,2	13,4	0,275	1,4	3,1	-4,0
1935	7,3	13,7	0,252	1,4	2,2	-8,5
1936	7,4	13,9	0,271	1,4	1,5	7,7
1937	7,7	14,4	0,341	4,1	3,6	25,7
1938	7,7	14,1	0,387	0,0	-2,1	13,6
1939	7,7	13,9	0,414	0,0	-1,4	7,0
1940	8,0	14,0	0,488	3,9	0,7	17,8
1941	8,5	14,7	0,573	6,3	5,0	17,5
1942	8,8	16,3	0,689	3,5	10,9	20,3
1943	9,0	17,3	0,856	2,3	6,1	24,2
1944	9,1	17,6	1,046	1,1	1,7	22,2
1945	9,2	18,0	1,554	1,1	2,3	48,5
1946	9,4	19,5	2,371	2,2	8,3	52,6
1947	10,3	22,3	3,538	9,6	14,4	49,2
1948	11,8	24,1	5,614	14,6	8,1	58,7
1949	12,2	23,8	6,356	3,4	-1,2	13,2
1950	12,5	24,1	6,991	2,5	1,3	10,0
1951	13,8	26,0	8,124	10,4	7,9	16,2
1952	14,2	26,5	9,090	2,9	1,9	11,9
1953	14,0	26,7	8,936	-1,4	0,8	-1,7
1954	14,1	26,9	8,972	0,7	0,7	0,4
1955	14,1	26,8	9,052	0,0	-0,4	0,9
1956	14,3	27,2	9,432	1,4	1,5	4,2
1957	14,8	28,1	9,715	3,5	3,3	3,0
1958	15,2	28,9	11,18	2,7	2,8	15,1
1959	15,3	29,1	11,88	0,7	0,7	6,2
1960	15,5	29,6	12,30	1,3	1,7	3,6
1961	15,7	29,9	12,71	1,3	1,0	3,3
1962	15,9	30,2	13,32	1,3	1,0	4,8
1963	16,1	30,6	13,96	1,3	1,3	4,8
1964	16,4	31,0	14,43	1,9	1,3	3,4
1965	16,8	31,5	14,79	2,4	1,6	2,5
1966	17,5	32,4	15,19	4,2	2,9	2,7
1967	18,1	33,4	15,60	3,4	3,1	2,7
1968	18,8	34,8	16,31	3,9	4,2	4,5
1969	19,7	36,7	17,37	4,8	5,5	6,5
1970	20,3	38,8	18,27	3,0	5,7	5,2
1971	20,9	40,5	19,31	3,0	4,4	5,7
1972	21,9	41,8	20,51	4,8	3,2	6,2
1973	23,6	44,4	22,39	7,8	6,2	9,2
1974	26,2	49,3	25,46	11,0	11,0	13,7
1975	29,0	53,8	28,47	10,7	9,1	11,8
1976	31,1	56,9	31,20	7,2	5,8	9,6
1977	33,6	60,6	34,13	8,0	6,5	9,4

1978	36,6	65,2	37,24	8,9	7,6	9,1
1979	40,0	72,6	41,26	9,3	11,3	10,8
1980	44,0	82,4	46,87	10,0	13,5	13,6
1981	49,5	90,9	53,15	12,5	10,3	13,4
1982	54,9	96,5	59,42	10,9	6,2	11,8
1983	58,1	99,6	65,13	5,8	3,2	9,6
1984	60,6	103,9	69,95	4,3	4,3	7,4
1985	63,0	107,6	74,01	4,0	3,6	5,8
1986	65,6	109,6	76,00	4,1	1,9	2,7
1987	68,5	113,6	78,36	4,4	3,6	3,1
1988	71,2	118,3	80,48	3,9	4,1	2,7
1989	74,8	124,0	83,37	5,1	4,8	3,6
1990	78,4	130,7	86,21	4,8	5,4	3,4
1991	82,8	136,2	88,97	5,6	4,2	3,2
1992	84,0	140,3	91,10	1,4	3,0	2,4
1993	85,6	144,5	93,01	1,9	3,0	2,1
1994	85,7	148,2	94,60	0,1	2,6	1,7
1995	87,6	152,4	96,20	2,2	2,8	1,7
1996	88,9	156,9	98,13	1,5	3,0	2,0
1997	90,4	160,5	99,30	1,7	2,3	1,2
1998	91,3	163,0	100,0	1,0	1,6	0,7
1999	92,9	166,6	100,5	1,8	2,2	0,5
2000	95,4	172,2	102,2	2,7	3,4	1,7
2001	97,8	177,1	103,9	2,5	2,8	1,7
2002	100,0	179,9	105,9	2,2	1,6	1,9
2003	102,8	184,0	108,1	2,8	2,3	2,1
2004	104,7	188,9	110,4	1,8	2,7	2,1
2005	107,0	195,3	112,4	2,2	3,4	1,8
2006	109,1	201,6	114,2	2,0	3,2	1,6
2007	111,5	207,3	115,9	2,2	2,8	1,5
2008	114,1	215,3	119,2	2,3	3,8	2,8
2009	114,4	214,5	119,3	0,3	-0,4	0,1
2010	116,5	218,1	121,1	1,8	1,6	1,5
2011	119,9	224,9	123,7	2,9	3,2	2,1
2012	121,7	229,6	126,1	1,5	2,1	2,0
2013	122,8	233,0	127,2	0,9	1,5	0,9

Sources :
Canada : Statistique Canada
États-Unis : Bureau of Labor Statistics
France : Insee