



N° 12-539-XIF au catalogue

Statistique Canada Lignes directrices concernant la qualité

Quatrième édition – Octobre 2003



Comment obtenir d'autres renseignements

Toute demande de renseignements au sujet du présent produit ou au sujet de statistiques ou de services connexes doit être adressée à : Services aux clients, Direction de la méthodologie, Statistique Canada, Ottawa, Ontario, K1A 0T6 (téléphone : 1 800 263-1136).

Pour obtenir des renseignements sur l'ensemble des données de Statistique Canada qui sont disponibles, veuillez composer l'un des numéros sans frais suivants. Vous pouvez également communiquer avec nous par courriel ou visiter notre site Web.

Service national de renseignements	1 800 263-1136
Service national d'appareils de télécommunications pour les malentendants	1 800 363-7629
Renseignements concernant le Programme des bibliothèques de dépôt	1 800 700-1033
Télécopieur pour le Programme des bibliothèques de dépôt	1 800 889-9734
Renseignements par courriel	infostats@statcan.ca
Site Web	www.statcan.ca

Renseignements sur les commandes et les abonnements

Le produit n° 12-539-XIF au catalogue est gratuit sur Internet. Les utilisateurs sont priés de se rendre à http://www.statcan.ca/cgi-bin/downpub/freepub_f.cgi.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois, et ce, dans la langue officielle de leur choix. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1 800 263-1136.



Statistique Canada
Direction de la méthodologie

Statistique Canada

Lignes directrices concernant la qualité

Quatrième édition – Octobre 2003

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2003

Tous droits réservés. Il est interdit de reproduire ou de transmettre le contenu de la présente publication, sous quelque forme ou par quelque moyen que ce soit, enregistrement sur support magnétique, reproduction électronique, mécanique, photographique, ou autre, ou de l'emmagasiner dans un système de recouvrement, sans l'autorisation écrite préalable des Services de concession des droits de licence, Division du marketing, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Octobre 2003

N° 12-539-XIF au catalogue

Périodicité : irrégulier

ISSN 1708-6264

Ottawa

This publication is available in English upon request (Catalogue no. 12-539-XIE)

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Préface

Statistique Canada accorde beaucoup d'importance à la gestion de la qualité des données afin de s'assurer que ses produits statistiques satisfont aux besoins des utilisateurs. Les nombreuses mesures mises en œuvre pour gérer les divers aspects de la qualité sont décrites dans le cadre de l'assurance de la qualité du Bureau (Statistique Canada, 2002c). Dans ce cadre global, l'élaboration et la mise en œuvre de programmes d'acquisition de données statistiques au moyen d'enquêtes par sondage, de recensements ou via la consultation de dossiers administratifs, jouent un rôle essentiel dans la gestion de la qualité des données d'une agence statistique. L'adoption d'une saine méthodologie d'enquête est à la base de la production de données statistiques à la fois conformes aux besoins d'information des clients et d'une précision mesurable.

La méthodologie d'enquête ne s'appuie pas sur un code global de la profession ou sur un ensemble de normes régissant les règles de l'art qui s'appliquent dans tous les genres de situation. De fait, elle repose sur un ensemble de règles appuyées par la théorie et la pratique et auxquelles les professionnels du domaine ont recours pour effectuer des choix raisonnables dans des contextes particuliers. Dans la mesure du possible, ces choix doivent refléter le juste milieu parmi les exigences relatives à la pertinence, à la précision, au temps, aux coûts et au fardeau de déclaration.

Ce document regroupe une série d'expériences et de conclusions qui, prises individuellement dans des contextes particuliers, sont reconnues à titre de « bonnes pratiques ». Elles ne s'appliqueront pas à tout processus d'acquisition. Il faut donc juger minutieusement de leur pertinence et de leur importance en fonction des exigences et des contraintes propres à chaque programme. L'utilisation de ce document commande donc beaucoup d'attention professionnelle et de jugement.

Bien que les lignes directrices de ce document ne remplacent pas l'expertise et le jugement de l'équipe de conception d'enquête, le souci de la qualité doit néanmoins se refléter dans toutes nos activités. Il incombe aux gestionnaires de programmes statistiques de s'assurer que l'importance qu'accorde Statistique Canada à la qualité se reflète de façon juste dans la conception et la mise en œuvre des méthodes et des procédures statistiques qu'ils gèrent.

La première édition de Lignes directrices concernant la qualité est parue en 1985. C'est en 1987 que la seconde édition a été publiée, apportant ainsi des révisions mineures. La méthodologie d'enquête a beaucoup évolué depuis. À titre d'exemple, l'informatique et les communications électroniques ont été intégrées à la collecte et au traitement des données. Le tout a occasionné d'une part des changements au déroulement des opérations d'enquête et d'autre part l'adoption de nouvelles approches de stockage des données diffusées. Cette évolution dans la méthodologie d'enquête a donné lieu à la troisième édition des lignes directrices en 1998. Une importante mise à jour est à l'origine du présent document qui reflète les progrès de la méthodologie au cours des cinq dernières années.

Nous devons des remerciements aux nombreux experts de Statistique Canada qui ont contribué à la préparation de ces lignes directrices au fil du temps. Les conseils du Comité des méthodes et des normes ainsi que les commentaires du Comité consultatif des méthodes statistiques ont permis d'améliorer ce document.

Ivan P. Fellegi



Statisticien en chef

Table des matières

1.	Introduction : Définition de la qualité.....	6
2.	Étapes de l'enquête	10
2.1	Objectifs, utilisations et utilisateurs.....	11
2.2	Concepts, variables et classifications.....	14
2.3	Couverture et bases de sondage	17
2.4	Échantillonnage.....	22
2.5	Conception du questionnaire.....	27
2.6	Réponse et non-réponse	31
2.7	Opérations de collecte et de saisie des données.....	35
2.8	Vérification	40
2.9	Imputation.....	44
2.10	Estimation	48
2.11	Désaisonnalisation et estimation de la tendance-cycle	56
2.12	Évaluation de la qualité des données	61
2.13	Contrôle de la divulgation.....	65
2.14	Diffusion des données.....	72
2.15	Analyse (et présentation) des données.....	75
2.16	Documentation.....	80
2.17	Utilisation des données administratives.....	84
3.	Contexte de gestion.....	91
3.1	La qualité à Statistique Canada.....	91
3.2	Cadre de gestion.....	91
3.3	Cadre d'assurance de la qualité.....	92
3.4	Conclusion	100
4.	Liste d'acronymes	101
5.	Bibliographie.....	102
6.	Annexes.....	116
6.1	Annexe 1: Politique concernant l'examen et la mise à l'essai des questionnaires	116
6.2	Annexe 2: Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie.....	120

1. Introduction : Définition de la qualité

Ce document réunit des lignes directrices et des listes de contrôles liées à de nombreuses questions dont on doit tenir compte dans la poursuite des objectifs de qualité que sous-tend l'exécution des activités statistiques. Le document s'attarde principalement à la façon d'assurer la qualité grâce à la conception ou à la restructuration efficace et adéquate d'un projet ou d'un programme statistique, des débuts jusqu'à l'évaluation, la diffusion et la documentation des données. Ces lignes directrices sont fondées sur les connaissances et l'expérience collective d'un grand nombre d'employés de Statistique Canada. On espère que les Lignes directrices concernant la qualité seront utiles au personnel chargé de la planification et de la conception des enquêtes et d'autres projets statistiques, ainsi qu'à ceux qui évaluent et analysent les résultats de ces projets.

Depuis la publication de la première édition des Lignes directrices concernant la qualité, en 1985, on a amplement discuté, au sein des organismes statistiques nationaux et internationaux, de la question de la qualité. Cette discussion se poursuit encore aujourd'hui. Bien qu'il n'existe aucune définition normalisée de la qualité en matière de statistiques officielles, on convient généralement, dans ces organismes, que la qualité englobe une notion élargie d'« adaptation des données à leur utilisation ». Cette notion comprend non seulement les concepts de qualité statistique de variance et de biais, mais également d'autres caractéristiques, comme la pertinence et l'actualité, qui déterminent la façon d'utiliser efficacement l'information statistique.

Cette définition générale de la qualité côtoie des points de vue semblables suggérés par le mouvement de *gestion de la qualité totale (GQT)*. Il s'agit, en partie, d'atteindre et de maintenir un niveau de qualité ou d'adaptation acceptable pour les défenseurs de la GQT, soit de connaître et de comprendre les besoins des clients, de faire participer les employés au processus décisionnel rattaché à la concrétisation de ces besoins et de chercher constamment à améliorer les méthodes et les processus. L'attention accordée à ces trois principes contribuera à rehausser la qualité, et cela est aussi vrai pour un organisme statistique que pour tout autre organisme. Les Lignes directrices concernant la qualité reflètent ces trois principes, ainsi que les efforts que déploie Statistique Canada depuis longtemps pour recueillir et diffuser de l'information statistique fiable et objective qui répond aux besoins essentiels et anticipe ces derniers.

Éléments de qualité

Statistique Canada définit la qualité ou l'« adaptation des données à leur utilisation » de l'information statistique en fonction de six dimensions ou éléments constitutifs, soit la pertinence, l'exactitude, l'actualité, l'accessibilité, l'intelligibilité et la cohérence (Statistique Canada, 2002c).

La *pertinence* de l'information statistique reflète la mesure dans laquelle cette information répond aux besoins réels des clients. C'est en examinant cet élément qu'on détermine si l'information disponible permet de mieux comprendre les enjeux qui sont importants pour les utilisateurs. L'évaluation de la pertinence est subjective et est

fonction des différents besoins des utilisateurs. Le défi du Bureau consiste à pondérer et à équilibrer les besoins conflictuels des utilisateurs actuels et potentiels afin d'élaborer un programme qui répond aux besoins les plus importants dans un contexte de ressources limitées.

L'*exactitude* de l'information statistique est la mesure dans laquelle l'information décrit correctement le phénomène qu'elle devait évaluer. Généralement, elle est caractérisée par l'erreur dans les estimations statistiques et est décomposée en composantes de biais (erreur systématique) et de variance (erreur aléatoire). L'exactitude peut également être décrite en fonction des sources d'erreur majeures qui peuvent mener à l'inexactitude (p. ex., couverture, échantillonnage, non-réponse, réponse).

L'*actualité* de l'information statistique renvoie à l'intervalle entre le point de référence (ou la fin de la période de référence) auquel se rapporte l'information et la date à laquelle l'information est diffusée. Habituellement, l'actualité se trouve en relation d'équilibre avec l'exactitude. L'actualité de l'information influera sur sa pertinence.

L'*accessibilité* de l'information statistique renvoie à la facilité avec laquelle on peut obtenir l'information auprès du Bureau. Cela comprend la facilité avec laquelle on peut certifier l'existence de l'information, ainsi que l'à-propos de la forme ou du médium par le biais duquel on peut accéder à l'information. En outre, le coût de l'information peut représenter un facteur d'accessibilité pour certains utilisateurs.

L'*intelligibilité* de l'information statistique reflète la disponibilité de l'information et des métadonnées supplémentaires nécessaires à l'interprétation et à l'utilisation appropriées des renseignements. Normalement, cette information comprend les variables, les classifications et les concepts sous-jacents utilisés, la méthode de collecte des données et le traitement, ainsi que les indications ou les mesures de l'exactitude de l'information statistique.

La *cohérence* de l'information statistique reflète la mesure dans laquelle on peut réussir à regrouper cette information avec d'autres renseignements statistiques dans un cadre analytique général et au fil du temps. L'utilisation de classifications, des populations cibles et de concepts normalisés favorise la cohérence, tout comme l'utilisation d'une méthode commune d'une enquête à l'autre. La cohérence ne sous-tend pas nécessairement l'uniformisation numérique complète.

Ces dimensions de qualité se chevauchent et se recoupent. Aucun modèle général ne les réunit pour optimiser ou pour prescrire un niveau de qualité donné. Pour atteindre un niveau de qualité acceptable, il faut considérer, gérer et équilibrer ces éléments de qualité au fil du temps en accordant une grande attention aux objectifs des programmes, aux coûts, au fardeau des répondants et à d'autres facteurs qui peuvent influencer sur la qualité de l'information ou sur les attentes des utilisateurs. Cet équilibre représente un des aspects essentiels de la conception des enquêtes du Bureau.

Enquête

Dans le présent document, le terme *enquête* est un générique qui s'applique à toute activité visant à recueillir ou à acquérir des données statistiques. Cela comprend les éléments suivants :

- Un *recensement*, par lequel on tente de recueillir des données sur tous les membres de la population.
- Une *enquête par sondage*, dans laquelle on recueille des données sur un échantillon (habituellement aléatoire) des membres de la population.
- La collecte des données provenant des *dossiers administratifs*, où les données sont tirées des documents initialement conservés à des fins non statistiques.
- Une *activité statistique dérivée*, dans laquelle on estime et on modèle des données qui peuvent être tirées des sources de données statistiques existantes.

Les lignes directrices concernent principalement les recensements et les enquêtes par sondage. Bien qu'un grand nombre de lignes directrices s'appliqueront également au traitement des dossiers administratifs, on a ajouté une section additionnelle (2.17) sur ce sujet afin de mettre en évidence les éléments propres à cette activité. Il est manifeste que la qualité des activités statistiques dérivées est grandement déterminée par la qualité des parties des composantes. Par conséquent, le présent document ne s'attarde pas directement à ces activités.

Conception

Dans le présent document, le terme *conception* s'applique à la délimitation de tous les aspects d'une enquête, de la détermination d'un besoin de données jusqu'à la production des résultats finals (le fichier de microdonnées, les séries statistiques et l'analyse).

Le corps du présent document (section 2) traite principalement des questions de qualité liées à la conception des enquêtes. Toutefois, il importe de se rappeler que le contexte dans lequel on prépare une enquête impose des contraintes à la conception de cette enquête. Chaque nouvelle enquête, tout en visant à répondre à certains besoins immédiats en matière d'information, ajoute également de l'information à une base de données statistiques qu'on peut utiliser à des fins qui sont beaucoup plus nombreuses que celles établies au moment de la conception de l'enquête. Il est donc important de s'assurer que le résultat de chaque enquête peut, dans la mesure du possible, être intégré à des données sur des sujets connexes tirées d'autres enquêtes et utilisé en conjonction avec ces données. Cela sous-tend la nécessité d'examiner et de respecter les normes statistiques sur le contenu et les domaines qui ont été mises en place pour offrir une cohérence et une harmonie des données dans le système statistique national. Parmi ces normes, on trouve des cadres statistiques (comme le Système de comptabilité nationale), des systèmes de classification statistique (comme ceux rattachés à l'industrie ou à la géographie), ainsi que d'autres concepts et définitions qui précisent les variables statistiques à mesurer.

L'utilité des nouvelles données statistiques est accrue jusqu'au point où on peut les utiliser en conjonction avec des données existantes.

Le processus de conception est également mis en place dans un contexte organisationnel donné. Les présentes lignes directrices s'appliquent à une agence statistique centralisée au sein de laquelle la conception d'une enquête est normalement confiée à une équipe de projet multidisciplinaire. Les principaux membres de l'équipe de projet sont le gestionnaire de projet et un groupe de spécialistes. Parmi les spécialistes, on trouve généralement un spécialiste du domaine, un méthodologiste, un spécialiste en informatique et un spécialiste des opérations. Parfois, un membre de l'équipe assumera plus d'un rôle et parfois, on devra ajouter plusieurs autres responsabilités dans l'équipe. Par exemple, des spécialistes peuvent être affectés aux systèmes géographiques, aux communications publiques et à la diffusion.

La section 3 décrit le contexte de gestion dans lequel s'appliquent les Lignes directrices concernant la qualité. Fondée sur le cadre d'assurance de la qualité, cette description regroupe les politiques, les processus de gestion, les mécanismes de consultation et les procédures techniques qui ont une influence sur la gestion de la qualité à Statistique Canada. Alors que la section 2 s'attarde principalement à la conduite des activités statistiques individuelles, la section 3 offre une perspective plus générale sur l'assurance de la qualité au sein du Bureau.

Références

Brackstone, G. (1999). La gestion de la qualité des données dans un bureau de statistique. **Techniques d'enquête**, 25, 157-171.

Statistique Canada (2002c). Le cadre d'assurance de la qualité de Statistique Canada - 2002. Publication n° 12-586-XIF au catalogue, Statistique Canada.

2. Étapes de l'enquête

Cette section est divisée en sous-sections qui correspondent aux activités principales d'une enquête type. Les sous-sections présentent toutes la même structure : elles décrivent la *portée et l'objet*, les *principes* et les *lignes directrices* liées à chaque activité. La première sous-section traite de l'étape à laquelle on détermine les objectifs, les utilisations et les utilisateurs (2.1). Les sous-sections qui suivent décrivent les autres étapes de l'enquête dans l'ordre chronologique où elles se présentent habituellement. Cependant, il existe des interrelations importantes entre certaines étapes, comme par exemple entre la conception du questionnaire (2.5) et les opérations de collecte et de saisie des données (2.7). Pour cette raison, certaines sous-sections renvoient à d'autres sous-sections. En outre, les étapes d'enquête dont il est question dans ce document ne se déroulent pas nécessairement selon une séquence stricte. Certaines activités peuvent être menées simultanément, par exemple, la constitution de la base de sondage (2.3), les plans d'échantillonnage (2.4) et la conception du questionnaire (2.5). D'autres étapes, telles que l'évaluation de la qualité des données (2.12) et la documentation (2.16) sont rattachées à la majorité des autres activités et ne constituent pas des activités distinctes en soi. Enfin, on aborde séparément l'utilisation des données administratives (2.17) afin d'examiner les questions qui sont propres à ce type de sources de données.

Portée et objet

Sous la rubrique *Portée et objet*, on donne une description de l'activité et on indique son impact potentiel sur la qualité. Essentiellement, on établit une définition et un contexte.

Principes

Les *principes* sont les politiques, les approches et les orientations générales qui régissent la conception d'une activité donnée; on accorde une priorité aux principes qui sont rattachés à la qualité.

Lignes directrices

Les *lignes directrices* sont des règles de pratique reconnues qui ont été élaborées lors de la conception et de la mise en œuvre des enquêtes statistiques. Toutes ces lignes directrices ne s'appliquent pas à chaque enquête, mais offrent des listes de contrôle pour faciliter la conception de l'enquête. Il faut toutefois faire preuve de jugement afin d'évaluer les considérations suggérées par ces lignes directrices.

D'un autre côté, Statistique Canada applique des politiques qui ont une incidence sur de nombreux aspects des activités statistiques du Bureau et qui peuvent comporter des exigences liées à la mise en œuvre d'activités particulières. Ces politiques sont décrites dans le Manuel des politiques de Statistique Canada. Lorsqu'une politique a une incidence sur un sujet particulier traité dans les lignes directrices, on indique l'existence et la pertinence de cette politique.

2.1 Objectifs, utilisations et utilisateurs

Portée et objet

En général, lorsqu'une nouvelle activité statistique ou la refonte d'une activité existante a été approuvée, on a démontré que l'information répond à un besoin et que le projet ou l'activité proposé peut être réalisé. De plus, le processus de planification a permis de formuler des objectifs généraux, de cibler les utilisateurs et de cerner les grandes questions qui doivent être analysées dans le cadre de l'activité. Afin de transformer cet exercice de planification initial en véritable mécanisme de collecte, on peut maintenant indiquer les objectifs et les utilisations avec plus de précision pour s'assurer que la nouvelle activité ou l'activité remaniée répondra aux besoins spécifiques des utilisateurs.

Les *objectifs* sont les besoins pour lesquels on requiert l'information, établis dans le contexte du programme, du problème de recherche ou des hypothèses qui donnent naissance au besoin en information. Les *utilisations* limitent et indiquent plus précisément les besoins en information, par exemple, en décrivant les décisions qui peuvent être prises à la lumière de l'information recueillie et la façon dont cette information appuiera ces décisions. Pour les enquêtes périodiques, d'autres utilisations peuvent évoluer au fil du temps. Les *utilisateurs* sont les organisations, les organismes, les groupes ou les personnes qui devraient utiliser l'information. Le consensus sur les utilisations et les objectifs précis facilite la prise de décisions rationnelles liées à la conception de l'enquête.

Principes

La précision des objectifs et des utilisations permet l'élaboration d'un plan détaillé de la nouvelle activité, en consultation avec les utilisateurs de l'information et les participants au projet. Les utilisateurs peuvent aider à préparer la description des buts de l'activité. Les participants au projet peuvent déterminer les problèmes conceptuels, méthodologiques et opérationnels qu'ils doivent résoudre et peuvent proposer un calendrier raisonnable.

Il est important de bien comprendre les objectifs et de les énoncer concrètement en indiquant les hypothèses à mettre à l'épreuve et les exigences de données spécifiques. Cet énoncé doit également inclure la qualité prévue, les contraintes budgétaires et les dates de livraison prévues.

Un énoncé des objectifs permettra aux utilisateurs subséquents qui poursuivent des objectifs différents, d'évaluer la mesure dans laquelle un produit tiré d'une activité statistique peut répondre à leurs propres besoins. En outre, cet énoncé est un excellent moyen de communiquer, aux utilisateurs (potentiels), les attentes qu'ils peuvent avoir à l'égard des produits d'une activité statistique et le degré de prudence qu'ils devront accorder à leur utilisation des données lorsque celle-ci dépasse les limites que l'activité prévoyait.

Lignes directrices

- Orienter l'analyse des besoins des utilisateurs sur la recherche des solutions les plus rentables à court et à long terme. Avant de commencer la conception d'une nouvelle activité statistique (ou la restructuration d'une activité existante), analyser les statistiques accessibles dans le secteur sur le plan de la provenance, de la fréquence, de la qualité, de l'actualité, etc. Examiner la relation d'équilibre entre l'à-propos des statistiques accessibles pouvant répondre aux besoins des clients et le coût et le temps requis pour entreprendre une nouvelle activité de production de statistiques qui n'existent pas encore.
- Établir les objectifs de l'enquête conjointement avec les utilisateurs et les intervenants importants. Établir et entretenir des relations avec les utilisateurs de l'information dans les secteurs public et privé et avec le public général afin de rehausser la pertinence de l'information produite et d'améliorer la commercialisation des produits et des services. Parmi les utilisateurs importants, mentionnons les représentants des marchés potentiels, les décideurs et les agents qui ont besoin de l'information pour une utilisation autorisée par la loi. Avant les conceptions et les restructurations d'envergure, mener systématiquement une consultation élargie et axée sur les utilisateurs afin de déterminer les options de contenu et de susciter un appui public à l'égard du programme lorsque ce dernier atteint l'étape de collecte des données. Étant donné que même les statistiques pertinentes et exactes ne sont pas utilisées si elles ne sont pas considérées comme fiables, il faut adopter une approche très ouverte en élaborant ou en révisant les programmes afin de susciter la confiance dans le processus et les produits d'une activité statistique.
- En déterminant la mesure dans laquelle une enquête répondra aux besoins des utilisateurs, il faut trouver un équilibre raisonnable entre ces besoins et les questions liées au budget, au fardeau de réponse et à la protection des renseignements personnels. Bien que le pouvoir discrétionnaire du Bureau soit minime lorsqu'il y a une prescription d'une loi, il est souhaitable, dans les autres cas, de formuler des approches méthodologiques de rechange, des moyens et des modes de collecte des données, des fréquences, un degré de détail géographique, etc. en vue de trouver une solution optimale.
- Passer en revue les activités statistiques en cours à intervalles réguliers. Pour que le Bureau puisse suivre le rythme des nouvelles exigences des utilisateurs qu'il sert ou des nouveaux utilisateurs, les programmes statistiques doivent évoluer, s'adapter et innover. Le but de l'activité ou l'énoncé de ses objectifs doit être réexaminé périodiquement, ce qui permet d'améliorer la pertinence du produit statistique par rapport aux besoins des utilisateurs, au fur et à mesure qu'ils évoluent ou changent. Parfois, il peut être souhaitable de restructurer les enquêtes existantes pour conserver la fiabilité des principales séries statistiques, surtout si les sources d'information ont changé ou si on a restructuré ou repensé la façon de les diffuser.

- Lorsqu'il existe des objectifs de qualité des données explicites, il faut les intégrer à l'énoncé des objectifs de l'enquête en indiquant les aspects mesurables de la qualité. On peut établir des objectifs pour mesurer, par exemple, l'erreur d'échantillonnage, les taux de couverture, les taux de réponse et l'actualité. En ce qui concerne les données administratives et les activités statistiques dérivées, la qualité des données sera directement liée à la qualité des sources de données d'entrée.
- En déterminant les objectifs et les utilisations d'une activité statistique précise, il faut également tenir compte des objectifs et des utilisations des activités statistiques subséquentes ou des cadres statistiques (p. ex., le Système de comptabilité nationale).

Références

- Blanc, M., Radermacher, W. et Körner, T. (2001). Quality and users. Session 15.1 of **the International Conference on Quality in Official Statistics**, Stockholm, Sweden. Voir aussi <http://method/English/research/Proceed/Q-2001/15.1.pdf> (Site intranet de SC).
- Brackstone, G.J. (1993). Data relevance: keeping pace with user needs. **Journal of Official Statistics**, 9, 49-56.

2.2 Concepts, variables et classifications

Portée et objet

Les *concepts* sont les sujets d'enquête et d'analyse qui intéressent les utilisateurs. Ils renvoient à des caractéristiques ou à des attributs généraux d'une unité statistique ou d'un groupe d'unités statistiques semblables. Les *variables* sont les indicateurs qu'on utilise pour évaluer les concepts. Il importe que les définitions des concepts et des variables précises requises pour l'évaluation des concepts soient sans équivoque et clairement précisées pour les besoins de l'analyse en prévision de laquelle on recueille les données. Comme toutes les données statistiques catégoriques doivent être classées en vue de l'analyse, les critères de *classification* choisis pour regrouper les données de façon systématique doivent être pertinents pour les besoins de l'analyse.

Principes

Afin de tirer des conclusions à partir d'un ensemble de données, il est extrêmement important que les utilisateurs aient une connaissance des concepts qui sous-tendent les données, c.-à-d. ce que les données devraient mesurer. Bien que l'utilisation des définitions harmonisées de concepts, de variables et de classifications aide les utilisateurs à comparer et à intégrer les données, il peut être nécessaire de modifier ces définitions dans le but de respecter les utilisations prévues des données. Les définitions des concepts, des variables et des classifications doivent être soigneusement documentées. Par ailleurs, il faut noter toute différence par rapport aux normes ou aux définitions utilisées pour produire les données connexes. Cette documentation est particulièrement importante pour les utilisateurs qui désirent se servir des données pour un usage autre que celui prévu.

Parfois, il y a plus d'une façon d'évaluer un concept. Les variables et les classifications utilisées pour évaluer un concept devront également tenir compte de facteurs comme la facilité d'obtenir l'information nécessaire, le fardeau du répondant imposé, la méthode de collecte, le contexte dans lequel la ou les questions doivent être posées, le traitement des données (et particulièrement les techniques de révision, d'imputation et de pondération), la présence de l'information dans les dossiers administratifs et les coûts associés à la collecte et au traitement. Par conséquent, l'approche d'évaluation adoptée peut, avec plus ou moins de succès, permettre l'interprétation souhaitée du concept. Une variable choisie à un certain moment peut devenir désuète plus tard si de nouveaux facteurs entrent en ligne de compte. Il faut alors les modifier ou les changer.

Lignes directrices

- Préciser clairement les concepts et les variables et indiquer leur utilisation prévue. Lorsque cela est possible, utiliser les définitions normalisées de concepts, variables, classifications, unités statistiques et populations établies en vertu de la Politique concernant les normes de Statistique Canada (Statistique Canada, 1998b). En choisissant les règles d'affectation des noms, il faut tenir compte de la

ressemblance et de la dissemblance des normes existantes et de l'usage. Les titres des normes existantes ne doivent servir que pour ce qui est défini dans les normes.

- Pour optimiser la souplesse de l'utilisation, coder les microdonnées et conserver les fichiers au plus faible niveau de classification appropriée possible. Il peut être nécessaire de procéder à une agrégation à un niveau plus élevé pour répondre aux besoins particuliers de l'analyse ou pour respecter les contraintes de confidentialité ou de fiabilité des données. Lorsque cela est possible, utiliser une stratégie de regroupement commune pour effectuer ces agrégations et définir celles-ci par rapport aux classes ou aux agrégations de niveau supérieur de la norme. Documenter les différences entre la norme et les niveaux de classifications/agrégations adoptés et utilisés. Appliquer des classifications qui reflètent les niveaux les plus détaillés et les niveaux regroupés. Préciser aux utilisateurs comment ces classifications s'intègrent aux classifications de niveau supérieur (c.-à-d. moins détaillé).
- Utiliser des définitions normalisées pour permettre la comparaison des données recueillies à partir de différentes sources et pour intégrer les données d'une source à l'autre (Statistique Canada, 1998b). Statistique Canada a établi des classifications normalisées des industries, des produits, des professions et de la géographie (Statistique Canada, 2001b, 2001c, 2002b, 2003c) ainsi que d'un grand nombre d'autres variables utilisées dans les statistiques sociales et économiques (Statistique Canada, 2000c).
- Outre les classifications normalisées de Statistique Canada, il existe des classifications internationales normalisées produites par le Bureau de statistique des Nations Unies, par le Bureau international du Travail, par Eurostat et par d'autres organismes régionaux et internationaux. La Division des normes a produit des concordances officielles avec un certain nombre de classifications internationales normalisées. Lorsque qu'il est nécessaire de transmettre des données à des organismes internationaux, il faut utiliser les concordances officielles qui existent.
- Utiliser des unités d'observation normalisées pour faciliter la comparaison des données. Les classifications sont généralement conçues à l'aide d'unités d'observation particulières. Par exemple, la classification des industries de l'Amérique du Nord sert principalement à classer des établissements.
- Il faut connaître les activités statistiques dérivées ou les cadres statistiques (p. ex., le Système de comptabilité nationale) dont les définitions des concepts et des variables peuvent avoir une incidence considérable sur des activités de collecte de données précises (Statistique Canada, 1989).
- En l'absence d'une norme officielle, examiner les concepts, les variables et les classifications utilisés dans le cadre des programmes statistiques connexes et consulter la Division des normes s'il y a lieu.

Références

Statistique Canada (1989). **Guide de l'utilisateur pour le Système de comptabilité nationale**. Publication n° 13-589 au catalogue, Statistique Canada.

Statistique Canada (1998b). Politique concernant les normes. **Manuel de politiques**, 2.10.

Statistique Canada (2000c). Base de métadonnées intégrée: http://stdsweb/standards/imdb/imdb-menu_f.htm (Site intranet de SC).

Statistique Canada (2001b). **Classification nationale des professions pour statistiques (CNP-S) 2001**. Publication n° 12-583-XPF au catalogue, Statistique Canada.

Statistique Canada (2001c). **Classification type des biens 2001** (basée sur le Système harmonisé de désignation et de codification des marchandises). Publication n° 12-580 au catalogue, Statistique Canada.

Statistique Canada (2002b). **Classification géographique type (CGT) 2001**. Publication n° 12-571-XPB au catalogue, Statistique Canada

Statistique Canada (2003c). **Système de classification des industries de l'Amérique du Nord, Canada 2002**. Publication n° 12-501-XPF au catalogue, Statistique Canada.

2.3 Couverture et bases de sondage

Portée et objet

La *population cible* est l'ensemble des éléments sur lesquels on désire obtenir des renseignements et sur lesquels on doit produire des estimations. Des aspects pratiques peuvent nous dicter d'exclure délibérément certaines unités (p. ex., les personnes institutionnalisées, les sans-abri et les personnes qu'il ne serait pas possible de joindre sans coût excessif). Cette situation donne naissance au concept de *population observée*, soit l'ensemble des unités que les contraintes pratiques nous obligent à restreindre en excluant certaines unités difficiles d'accès ou dont l'accès est onéreux. Les différences entre la population cible et la population observée résultent de restrictions de couverture délibérées. Si les deux éléments diffèrent, on peut effectuer une inférence statistique valide par échantillonnage probabiliste à l'égard de la population observée, mais pas à l'égard de la population cible en soi.

Une *base de sondage* est une liste, un document ou un dispositif qui délimite et identifie les éléments de la population observée et qui en permet l'accès. De façon générale, il existe deux types de bases de sondage : les bases aréolaires et les bases de liste. Habituellement, les bases aréolaires font partie d'une hiérarchie d'unités géographiques, c.-à-d. que les sous-sections de la base de sondage d'un niveau peuvent être subdivisées pour former des unités du niveau suivant. Tous les éléments compris dans la base de sondage constituent la *population de la base de sondage*. Les divergences entre la population observée et la population de la base de sondage sont appelées les *erreurs de couverture*.

Principes

La base de sondage doit être conforme à la population observée et contenir un sous-dénombrement et un surdénombrement minimaux. L'information de la base de sondage doit être mise à jour. Les erreurs de couverture sont attribuables à des omissions, des inclusions erronées, des doublons ou des erreurs de classification des unités dans la base de sondage.

Les caractéristiques des unités de la base de sondage (p. ex., identification, personne-ressource, classification, adresse, taille, cartes pour les unités géographiques) doivent être de grande qualité, car elles servent à la stratification, à la sélection des échantillons, à la collecte, au suivi, au traitement des données, à l'imputation, à l'estimation, au couplage des enregistrements, à l'évaluation de la qualité et à l'analyse. Les imperfections de la base de sondage, telles que les erreurs de couverture et les caractéristiques désuètes, sont susceptibles de biaiser ou de diminuer la fiabilité des estimations de l'enquête et de gonfler les coûts de collecte des données.

Lignes directrices

- Mettre à l'essai des bases de sondage possibles à l'étape de la planification d'une enquête afin d'en évaluer l'à-propos et la qualité.
- En présence de plusieurs bases de sondage, dont certaines sont incomplètes mais moins coûteuses à utiliser et d'autres sont plus complètes mais très onéreuses, il faut envisager l'utilisation de bases de sondage multiples (Hartley, 1962; Bankier, 1986; Sirken et Casady, 1988; Kott et Vogel, 1995). On peut également utiliser le système d'appel aléatoire (SAA) pour certaines enquêtes téléphoniques, soit seul soit en combinaison avec d'autres bases aréolaires ou bases de liste.
- À Statistique Canada, pour les enquêtes-entreprises et les enquêtes institutionnelles, le Registre des entreprises est la base de sondage habituelle. Pour les enquêtes agricoles, le registre des fermes est la base de sondage habituelle. Pour les enquêtes-ménages, le Registre des adresses, la base de sondage de l'Enquête sur la population active (qui est une base aréolaire), les fichiers téléphoniques et les unités géographiques du recensement de la population sont des options à envisager.
- Lorsque cela est possible, utiliser une même base de sondage pour les enquêtes ayant la même population cible afin d'éviter les incohérences, de faciliter la combinaison des estimations tirées des enquêtes et de réduire les coûts de la mise à jour et de l'évaluation de la base de sondage.
- Conserver et stocker l'information sur l'échantillonnage, la rotation et la collecte des données afin de procéder à une coordination des enquêtes et de mieux gérer les relations avec les répondants et le fardeau de réponse. Par exemple, il faut indiquer à quelle fréquence chaque unité est sélectionnée par chacune des enquêtes utilisant la même base de sondage.
- Pour rehausser ou maintenir le niveau de qualité de la base de sondage, incorporer les procédures d'élimination des doublons et d'actualisation des naissances, des décès, des unités hors du champ d'observation et des changements de caractéristique.
- Surveiller la qualité de la base de sondage en évaluant périodiquement sa couverture et la qualité de l'information rattachée aux caractéristiques des unités. Il existe de nombreuses techniques pour y arriver :
 - Appairer la base de sondage ou un de ses échantillons avec d'autres sources comparables, souvent fournies par des dossiers administratifs, pour établir la population observée ou des sous-ensembles de cette dernière.
 - Analyser les résultats d'enquête pour mettre en évidence les doublons, les décès, les unités hors du champ d'observation et les changements de caractéristique.

- Utiliser des questions précises du questionnaire pour faciliter le suivi de l'information sur la couverture et la classification. Vérifier auprès des autorités locales (p. ex., les bureaux régionaux, le personnel des enquêtes sur le terrain, les unités d'enquête en soi).
 - Vérifier la base de sondage ou des sous-ensembles de cette dernière sur le terrain (ce qui pourrait sous-tendre la vérification des unités hors du champ d'observation).
 - Comparer la base de sondage avec un échantillon des unités d'une base aréolaire correspondante.
 - Actualiser la base de sondage afin d'en déterminer les changements.
 - Vérifier l'uniformité des chiffres dans d'autres sources ou avec des données tirées de répliques spécialement conçues.
 - Utiliser l'information évaluative tirée des autres enquêtes ayant la même base de sondage (Lessler et Kalsbeek, 1992).
- Effectuer un suivi de la base de sondage entre le moment de la sélection de l'échantillon et la période de référence de l'enquête.
 - Intégrer les mises à jour de la base de sondage de la façon la plus opportune qui soit.
 - Réduire au minimum les erreurs de base de sondage en offrant une formation efficace au personnel, en insistant sur l'importance de la couverture et en mettant en œuvre les procédures d'assurance de la qualité des activités liées à la base de sondage.
 - En ce qui concerne les bases aréolaires, procéder à des vérifications de carte pour assurer la délimitation précise et sans chevauchement des aires géographiques utilisées dans la conception de l'échantillonnage (p. ex., par des vérifications sur le terrain ou l'utilisation d'autres sources de carte). S'il y a lieu, utiliser le Système automatisé de regroupement des territoires (SARTE) pour créer des unités de base de sondage géographiques.
 - Pour les activités statistiques rattachées à des sources administratives ou pour les activités statistiques dérivées, où les changements de couvertures peuvent échapper au contrôle du gestionnaire immédiat, déterminer et surveiller la couverture, et négocier les changements requis avec le gestionnaire de la source.
 - Apporter des rajustements aux données ou utiliser des données supplémentaires tirées d'autres sources pour compenser l'erreur de couverture de la base de sondage.
 - Intégrer, dans la documentation sur l'enquête, la description des populations cible et observée, les différences entre la population cible et la population observée, ainsi que la description de la base de sondage et ses erreurs de couverture.

Références

- Archer, D. (1995). Maintenance of business registers. Dans **Business Survey Methods**, B.G. Cox et coll. (éd.), Wiley-Interscience, New York, 85-100.
- Bankier, M.D. (1986). Estimators based on several stratified samples with applications to multiple frame surveys. **Journal of the American Statistical Association**, 81, 1074-1079.
- Burgess, R.D. (1988). Évaluation des estimations du sous-dénombrement obtenues par la contre-vérification des dossiers du recensement du Canada. **Techniques d'enquête**, 14, 147-167.
- Colledge, M.J. (1995). Frames and business registers: an overview. Dans **Business Survey Methods**, B.G. Cox et coll. (éd.), Wiley-Interscience, New York, 21-47.
- Gambino, J.G., Singh, M.P., Dufour, J., Kennedy, B. et Lindeyer, J. (1998). **Méthodologie de l'Enquête sur la population active du Canada**. Publication n° 71-526-XPB au catalogue, Statistique Canada.
- Gosselin, J.-F., Chinnappa, B.N., Ghangurde, P.D. et Tourigny, J. (1978). Couverture. Chapitre 2 dans **Répertoire de méthodes d'évaluation des erreurs dans les recensements et les enquêtes**, Publication n° 13-564F au catalogue, 7-10. Statistique Canada.
- Hartley, H.O. (1962). Multiple frame surveys. **Proceedings of the Social Statistics Section**, American Statistical Association, 203-206.
- Kott, P.S. et Vogel, F.A. (1995). Multiple-frame business surveys. Dans **Business Survey Methods**, B.G. Cox et coll. (éd.), Wiley-Interscience, New York, 185-203.
- Laniel, N. et Finlay, H. (1991). Data quality concerns with sub-annual business survey frames. **Proceedings of the Section on Survey Research Methods**, American Statistical Association, 202-207.
- Lessler, J.T. et Kalsbeek, W.D. (1992). **Nonsampling Errors in Surveys**. Wiley, New York.
- Massey, J.T. (1988). An overview of telephone coverage. Dans **Telephone Survey Methodology**, R.M. Groves et coll. (éd.), Wiley, New York, 3-8.
- Pollock, K.H., Turner, S.C. et Brown, C.A. (1994). Techniques de saisie-ressaisie pour l'estimation de la taille de la population et de totaux de population lorsqu'on ne dispose pas d'une base de sondage complète. **Techniques d'enquête**, 20, 121-128

- Satin, A. et Shastry, W. (1993). Éléments d'un plan de sondage. Chapitre 2 dans **L'échantillonnage : un guide non mathématique - deuxième édition**, 7-10, Publication n° 12-602-XPF au catalogue, Statistique Canada.
- Sirken, M.G. et Casady, R.J. (1988). Sampling variance and nonresponse rates in dual frame, mixed mode surveys. Chapitre 11 dans **Telephone Survey Methodology**, P. Biemer, R.M. Groves, L. Lyberg, J. Massey, W. Nicholls et J. Waksberg (éd.), Wiley, New York, 175-188.
- Swain, L., Drew, J.D., Lafrance, B et Lance, K. (1992). La création d'un registre des adresses résidentielles pour améliorer la couverture du recensement du Canada de 1991. **Techniques d'enquête**, 18, 139-155.

2.4 Échantillonnage

Portée et objet

L'*échantillonnage* est la sélection d'un ensemble d'unités tirées d'une population observée. Cet ensemble d'unités est appelé l'*échantillon*. Le choix de la méthode d'échantillonnage influe directement sur la qualité des données. La qualité est déterminée par de nombreux facteurs, dont le degré souhaité de précision et de détail de l'information à produire, la disponibilité des bases de sondage appropriées, la disponibilité de variables auxiliaires convenables pour la stratification et la sélection de l'échantillon, les méthodes d'estimation qui seront appliquées et les budgets prévus.

Principes

On utilise l'*échantillonnage probabiliste* pour choisir un échantillon de la population observée. L'objectif consiste à recueillir de l'information utile à partir des unités échantillonnées afin de permettre les inférences liées à la population observée. L'échantillonnage probabiliste sous-tend une sélection probabiliste d'unités tirées de la base de façon que toutes les unités de la population observées correspondent à des *probabilités d'inclusion* connues et positives. La taille de l'échantillon est fonction de la précision requise et du budget affecté à l'observation des unités sélectionnées. La distribution de la probabilité qui régit la sélection de l'échantillon, ainsi que les étapes et les unités de l'échantillonnage, la stratification, etc. sont collectivement appelées le *plan d'échantillonnage*. On choisit une combinaison de plan d'échantillonnage et de méthode d'estimation (voir la section 2.10) afin que les estimations obtenues offrent la plus grande précision possible dans les limites du budget ou entraînent le coût le plus faible qui soit pour une précision fixe. Afin de rehausser la précision des estimations, on peut remplacer, à l'étape de l'estimation et s'il y a lieu, les renseignements recueillis pour les unités échantillonnées par de l'*information auxiliaire* provenant d'autres sources que l'enquête en soi (p. ex., dossiers administratifs et prévisions du recensement). Le choix du plan d'échantillonnage tiendra compte de l'accessibilité à cette information auxiliaire. Ces concepts sont traités par Särndal, Swensson et Wretman (1992), et par Tillé (2001).

Lignes directrices

- La *stratification* consiste à diviser la population en sous-ensembles (appelés des *strates*) dans chacun desquels on choisit un échantillon indépendant. Le choix des strates est déterminé par l'objectif de l'enquête, les caractéristiques de distribution de la variable intéressante et la précision des estimations souhaitée. On mène la majorité des enquêtes pour produire des estimations sur divers *domaines* d'intérêt (p. ex., provinces). Lorsque cela est possible, il faut tenir compte de cette réalité dans la conception en stratifiant de façon convenable (p. ex., par province). Autrement, il faudra envisager des méthodes spéciales à l'étape de l'estimation pour produire des estimations sur ces domaines (voir la section 2.10). Afin d'atteindre une efficacité statistique, créer des strates de façon que chacune contienne des unités le plus homogène possible à l'égard de l'information

demandée dans l'enquête. En ce qui concerne les enquêtes longitudinales, choisir des variables de stratification qui correspondent à des caractéristiques qui demeurent stables au fil du temps.

- Pour les populations dont la distribution est très asymétrique, créer une strate de grandes unités qui seront incluses de façon certaine dans l'enquête. Normalement, ces grandes unités devraient représenter une partie importante des estimations des totaux de population.
- Parfois, l'information nécessaire à la stratification de la population n'est pas accessible dans la base. Dans de tels cas, on peut utiliser un plan d'échantillonnage à deux phases, dans lequel on sélectionne un grand échantillon dans la première phase afin d'obtenir l'information requise sur la stratification. Le premier échantillon est par la suite stratifié et, dans la deuxième phase, on choisit un sous-échantillon dans chacune des strates du premier échantillon. Examiner le coût de l'échantillonnage à chaque phase, la disponibilité de l'information requise à chaque phase, ainsi que le gain de précision tiré de la stratification de l'échantillon de la première phase.
- En pratique, et particulièrement en ce qui concerne les bases aréolaires, il est parfois difficile ou non rentable de sélectionner ou peu utile de sélectionner et de contacter directement les unités qui signaleront l'information demandée. Dans de tels cas, on peut utiliser un plan d'échantillonnage à deux degrés en choisissant d'abord les *grappes* (appelées les *unités primaires d'échantillonnage*) des unités déclarantes, puis en effectuant un sous-échantillonnage dans chacune des unités de sondage du premier degré sélectionnées pour obtenir un échantillon des unités déclarantes. Les contraintes budgétaires ou d'autres contraintes peuvent nécessiter plus de deux degrés. Déterminer le nombre d'étapes d'échantillonnage nécessaires et les unités d'échantillonnage appropriées à chaque étape. Pour chaque type d'unité possible, examiner la disponibilité d'une base adéquate d'unités à chaque étape ou la possibilité de créer une base pour l'enquête, la facilité de contact et de collecte/mesure des données, la qualité des données fournies par l'unité et le coût de la collecte.
- Dans la détermination de la taille de l'échantillon, tenir compte des niveaux de précision nécessaires à la production des estimations de l'enquête, du type de plan et d'estimateur à utiliser, de l'accessibilité à l'information auxiliaire, des contraintes budgétaires, ainsi que des facteurs d'échantillonnage (p. ex., mise en grappe, stratification) et des facteurs hors échantillonnage (p. ex., non-réponse, présence d'unités hors du champ de l'enquête, attrition dans les enquêtes longitudinales). Pour les enquêtes périodiques, tenir compte des naissances et des décès prévus liés aux unités au sein de la population observée en évolution.
- Il importe de se rappeler que la plupart des enquêtes produisent des estimations à l'égard d'un grand nombre de variables différentes et que le fait d'optimiser l'échantillon pour une variable particulière peut avoir des effets négatifs sur

d'autres variables importantes. Il faut résoudre ce problème en déterminant d'abord les variables les plus importantes, puis utiliser ce sous-ensemble de variables pour établir la stratégie d'échantillonnage à adopter, laquelle sous-tend souvent un compromis entre les stratégies optimales à utiliser pour les variables du sous-ensemble.

- Dans la détermination de la répartition et de la taille des échantillons stratifiés, tenir compte des taux d'erreur de classification prévus dans les unités et des autres anomalies de la base. Si elles sont prises à la légère à l'étape de l'échantillonnage, les estimations de l'enquête ne seront pas aussi précises que prévu. Il faut donc s'attaquer à ce problème à l'étape de l'estimation (voir la section 2.10).
- Mener des études pour évaluer les autres méthodes d'échantillonnage, les options de stratification et les possibilités de répartition. L'utilité de ces études est fonction de la disponibilité et de l'authenticité des données utilisées, qu'il s'agisse de données administratives ou d'enquêtes ou de recensements antérieurs, et de leur relation avec les variables importantes de l'enquête.
- À l'étape de la mise en œuvre, comparer la taille et les caractéristiques de l'échantillon réel par rapport aux attentes. Comparer la précision des estimations par rapport aux objectifs prévus.
- Pour les enquêtes périodiques qui sont fondées sur des plans dans lequel la taille de l'échantillon augmente parallèlement à la taille de la population, il est souvent approprié d'élaborer une méthode visant à maintenir stables la taille de l'échantillon et, par conséquent, les coûts de collecte. Généralement, la taille de l'échantillon total influe davantage que la *fraction d'échantillonnage* (ratio de la taille de l'échantillon par rapport à la taille de la population) sur la *précision* des estimations de l'enquête.
- Pour les enquêtes périodiques, élaborer un plan des plus souples afin de faire face aux changements futurs, comme les augmentations ou les réductions de la taille de l'échantillon, la restratification, le rééchantillonnage et l'actualisation des probabilités de sélection. Si des estimations sont requises pour des domaines d'intérêt précis (p. ex., les estimations infraprovinciales), former les strates en combinant de petites unités stables liées aux domaines identifiés (p. ex., petites régions géographiques), si possible. Ainsi, il sera plus facile de composer avec les futurs changements dans les définitions des strates.
- Pour les enquêtes périodiques, si des estimations efficaces du changement sont requises ou si le fardeau de réponse pose un problème, utiliser un plan d'échantillonnage avec *rotation* qui remplace une partie de l'échantillon dans chaque période. Le choix du taux de rotation sera un compromis entre la précision exigée pour les estimations du changement et le fardeau de réponse associé aux unités déclarantes. La réduction du taux de rotation augmentera la précision des estimations du changement, mais peut faire diminuer le taux de réponse au fil du

temps. Un taux de rotation faible offre l'avantage additionnel de réduire les coûts si le premier contact est beaucoup plus onéreux que les contacts subséquents.

- Pour les enquêtes périodiques, élaborer des procédures pour surveiller la qualité du plan d'échantillonnage au fil du temps. Mettre en place une stratégie d'actualisation pour le remaniement sélectif des strates qui ont subi une grave détérioration.
- En ce qui concerne les enquêtes à panel longitudinal, déterminer la longueur du panel (sa durée dans l'échantillon) en équilibrant la nécessité des données de durée et les effets d'attrition et de conditionnement. Utiliser un plan fondée sur des panels chevauchants (c.-à-d. ayant une durée en chevauchement) où il est nécessaire de produire des estimations transversales de pair avec les estimations longitudinales.
- Utiliser un logiciel de sélection d'échantillon généralisée plutôt que des systèmes personnalisés. Il peut s'agir du Système généralisé d'échantillonnage (SGÉCH) mis au point par Statistique Canada. Le SGÉCH est particulièrement utile pour gérer la sélection et la rotation des échantillons dans le cadre des enquêtes périodiques. Le logiciel MICROSTRATE est une autre solution développée par Eurostat pour contrôler le chevauchement des échantillons. En utilisant des systèmes généralisés, on peut s'attendre à commettre moins d'erreurs de programmation et réduire, dans une certaine mesure, les coûts et le temps de développement.

Références

- Bethel, J. (1989). Répartition de l'échantillon dans les enquêtes à plusieurs variables. **Techniques d'enquête**, 15, 49-60.
- Cochran, W.G. (1977). **Sampling Techniques**. Wiley, New York.
- Gambino, J.G., Singh, M.P., Dufour, J., Kennedy, B. et Lindeyer, J. (1998). **Méthodologie de l'Enquête sur la population active du Canada**. Publication n° 71-526-XPB au catalogue, Statistique Canada.
- Hidiroglou, M.A., (1994). Sampling and estimation for establishment surveys: stumbling blocks and progress. **Proceedings of the Section on Survey Research Methods**, American Statistical Association, 153- 162.
- Hidiroglou, M.A, et Srinath, K.P. (1993). Problems associated with designing sub-annual business surveys. **Journal of Economic Statistics**, 11, 397-405.
- Kalton, G. et Citro, C.F. (1993). Enquêtes par panel: ajout d'une quatrième dimension. **Techniques d'enquête**, 19, 217-227.

Kish, L. (1965). **Survey Sampling**. Wiley, New York.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). **Model Assisted Survey Sampling**. Springer-Verlag, New York.

Tillé, Y. (2001). **Théorie des sondages - Échantillonnage et estimation en populations finies**. Dunod, Paris.

2.5 Conception du questionnaire

Portée et objet

Un *questionnaire* est une série de questions conçues pour recueillir de l'information auprès d'un répondant. Le questionnaire peut être rempli par un intervieweur ou par le répondant lui-même, par des méthodes de collecte de données au moyen de papier et de crayons, par des méthodes téléphoniques ou par des modes assistés par ordinateur; la collecte de données peut se faire par Internet. Les questionnaires jouent un rôle de premier plan dans le processus de collecte des données. Ils influent considérablement sur le comportement des répondants, le rendement de l'intervieweur, le coût de collecte et les relations avec les répondants. Ainsi, ils déterminent la qualité des données.

Principes

La conception des questionnaires tient compte des besoins des utilisateurs de données en matière de statistique, des exigences administratives de l'organisation de l'enquête, des exigences du traitement des données, ainsi que de la nature et des caractéristiques de la population des répondants. Les questionnaires efficaces imposent un faible fardeau de réponse et sont conviviaux tant pour les intervieweurs que pour les répondants. Ils posent des questions pertinentes et permettent de recueillir les données avec efficacité et avec un minimum d'erreurs, tout en facilitant le codage et la saisie des données et en réduisant au minimum les tâches de vérification et d'imputation requises.

Les questionnaires rattachés aux enquêtes permanentes doivent être évalués périodiquement. Tous les questionnaires préparés à Statistique Canada doivent être mis à l'essai en français et en anglais avant d'être utilisés, en vertu de la Politique concernant l'examen et la mise à l'essai des questionnaires du Bureau (Statistique Canada, 2002a; voir annexe 1).

Lignes directrices

- Dans les questionnaires, utiliser des mots et des concepts qui ont la même signification pour les répondants et les concepteurs et, dans le cas des entreprises, choisir des questions, des périodes de référence en temps et des catégories de réponse qui sont compatibles avec les pratiques de tenue des livres de l'établissement. Dans la mesure du possible, harmoniser les concepts et la terminologie avec ceux qui sont déjà utilisés. S'il y a lieu, réutiliser les questions tirées des autres enquêtes.
- Choisir une conception des questions et une terminologie qui encouragent les répondants à remplir le questionnaire le plus exactement possible. Dans cet esprit, le questionnaire doit s'attarder au sujet de l'enquête, être le plus bref possible, comporter des questions qui suivent un ordre logique afin de faciliter le rappel des répondants et diriger ces derniers vers la source d'information appropriée.

- Dans l'introduction du questionnaire, indiquer le titre ou le sujet de l'enquête, identifier le commanditaire, expliquer l'objectif de l'enquête et demander la collaboration des répondants. Indiquer également l'autorité sous laquelle l'enquête est administrée et les mesures de protection de la confidentialité, les plans de couplage des enregistrements et les ententes de partage des données en place (Statistique Canada, 1998a). Les questions d'ouverture doivent s'appliquer à tous les répondants, être faciles à comprendre et intéressantes à répondre, et établir le fait que le répondant fait partie de la population observée.
- S'assurer que l'avantage de communiquer de l'information est clairement indiqué aux répondants et expliquer la raison pour laquelle il est important de remplir le questionnaire et la façon dont les données de l'enquête seront utilisées.
- Concevoir des questionnaires à remplir soi-même qui sont attrayants et faciles à remplir. À cette fin, donner une première impression positive dans la lettre d'accompagnement et la page couverture, et préparer un questionnaire qui semble professionnel. Si le questionnaire doit être administré par un intervieweur, il doit être convivial pour ce dernier.
- Pour réduire au minimum la possibilité d'erreurs de déclaration, s'assurer que les instructions à l'intention des répondants ou des intervieweurs sont courtes, précises et faciles à trouver. Donner des définitions au début du questionnaire ou à l'égard de questions précises, s'il y a lieu. S'assurer que les périodes de référence en temps et les unités de réponse sont claires pour le répondant, utiliser le caractère gras pour mettre en évidence les éléments importants, préciser « incluez » ou « excluez » dans les questions (et non dans des directives distinctes) et s'assurer que les catégories de réponse sont mutuellement exclusives et exhaustives.
- En ce qui concerne la présentation du questionnaire, donner des titres ou des entêtes à chaque section du questionnaire et intégrer des directives et des espaces de réponse qui facilitent la réponse exacte aux questions. Utiliser de la couleur, des ombrages, des illustrations et des symboles pour attirer l'attention des répondants ou des intervieweurs et pour les guider dans les parties du questionnaire qui doivent être lues et pour indiquer où doivent figurer les réponses. À la fin du questionnaire, offrir un espace pour les commentaires additionnels des répondants et intégrer une formule d'appréciation destinée aux répondants.
- Choisir parmi une vaste gamme de méthodes pour évaluer et mettre à l'essai le questionnaire. L'à-propos et l'intensité de son utilisation dépend de divers facteurs et de diverses circonstances, dont le type et l'ampleur de l'enquête, le contenu de l'enquête, l'utilisation de questions d'enquêtes antérieures ou de questions normalisées, le statut de collecte permanente ou non, la méthode de collecte des données, le calendrier du projet, le budget et l'accessibilité des ressources (Statistique Canada, 2002a; voir annexe 1).

- Envisager deux phases d'essai du questionnaire ou plus. Cela sous-tend la mise à l'essai du questionnaire à une des premières étapes de son élaboration, les révisions au questionnaire à la lumière des résultats, puis la mise à l'essai du questionnaire révisé. Ce processus peut être répété dans le cadre de deux ou trois phases d'essai, voire davantage. On peut utiliser différentes méthodes de mise à l'essai du questionnaire pendant chacune des phases.
- Effectuer des essais qualitatifs pour donner un aperçu de la façon dont les répondants réagissent au questionnaire. Parmi les méthodes, on trouve les groupes de discussion et les interviews approfondies, les méthodes cognitives comme les interviews non dirigées et la reformulation, ainsi que le codage des comportements. On utilise les groupes de discussion et les interviews approfondies en personne pour tester et évaluer la formulation, la séquence et le format des questions. On utilise les méthodes cognitives pour examiner les processus de réflexion des répondants au fur et à mesure qu'ils répondent aux questions de l'enquête et pour vérifier s'ils comprennent les questions et sont en mesure de donner des réponses exactes. Le codage des comportements offre un moyen systématique et objectif d'examiner l'efficacité du questionnaire par l'analyse de l'interaction entre l'intervieweur et le répondant. En outre, on peut effectuer des essais qualitatifs pour déterminer le contenu du questionnaire grâce à l'évaluation et à l'exploration des principaux concepts.
- Effectuer des essais officieux (essais préalables) du questionnaire afin de déceler les lacunes dans la formulation et l'ordre des questions, les erreurs dans la présentation ou dans les directives du questionnaire et les problèmes causés par l'incapacité ou le refus du répondant de répondre aux questions. Utiliser les essais officieux pour proposer des catégories de réponse additionnelles qui peuvent être précodées dans le questionnaire et pour donner une indication préliminaire de la durée de l'interview et des problèmes de non-réponse possibles.
- Après la mise à l'essai du questionnaire, organiser des séances de compte rendu avec les intervieweurs. Laisser les intervieweurs discuter de leur expérience d'interaction avec les répondants et du « rendement » du questionnaire. Ils peuvent nommer des sources de réponse potentielles et des erreurs de non-réponse, ainsi que les sections du questionnaire à améliorer.
- Effectuer des essais avec échantillon fractionné lorsqu'il existe au moins deux versions du questionnaire et qu'on veut déterminer laquelle est la « meilleure ». Cela sous-tend une conception expérimentale intégrée au processus de collecte des données et permettant d'examiner des aspects tels que la formulation des questions, la séquence des questions et les procédures de collecte des données.
- Effectuer des essais pilotes à la suite d'un essai approfondi du questionnaire afin d'observer la façon dont se déroulent, en pratique, les opérations des enquêtes, ce qui comprend l'administration du questionnaire. L'essai pilote reproduit la conception de l'enquête finale à petite échelle, des débuts jusqu'à la fin, ce qui

comprend le traitement et l'analyse des données. Cet essai offre la possibilité de perfectionner le questionnaire avant son utilisation dans le cadre de l'enquête principale.

- Vérifier les versions anglaise et française du questionnaire pour en assurer la correspondance.

Références

- Converse, J.M. et Presser, S. (1986). **Survey Questions: Handcrafting the Standardized Questionnaire**. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-063, Sage Publications, Thousand Oaks, California.
- DeMaio, T.J. (éd.) (1983). Approaches to developing questionnaires. Statistical Policy Working Paper 10, United States Office of Management and Budget, Washington, D.C.
- Fowler, F.J. Jr. (1995). **Improving Survey Questions: Design and Evaluation**. Applied Social Research Methods Series, 38, Sage Publications, Thousand Oaks, California.
- Gower, A.R. (1994). Conception des questionnaires d'enquêtes-entreprises. **Techniques d'enquête**, 20, 129-141.
- Platek, R., Pierre-Pierre, F.K. et Stevens, P. (1985). **Élaboration et conception des questionnaires d'enquête**. Publication n° 12-519F au catalogue. Statistique Canada.
- Schuman, H. et Presser, S. (1996). **Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context**. Sage Publications, Thousand Oaks, California.
- Statistique Canada (1998a). Politique d'information des répondants aux enquêtes. **Manuel de politiques**, 1.1.
- Statistique Canada (2002a). Politique concernant l'examen et la mise à l'essai des questionnaires. **Manuel des politiques**, 2.8. (Reproduite à l'annexe 1).
- Sudman, S. et Bradburn, N.M. (1982). **Asking Questions: A Practical Guide to Questionnaire Design**. Jossey-Bass, San Francisco.
- Tanur, J.M. (éd.) (1992). **Questions about Questions: Inquiries into the Cognitive Bases of Surveys**. Russell Sage Foundation, New York.

2.6 Réponse et non-réponse

Portée et objet

Malgré les plus grands efforts que fournissent les gestionnaires d'enquête et le personnel des opérations pour optimiser la réponse, il y aura toujours un certain taux de non-réponse. Pour qu'une unité soit classée comme répondante, le degré de *réponse d'item* ou de *réponse partielle* (où on obtient une réponse exacte seulement à l'égard de certaines données exigées du répondant) doit correspondre à un seuil minimal en deçà duquel la réponse serait rejetée et considérée comme une *non-réponse d'unité*. Dans un tel cas, la personne, le ménage, l'entreprise, l'institution, l'exploitation agricole ou toute autre unité échantillonnée est considérée comme n'ayant fourni aucune réponse.

La non-réponse a deux effets sur les données : premièrement, elle introduit un biais dans les estimations lorsque les non-répondants diffèrent des répondants par rapport aux caractéristiques mesurées; deuxièmement, elle contribue à faire augmenter la variance due à l'échantillonnage des estimations, car la taille observée de l'échantillon est réduite par rapport à la taille initialement prévue.

Principes

Le degré des efforts fournis pour obtenir une réponse d'un non-répondant est fonction des contraintes de budget et de temps, de son incidence sur la qualité générale et du risque de biais de non-réponse. Si la non-réponse persiste, on apporte des rajustements ultérieurs aux données pour contrebalancer l'effet de non-réponse. Les décisions concernant le degré de recherche convenable à instaurer pour élaborer des techniques de rajustement de la non-réponse sont susceptibles d'être influencées par les questions de budget, de temps, d'utilisation des données et de risque de biais. On effectue une évaluation de la non-réponse pour donner des commentaires au personnel de l'enquête sur les mesures immédiates et futures. La non-réponse est également signalée aux utilisateurs des données de l'enquête. Un programme efficace de relations avec les répondants et un questionnaire bien conçu sont des éléments essentiels à l'optimisation de la réponse (voir la section 2.5).

Lignes directrices

- On obtient un bon *taux de réponse* en s'assurant un degré de qualité acceptable durant toutes les étapes de planification et de mise en œuvre de l'enquête. Il faut adopter une approche intégrée afin de ne pas reproduire les techniques de gestion de la non-réponse. Pour obtenir le taux de réponse souhaité, il faut garder à l'esprit les facteurs suivants :
 - la qualité de la base de sondage (sur le plan de la couverture de la population et de la facilité à établir le contact avec le répondant);
 - la population observée;
 - la méthode de collecte des données (par exemple, par la poste, par une interview personnelle, par une interview téléphonique, par une interview

- assistée par ordinateur);
 - la méthode d'échantillonnage;
 - la période de l'année et la longueur de la période de collecte;
 - le fardeau de réponse imposé (longueur de l'interview, difficulté du sujet, choix du moment et périodicité de l'enquête);
 - la nature du sujet (la délicatesse des sujets);
 - la longueur et la complexité du questionnaire;
 - l'efficacité et la portée de la méthodologie de suivi;
 - les difficultés prévues dans le dépistage des répondants qui ont déménagé;
 - l'expérience antérieure du même type d'enquête;
 - l'expérience antérieure et les compétences prouvées du personnel de collecte;
 - la charge de travail du personnel de collecte;
 - les relations établies avec les répondants;
 - la stratégie de communications;
 - le budget total;
 - l'affectation du budget entre les diverses opérations;
 - la langue du questionnaire;
 - les antécédents culturels des répondants;
 - l'importance de l'enquête pour les utilisateurs et les répondants;
 - les facteurs liés aux intervieweurs eux-mêmes, comme la formation, l'expérience, les compétences en relations interpersonnelles, l'établissement de relations et le roulement;
 - l'utilisation et l'efficacité des mesures incitatives pour les répondants.
- Utiliser un essai préalable ainsi que des cycles des mêmes enquêtes ou d'enquêtes semblables antérieures, entre autres moyens, pour établir le taux de réponse souhaité.
 - Lorsque les contraintes opérationnelles le permettent, effectuer un suivi auprès des non-répondants (tous ou un sous-échantillon de ces derniers). Le suivi auprès des non-répondants augmente le taux de réponse et peut aider à vérifier si les répondants et les non-répondants offrent des caractéristiques mesurées semblables. Un tel suivi est particulièrement important dans le cadre des enquêtes longitudinales, où il est manifeste que l'investissement vise davantage le long terme et où l'échantillon est assujéti à l'attrition croissante (et possiblement au biais) en raison de la non-réponse dans chacun des cycles d'enquête. Dans ce cas, les activités de dépistage sont particulièrement importantes.
 - Pour les enquêtes longitudinales, il faut faciliter le dépistage de grande qualité. Obtenir des données de contact additionnelles pour les unités échantillonnées à chaque cycle d'enquête. Fournir une carte de changement d'adresse et demander à l'unité échantillonnée d'informer le Bureau si un déménagement a lieu entre les cycles d'enquête. Cela permettra d'obtenir des données de contact actualisées. En outre, les données administratives, les annuaires municipaux et téléphoniques, et

de nombreuses autres sources, dont le savoir local, sont précieux pour le personnel de dépistage.

- Établir la priorité des activités de suivi. Par exemple, dans les enquêtes-entreprises, effectuer d'abord un suivi des grandes unités ou des unités influentes, possiblement au risque de manquer les plus petites unités (voir la section 2.8). De même, accorder une plus grande priorité aux unités non-répondantes dans les domaines comportant un fort potentiel de biais de non-réponse. On peut utiliser une fonction « score » pour établir la priorité du suivi.
- Inscrire et évaluer les raisons de la non-réponse (p. ex., refus, non-contact, absence temporaire, problème technique). Le degré de biais de non-réponse peut différer en fonction de la raison. Évaluer la non-réponse par raison.
- Comme les différences entre les répondants et les non-répondants peuvent donner lieu à des biais dans les estimations, essayer de déterminer si de telles différences sont présentes. Même si cela est difficile à faire, on peut y arriver en partie en utilisant des sources de données externes (par exemple, les fichiers de données administratives) et en partie en examinant les réponses des non-répondants qui ont été convertis au cours d'une activité de suivi. Bien souvent, il est plus facile de comparer les caractéristiques connues des répondants et des non-répondants pour établir l'ampleur des différences. Dans le cas des enquêtes longitudinales (ou avec groupes de renouvellement), il est possible d'analyser les caractéristiques connues des répondants au cours d'un des cycles de l'enquête afin de comparer les caractéristiques des répondants et des non-répondants au cours d'un cycle subséquent. L'information ainsi obtenue peut influencer sur les méthodes de compensation de la non-réponse.
- Voici les deux principales approches servant à composer avec les données manquantes (Kalton et Kasprzyk, 1986) : rajuster le poids d'échantillonnage (voir la section 2.10) ou utiliser l'imputation (voir la section 2.9). S'il y a lieu, tenter d'évaluer la mesure dans laquelle les procédures corrigent le biais potentiel. Tenir compte de la non-réponse dans la production des estimations et de leurs estimations de variance connexes.
- Indiquer les taux de réponse et de non-réponse (Statistique Canada, 2000d; voir annexe 2). À Statistique Canada, on a élaboré des normes et des lignes directrices sur l'indication de la non-réponse (Statistique Canada, 2001d). Informer régulièrement les utilisateurs du taux de non-réponse en fournissant les estimations. Indiquer les taux de non-réponse pondérés et non pondérés à l'étape de l'estimation dans la Base de métadonnées intégrée (BMDI). Essayer de se conformer à la norme d'indication de la non-réponse afin de faciliter la comparabilité entre les enquêtes. Les lignes directrices stipulent que toutes les unités doivent être classées comme répondantes ou non répondantes. Indiquer clairement que des unités ont répondu partiellement et la façon dont ces unités ont été classées.

Références

- Cialdini, R., Couper, M. et Groves, R.M. (1992). Understanding the decision to participate in a survey. **Public Opinion Quarterly**, 56, 475-495.
- Couper, M.P. et Groves, R.M. (1992). Le rôle de l'intervieweur dans la participation aux enquêtes. **Techniques d'enquête**, 18, 279-294.
- Federal Committee on Statistical Methodology (2001). Measuring and reporting sources of error in surveys. Statistical Policy Working Paper 31. Voir aussi <http://www.fcsm.gov>.
- Groves, R.M., Dillman, D. A., Eltinge, J. L. et Little, R. J. A. (2002). **Survey Nonresponse**. Wiley, New York.
- Kalton, G. et Kasprzyk, D. (1986). Le traitement des données d'enquête manquantes. **Techniques d'enquête**, 12, 1-17.
- Latouche, M. et Michaud, S. (1997). Concerns pertaining to weighting of longitudinal surveys. **Proceedings of the Section on Government Statistics and Section on Social Statistics**, American Statistical Association, 111-119.
- Lévesque, I. et Franklin, S. (2000). Pondérations longitudinale et transversale de l'Enquête sur la dynamique du travail et du revenu, année de référence 1997. Statistique Canada, document de travail n°, 75F0002MIF - 00004.
- Statistique Canada (2000d). Politique visant à informer les utilisateurs de la qualité des données et la méthodologie. **Manuel des politiques**, 2.3. (Reproduite à l'annexe 2).
- Statistique Canada (2001d). Normes et lignes directrices pour la déclaration des taux de non-réponse. Rapport technique de Statistique Canada.
- Swain, L. et Dolson, D. (1997). Current issues in household survey nonresponse at Statistics Canada. **Statistics in Transition**, 3, 439-468.
- Tambay, J. L., Schiopu-Kratina, I., Mayda, J., Stukel, D. et Nadon, S. (1998). Traitement de la non-réponse du cycle deux de l'enquête nationale sur la santé de la population. **Techniques d'enquête**, 24, 159-169.

2.7 Opérations de collecte et de saisie des données

Portée et objet

La *collecte des données* renvoie à tout processus dont l'objectif est d'acquérir ou de faciliter l'acquisition des données. On procède à la collecte en demandant et en obtenant des données pertinentes auprès de personnes ou d'organismes par le truchement d'un mécanisme adéquat (voir la section 2.5). Si aucune information n'est obtenue au départ ou si les données sont jugées non convenables lors de la vérification préliminaire, on peut effectuer des contacts de suivi dans le cadre de la collecte des données (voir la section 2.8).

La *saisie des données* renvoie à tout processus qui convertit l'information communiquée par un répondant en un format électronique convenant à une utilisation dans des processus subséquents. Parfois, les données sont saisies dans le cadre du processus de collecte dans les enquêtes qui utilisent des instruments comme l'IPAO, l'ITAO et la CÉD. À d'autres moments, il faut procéder à une opération distincte pour saisir les données par introduction manuelle sur clavier ou par des moyens automatisés (p. ex., RIC). Bien souvent, cette conversion des données sous-tend un *codage* manuel ou automatisé et parfois, la transmission des données à un autre lieu.

L'incidence des opérations de collecte et de saisie des données sur la qualité des données est directe et essentielle, car ces données sont les principaux intrants d'un organisme de réalisation d'enquêtes. Par conséquent, la qualité de ces opérations influe grandement sur la qualité du produit final.

Principes

Les répondants, ou fournisseurs de données, et particulièrement les personnes et les organismes qui remplissent les questionnaires sans être payés, sont les ressources les plus précieuses d'un organisme de réalisation d'enquêtes. Pour assurer une collaboration constante, il est essentiel de réduire au minimum le fardeau des répondants. La meilleure façon de remédier aux lacunes ou aux incohérences des données est de consulter les répondants eux-mêmes au cours de la collecte de données ou très tôt après cette collecte. Compte tenu que les opérations de collecte et de saisie des données ont une forte incidence sur la qualité des données, on recommande vivement d'utiliser des outils de mesure de la qualité et du rendement pour gérer ces processus et de communiquer des mesures objectives aux superviseurs et aux clients. En cours de processus, il faut prendre les mesures appropriées pour conserver la confidentialité de l'information recueillie (voir la section 2.13).

Lignes directrices

- Les intervieweurs et les opérateurs de saisie des données sont essentiels au succès de la majorité des opérations de collecte et de saisie des données. Il faut s'assurer

qu'ils possèdent la formation et les outils adéquats (p. ex., manuels de formation, voir Burgess et Brierley, 1995).

- Exploiter la technologie accessible pour rehausser l'efficacité et la qualité des processus de collecte et de saisie des données. Les progrès en communications et en technologie informatique offrent des possibilités de réduire grandement les coûts et les risques que comportent ces processus. Par exemple, l'interview d'enquête assistée par ordinateur (p. ex., IPAO et ITAO) et la collecte électronique des données (CÉD) par Internet, l'entrée de données informatisée (au moyen du RIC) et le codage automatisé par reconnaissance de texte (CART) sont des approches qui bénéficient des technologies accessibles.
- Contrôler attentivement les opérations de livraison des questionnaires papier au cours des enquêtes par la poste pour veiller à ce que chaque unité qui a été sélectionnée dans l'enquête reçoive le questionnaire qui convient. Lorsque le questionnaire est retourné, vérifier l'exactitude de l'information de couverture et la qualité des données fournies. Dans certains cas, il peut être nécessaire d'effectuer des interviews de suivi. Lorsque le questionnaire n'est pas reçu, il est nécessaire d'entreprendre des activités de suivi pour établir le statut de l'unité (p. ex., occupé ou non occupé; en exploitation ou fermé) et d'obtenir l'information manquante. À toutes ces étapes, mettre en place un système pour indiquer l'état d'achèvement de chaque unité.
- Établir des procédures adéquates de contrôle des échantillons à l'égard de toutes les opérations de collecte des données. Ces procédures indiquent le statut des unités échantillonnées du début à la fin de la collecte des données, de sorte que les gestionnaires de la collecte des données et les intervieweurs peuvent évaluer la progression à tout moment. On utilise également des procédures de contrôle des échantillons et les commentaires en découlant pour s'assurer que chaque unité échantillonnée est traitée à toutes les étapes de la collecte et de la saisie des données et qu'on en indique le statut final.
- Mettre en place des systèmes de contrôle efficaces pour assurer la sécurité de la saisie, de la transmission et de la manipulation des données. Prévenir la perte d'information et la perte de qualité qui en découle attribuables aux pannes de système ou aux erreurs humaines.
- Au cours de la collecte des données, veiller à communiquer au moment convenu avec le répondant ou la personne désignée du ménage ou de l'organisme répondant. De cette façon, l'information sera rapidement accessible. Laisser le répondant communiquer les données selon la méthode et dans le format qui lui conviennent ou qui conviennent à son organisme. Cela permettra de faire augmenter les taux de réponse et rehaussera la qualité des renseignements obtenus auprès des répondants.

- Pendant la conception des processus de collecte des données, et particulièrement pendant la vérification et le codage, s'assurer que les procédures sont appliquées à toutes les unités d'étude le plus uniformément possible et qu'elles comportent le moins d'erreurs possible. L'automatisation est préférable. Permettre au personnel ou aux systèmes de soumettre les cas difficiles à un petit groupe de spécialistes compétents. Centraliser le traitement afin de réduire les coûts et de faciliter l'apport des connaissances spécialisées disponibles. Comme l'information recueillie peut donner lieu à des résultats imprévus, utiliser des processus adaptables pour apporter les changements qui s'imposent s'il y a lieu de le faire du point de vue de l'efficacité.
- Effectuer un suivi de la fréquence des rejets à la vérification, du nombre et du type de corrections apportées par strate, du mode de collecte, du type de traitement, des données élémentaires et de la langue utilisée pour la collecte. Cela aidera à évaluer la qualité des données et l'efficacité de la fonction de vérification.
- Les mesures des dépenses, du rendement et de la qualité établies pendant l'opération de collecte des données permettent au gestionnaire d'enquête de prendre des décisions sur la nécessité de modifier ou de restructurer le processus. Effectuer un suivi des coûts postaux réels, des appels téléphoniques, de la production de l'instrument de collecte, de l'informatisation et de la consommation en jours-personnes. Les mesures de la qualité importantes comprennent les taux de réponse, les taux d'erreur de traitement, les taux de suivi et la répartition de la non-réponse selon la cause. Lorsque ces mesures sont connues à tous les niveaux auxquels on produit des estimations et à toutes les étapes du processus, elles peuvent servir de mesures du rendement et de mesures de la qualité des données (voir la section 2.6).
- La saisie manuelle des données tirées des questionnaires papier ou d'images balayées par scanner est susceptible d'entraîner des erreurs d'introduction sur clavier. Intégrer des vérifications en ligne pour les erreurs que l'opérateur de saisie des données peut corriger (c.-à-d. vérifications qui mettront en évidence les erreurs d'introduction sur clavier). Enregistrer ces cas pour analyse et examen ultérieurs.
- Mettre en œuvre des procédures de vérification pour évaluer dans quelle mesure les opérateurs respectent les niveaux préétablis concernant les taux d'erreur d'introduction sur clavier.
- Utiliser des méthodes de contrôle de la qualité statistique pour évaluer et améliorer la qualité des opérations de collecte et de saisie. Examiner et analyser les mesures et les résultats du contrôle de la qualité de façon à faciliter la détermination des principales causes fondamentales des erreurs. Fournir des rapports de rétroaction pour les gestionnaires, le personnel, les spécialistes du sujet et les méthodologistes. Ces rapports doivent contenir de l'information sur la fréquence et les sources des erreurs (voir Mudryk et coll., 1994, 1996 et 2002;

Mudryk et Xiao, 1996). Divers outils logiciels existent pour faciliter cette tâche, dont le Système d'analyse des données du contrôle qualitatif (SADCQ) et NWA Quality Analyst (voir Mudryk, Bougie et Xie, 2002).

- Utiliser des mesures de la qualité et de la productivité pour communiquer des commentaires aux intervieweurs ou aux opérateurs, ainsi que pour déterminer les éléments inducteurs d'erreurs dans la conception de l'instrument de collecte ou des procédures de traitement connexes.
- Utiliser des processus d'enquête subséquents pour recueillir de l'information utile sur la qualité, qui servirait de signal indiquant que les procédures et les outils de collecte et de saisie peuvent nécessiter des changements en prévision des cycles d'enquête futurs. Par exemple, les étapes de la vérification ou de l'analyse des données (voir les sections 2.8 et 2.15) peuvent laisser supposer qu'il existe un biais de réponse ou d'autres problèmes liés à la collecte.
- Effectuer une évaluation rétrospective de toutes les opérations de collecte et de saisie des données, et documenter les résultats en prévision d'une utilisation ultérieure.

Références

- Burgess, M.J. et Brierly, R. (1995). A self-directed training course for monitors of CATI operations. Division de la recherche et du développement des opérations, Statistique Canada.
- Couper, M.P., Baker, R.P., Bethlehem, J., Clark, C.Z.F., Martin, J., Nicholls II, W.L. et O'Reilly, J. (éd.) (1998). **Computer Assisted Survey Information Collection**. Wiley, New York.
- Dielman, L. et Couper, M.P. (1995). Data quality in a CAPI survey: keying errors. **Journal of Official Statistics**, 11, 141-146.
- Dufour, J. (1996). Qualité des données à l'Enquête sur la population active. Statistique Canada, document de travail de la Direction de la méthodologie n° HSMD-96-002E/F.
- Dufour, J., Kaushal, R., Clark, C. et Bench, J. (1995). Converting the Labour Force Survey to computer-assisted interviewing. Statistique Canada, document de travail de la Direction de la méthodologie n° HSMD-95-009E.
- Groves, R.M. (1989). **Survey Errors and Survey Costs**. Wiley, New York.
- Groves, R.M., Biemer, P., Lyberg, L., Massey, J., Nicholls, W. et Waksberg, J. (éd.) (1988). **Telephone Survey Methodology**. Wiley, New York.

- Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N. et Trewin, D. (éd.) (1997). **Survey Measurement and Process Quality**. Wiley, New York.
- Mudryk, W. et Xie, H. (2002). Quality control application in ICR data capture for the 2001 Census of Agriculture. **Proceedings of the Section on Quality and Productivity**, American Statistical Association, 2424-2429.
- Mudryk, W. et Xiao, P. (1996). Quality control methodology for LFS industry and occupation coding operations. Rapport technique de Statistique Canada.
- Mudryk, W., Bougie, B. et Xie, H. (2002). Some guidelines for data analysis in quality control. Rapport technique de Statistique Canada.
- Mudryk, W., Burgess, M.J. et Xiao, P. (1996). Quality control of CATI operations in Statistics Canada. **Proceedings of the Section on Survey Research Methods**, American Statistical Association, 150-159.
- Mudryk, W., Croal, J. et Bougie, B. (1994). Système généralisé de collecte et de saisie des données (DC2): version 2.5.1, Comparaison des échantillons (CÉ). Rapport technique de Statistique Canada.
- Williams, K., Denyes, C., March, M. et Mudryk, W. (1996). Mesure de la qualité dans le traitement des données d'enquête. **Recueil du Symposium 96 : Erreurs non dues à l'échantillonnage**, Statistique Canada, 131-141.

2.8 Vérification

Portée et objet

La *vérification* des données est l'application de contrôles visant à détecter les entrées manquantes, invalides ou incohérentes ou à mettre en évidence les enregistrements de données qui sont susceptibles de contenir des erreurs. Certains de ces contrôles sous-tendent des relations logiques qui découlent directement des concepts et des définitions. D'autres sont de nature plus empirique ou sont obtenus par le résultat de l'application d'essais ou de procédures statistiques (p. ex., des techniques d'analyse des valeurs aberrantes). Les contrôles peuvent être fondés sur des données tirées de collectes antérieures de la même enquête ou d'autres sources.

La vérification englobe une vaste gamme d'activités, dont les vérifications des intervieweurs sur le terrain, les avertissements générés par ordinateur au moment de la collecte ou de la saisie des données, la détermination des unités en prévision du suivi, les vérifications de relations complexes, la localisation des erreurs pour les besoins de l'imputation et la validation des données. On abordera les deux derniers sujets dans les sections 2.9 et 2.12.

Principes

La vérification vise trois objectifs (Granquist, 1984) : servir de fondement à l'amélioration future de l'instrument d'enquête, fournir de l'information sur la qualité des données de l'enquête et purger les données. Nous avons de bonnes raisons de penser qu'une quantité disproportionnée de ressources est concentrée sur le troisième objectif, soit celui de « nettoyer les données ». Par conséquent, l'apprentissage tiré du processus de vérification joue souvent un rôle indu et secondaire.

On reconnaît que les *erreurs fatales* (p. ex., les entrées invalides ou incohérentes) devraient être retirées dans ensembles de données afin de maintenir la crédibilité du Bureau et de faciliter l'analyse et le traitement automatisés des données dans l'avenir. Toutefois, il faut faire preuve de prudence à l'égard de la surutilisation des *vérifications exploratoires* (celles qui mettent en évidence les enregistrements douteux qui peuvent contenir des erreurs). La vérification des données est fort probablement l'activité la plus coûteuse d'une enquête par sondage ou d'un cycle de recensement. On estime qu'elle représente au moins le quart du budget total de l'enquête, en moyenne, atteignant jusqu'à 40 % dans le cas des enquêtes-entreprises (Gagnon, Gough et Yeo, 1994). Lorsque cette vérification laborieuse, et souvent manuelle, a une incidence négligeable sur les estimations finales, on parle de *survérification*. Outre le fait qu'elle soit coûteuse sur le plan des finances et de la ponctualité et qu'elle augmente le fardeau de réponse, la survérification peut donner lieu à de graves biais, engendrés par l'intégration de données dans les modèles implicites qu'imposent les vérifications.

Lignes directrices

- S'assurer que toutes les vérifications sont cohérentes à l'interne (c.-à-d. non contradictoires).
- Appliquer de nouveau les vérifications aux unités qui ont subi des corrections pour s'assurer qu'aucune autre erreur n'a été introduite directement ou indirectement par le processus de correction.
- La vérification contribue efficacement à détecter les erreurs fatales (Granquist et Kovar, 1997) - puisque le processus peut facilement être informatisé. Exécuter cette activité le plus rapidement possible. Bien qu'une certaine intervention manuelle soit nécessaire, un logiciel généralisé et réutilisable peut être particulièrement utile pour cette tâche. Banff - la version SAS du Système généralisé de vérification et d'imputation (Statistique Canada, 2000a) - et SCANCIR - le Système canadien de contrôle et d'imputation du recensement (Bankier et coll., 1999) - sont des exemples de ce type de logiciel.
- Les *taux de succès* des vérifications, soit la proportion des vérifications d'avertissement ou d'interrogation qui mettent en évidence les véritables erreurs, se sont avérés peu efficaces, souvent aussi bas que 20 ou 30 % (Linacre et Trewin, 1989). En outre, l'incidence des erreurs s'est avérée très différentielle, particulièrement dans les enquêtes recueillant des données numériques. En d'autres mots, il est fréquent qu'un petit nombre d'erreurs soit à la source de la majorité des changements apportés dans les estimations. Envisager d'effectuer la vérification de façon sélective afin de réaliser des gains d'efficacité potentiels (Granquist et Kovar, 1997) sans incidence négative sur la qualité des données. Les priorités peuvent être établies en fonction des types ou de la gravité des erreurs ou en fonction de l'importance de la variable ou de l'unité déclarante.
- Pour les enquêtes-entreprises, élaborer une stratégie de suivi sélectif. L'utilisation d'une fonction de pointage (Latouche et Berthelot, 1992) concentre les ressources sur les unités d'échantillon importantes, les principales variables et les erreurs les plus graves.
- Il faut se rappeler que l'utilité de la vérification est limitée et que le processus peut en fait être improductif (voir, par exemple, Linacre et Trewin, 1989). Bien souvent, les changements de données fondés sur les vérifications sont considérées à tort comme des corrections de données. On peut prétendre qu'à un certain point du processus de vérification, on introduit autant d'erreurs qu'on en corrige. Identifier et respecter cette fin logique du processus.
- L'informatisation permet aux gestionnaires d'enquête d'augmenter la portée et le volume des contrôles qu'on peut effectuer, ce qui est tentant pour eux. Minimiser le nombre de ces augmentations si elles font peu de différence dans les estimations de l'enquête. Plutôt que de hausser l'effort de vérification, réorienter

les ressources vers des activités plus rentables (p. ex., l'analyse des données, l'analyse des erreurs de réponse).

- Limiter le recours à la vérification pour résoudre les problèmes déjà survenus, surtout dans le cas des enquêtes répétitives. La contribution de la vérification à la réduction des erreurs est limitée. Bien qu'il soit essentiel d'effectuer un peu de vérification, il faut en réduire la portée et réorienter l'objectif. Attribuer une grande priorité à l'apprentissage tiré du processus de vérification. Pour réduire le nombre d'erreurs, s'attarder aux premières phases de la collecte de données plutôt qu'au nettoyage effectué à la fin. Pratiquer la prévention plutôt que la correction des erreurs. Dans cet esprit, ramener l'étape de vérification aux premières phases du processus d'enquête, de préférence lorsque le répondant est encore disponible, par exemple, en utilisant les méthodes d'interview téléphonique, d'interview sur place ou d'auto-interview assistées par ordinateur.
- Il est possible que les vérifications ne permettent pas de détecter les petites erreurs systématiques introduites constamment dans les enquêtes répétitives, erreurs qui peuvent donner lieu à d'importants biais dans les estimations. Le « resserrement » des vérifications n'est pas la solution. Pour détecter ce genre d'erreurs systématiques, utiliser d'autres méthodes, comme les méthodes classiques de contrôle de la qualité, l'analyse et l'examen approfondis des concepts et des définitions, les études ultérieures aux interviews, la validation des données, la confrontation des données (voir la section 2.17) avec d'autres sources de données qui peuvent être disponibles pour certaines unités, etc.
- Déterminer les valeurs des données extrêmes d'une période d'enquête ou entre les périodes d'enquête (cet exercice est appelé le processus de détection des valeurs aberrantes). La présence de ce type de données se démarquant de la distribution est un signe précurseur d'erreurs potentielles. Utiliser des méthodes de détection univariées simples (Hidioglou et Berthelot, 1986) ou des méthodes plus complexes et explicites (de Waal, 2000).
- Au cours des suivis, ne pas surestimer la capacité des répondants de signaler ou de corriger les rapports. Leur agrégation peut être différente, leur mémoire limitée et leur apport négligeable. Limiter l'activité de suivi des répondants.
- Ne pas sous-estimer la capacité du processus de vérification d'intégrer les données signalées aux modèles qu'imposent les vérifications. Il y a un réel danger de créer de faux changements pour la seule raison de s'assurer que les données sont soumises aux vérifications. Contrôler le processus!
- Le processus de vérification est souvent très complexe. Lorsque la vérification se fait sous le contrôle du Bureau, communiquer les procédures détaillées et actualisées en offrant une formation adéquate à tout le personnel concerné et effectuer un suivi des travaux en soi. Envisager d'appliquer des procédures de contrôle de la qualité.

- La vérification peut être utile à l'épuration de certaines données, mais son rôle le plus utile est de permettre de fournir de l'information sur le processus d'enquête, soit en donnant des mesures de qualité pour l'enquête en cours, soit en suggérant des améliorations pour les enquêtes futures. Envisager la vérification comme une partie intégrante du processus de collecte des données dont l'objectif est de recueillir des renseignements sur le processus. Dans cette optique, la vérification peut être précieuse pour raffiner les définitions, pour améliorer l'instrument d'enquête, pour évaluer la qualité des données, pour déterminer les sources des erreurs non dues à l'échantillonnage, pour servir de fondement à l'amélioration future du processus d'enquête complet et pour alimenter le cycle d'apprentissage continu. Afin d'atteindre cet objectif, il faut superviser le processus et produire des pistes de vérification, des diagnostics et des mesures de rendement, et utiliser ces éléments pour établir les meilleures pratiques.

Références

- Bankier, M., Lachance, M. et Poirier, P. (1999), A generic implementation of the New Imputation Methodology. **Proceedings of the Survey Research Methods Section**, American Statistical Association, 548-553.
- De Waal, T., Van de Pol, F. et Renssen, R. (2000). Graphical macro editing: possibilities and pitfalls. **Proceedings of the Second International Conferences on Establishment Surveys**, Buffalo, New York.
- Gagnon, F., Gough, H. et Yeo, D. (1994). Survey of editing practices in Statistics Canada. Rapport technique de Statistique Canada.
- Granquist, L. (1984). On the role of editing. **Statistisk tidskrift**, 2, 105-118.
- Granquist, L. et Kovar, J.G. (1997). Editing of survey data: how much is enough? Dans **Survey Measurement and Process Quality**, Lyberg et coll. (éd.), Wiley, New York, 415-435.
- Hidioglou, M. A. et Berthelot, J.-M. (1986). Contrôle statistique et imputation dans les enquête-entreprises périodiques. **Techniques d'enquête**, 12, 79-89.
- Latouche, M. et Berthelot, J.-M. (1992). Use of a score function to prioritize and limit recontacts in editing business surveys. **Journal of Official Statistics**, 8, 389-400.
- Linacre, S. J. et Trewin, D. J. (1989). Evaluation of errors and appropriate resource allocation in economic collections. **Proceedings of the Annual Research Conference**, U.S. Bureau of the Census, 197-209.
- Statistique Canada (2000a). Description des fonctions du système généralisé de vérification et d'imputation. Rapport technique de Statistique Canada.

2.9 Imputation

Portée et objet

L'*imputation* est le processus utilisé pour déterminer et attribuer les valeurs de remplacement aux données manquantes, invalides et incohérentes qui ont été rejetées lors de la vérification. Pour ce faire, on change certaines des réponses ou on attribue des valeurs à ces dernières lorsqu'elles sont absentes de l'enregistrement vérifié afin de s'assurer que les estimations sont de bonne qualité et qu'on crée un enregistrement plausible et cohérent à l'interne. Un grand nombre de ces problèmes sont préalablement résolus par le suivi auprès du répondant ou par l'examen et la correction manuelle du questionnaire. Toutefois, il est généralement impossible de résoudre tous les problèmes au cours des premières étapes en raison des questions liées au fardeau de réponse, au coût et à l'actualité. Comme il est habituellement préférable de produire un fichier de microdonnées complet et cohérent contenant les données imputées, on utilise l'imputation pour traiter le reste des rejets lors de la vérification.

Bien que l'imputation puisse rehausser la qualité des données finales grâce à la correction des réponses manquantes, invalides ou incohérentes, il faut prendre soin de choisir une méthode d'imputation appropriée. Certaines méthodes d'imputation ne préservent pas les liens entre les variables et peuvent engendrer des distorsions au niveau des distributions sous-jacentes. Par conséquent, il faut tenir compte de l'imputation au cours de la production des estimations et des estimations de variance connexes.

Principes

Les personnes qui ont pleinement accès aux microdonnées et qui possèdent une bonne information auxiliaire sont les plus aptes à effectuer l'imputation. Cette dernière peut être informatisée, manuelle, ou une combinaison des deux. Une imputation efficace vise à tenter de réduire le biais introduit par le fait que toutes les valeurs souhaitées n'ont pas été observées. L'imputation efficace comporte une piste de vérification pour les besoins de l'évaluation et permet d'assurer que les enregistrements imputés sont cohérents à l'interne. Les bons processus d'imputation sont informatisés, objectifs, reproductibles et efficaces. Selon les principes de Fellegi-Holt (1976), on apporte des changements à un nombre minimal de champs afin de s'assurer que l'enregistrement complet rencontre chacune des règles de vérification.

On peut classer les méthodes d'imputation comme déterministes ou stochastiques, selon que les données imputées soient choisies de façon aléatoire ou non (Kalton et Kasprzyk, 1986; Kovar et Whitridge, 1995). Parmi les méthodes d'*imputation déterministes*, on trouve l'imputation logique, l'imputation historique, l'imputation par la moyenne, l'imputation par le ratio, l'imputation par la régression et l'imputation par le plus proche voisin. Ces méthodes peuvent elles-mêmes être divisées en méthodes strictement fondées sur la déduction de la valeur imputée à partir des données disponibles sur le non-répondant et les autres données auxiliaires (logiques et historiques) et en méthodes qui sont fondées sur les données observées chez les autres répondants d'une enquête donnée.

On peut directement utiliser les données observées dans les unités répondantes en transférant les données d'un enregistrement donneur choisi ou en utilisant des modèles (ratio et régression). Les méthodes d'*imputation stochastique* comprennent le « hot deck », l'imputation par la méthode du plus proche voisin où on effectue une sélection aléatoire à partir de plusieurs des plus proches voisins, la régression avec résidus aléatoires et tout autre méthode déterministe employant des résidus aléatoires.

Lignes directrices

- Évaluer le type de non-réponse, c'est-à-dire tenter de déterminer les variables auxiliaires qui peuvent expliquer le ou les mécanismes de non-réponse afin de les utiliser pour enrichir la méthode d'imputation. Intégrer ces variables auxiliaires dans la méthode d'imputation.
- Élaborer et mettre à l'essai avec soin l'approche d'imputation. Étudier la qualité et la pertinence des variables disponibles afin de déterminer celles qu'on peut utiliser comme variables auxiliaires ou variables de couplage ou pour établir des classes d'imputation. À cette fin, consulter des spécialistes du sujet et utiliser des techniques de modélisation.
- Tenir compte du type d'estimations à produire, comme les niveaux par rapport aux variations, les agrégats de niveau supérieur par rapport aux petits domaines et l'enquête transversale par rapport à l'enquête longitudinale.
- Tenter de faire en sorte que l'enregistrement imputé ressemble étroitement à l'enregistrement de rejet au contrôle. Pour ce faire, on impute le nombre minimal de variables dans un sens, ce qui permet de conserver le plus de données de répondant possible. L'hypothèse implicite est qu'un répondant est plus susceptible de commettre seulement une ou deux erreurs plutôt que plusieurs, même si cela n'est pas toujours vrai en pratique. Préparer des enregistrements imputés qui sont cohérents à l'interne.
- Dans le cadre de certaines enquêtes, il est nécessaire d'utiliser différents types de méthode d'imputation. Habituellement, on établit une hiérarchie informatisée des méthodes. Limiter le nombre des échelons et élaborer avec soin les méthodes utilisées à chaque échelon hiérarchique, en n'oubliant pas de les mettre à l'essai. De même, lorsqu'il faut regrouper des classes d'imputation, élaborer et mettre à l'essai avec soin les méthodes d'imputation rattachées aux nouvelles classes.
- Dans le cas d'une imputation par enregistrement donneur, essayer d'imputer les données dans un enregistrement qui comporte le moins de donneurs possible. Sur le plan opérationnel, cela peut être interprété comme un donneur par section du questionnaire, car il est pratiquement impossible de traiter toutes les variables d'un grand questionnaire simultanément. En outre, selon le nombre de donneurs disponibles, il faut faire en sorte que les mesures d'imputation équivalentes aient

des chances égales d'être sélectionnées afin d'éviter d'augmenter artificiellement la taille de certains groupes de la population.

- Dans le cas des enquêtes importantes, il peut être nécessaire d'échelonner le traitement des variables sur deux ou plusieurs cycles plutôt que dans un seul cycle, ce qui réduit les coûts informatiques. En outre, il est possible qu'un enregistrement contienne un très grand nombre d'erreurs de réponse. L'une ou l'autre de ces conditions peut nuire au strict respect des lignes directrices. Dans certains cas, on aura besoin de plus d'un donneur et on imputera plus que le nombre de variables minimal.
- Au cours de l'élaboration de la méthode d'imputation, remarquer qu'un certain nombre de systèmes généralisés utilisent une variété d'algorithmes pour les données continues ou catégoriques. Généralement, les systèmes sont faciles à utiliser lorsque les vérifications sont précisées et intègrent des algorithmes pour déterminer les champs à imputer. Ils sont bien documentés et conservent les pistes de vérification pour permettre l'évaluation du processus d'imputation. Statistique Canada possède actuellement deux systèmes, soit le Système généralisé de vérification et d'imputation (SGVI/BANFF) (Kovar et coll., 1988; Statistique Canada, 2000a) pour les variables économiques quantitatives et le Système canadien de contrôle et d'imputation du recensement (SCANCIR) (Bankier et coll., 1999) pour les variables qualitatives et quantitatives.
- Marquer les valeurs imputées et désigner clairement les méthodes et les sources d'imputation. Conserver les valeurs imputées et non imputées des champs de l'enregistrement pour les besoins de l'évaluation. Évaluer le degré et les effets de l'imputation. Envisager l'utilisation de techniques pour mesurer efficacement la variance d'échantillonnage dans le cadre de l'imputation et pour mesurer la variance ajoutée introduite par imputation (Lee et coll., 2002). Cette information est requise pour satisfaire aux exigences de la Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie (Statistique Canada, 2000d; voir l'annexe 2).
- Tenir compte du degré et de l'incidence de l'imputation au cours de l'analyse des données. Les méthodes d'imputation utilisées peuvent avoir une forte incidence sur les distributions des données. Par exemple, il est possible qu'il n'y ait pas eu de grands changements au niveau de l'agrégation mais que les valeurs d'un domaine aient systématiquement augmenté, alors que les valeurs d'un autre domaine aient diminué en créant un effet de compensation. En outre, même si le degré d'imputation est faible, les changements dans chacun des enregistrements peuvent avoir un effet considérable, par exemple lorsque les changements sont apportés à de grandes unités ou lorsque de grands changements sont apportés à un petit nombre d'unités. En général, plus le degré et l'incidence de l'imputation sont importants, plus l'analyste doit être judicieux dans l'utilisation des données. Dans de tels cas, les analyses peuvent être trompeuses si les valeurs imputées sont traitées comme des données observées.

- Soulignons que le Bulletin d'imputation produit par la Direction de la méthodologie présente le logiciel et les pratiques de Statistique Canada rattachés à l'imputation, ainsi que les progrès récents dans ce domaine. En outre, le Comité des pratiques d'imputation (CoPI) se réunit régulièrement pour discuter des questions liées à l'imputation et des mises en oeuvre précises d'activités d'imputation. Au cours de l'élaboration d'une stratégie d'imputation, on peut obtenir des commentaires et des suggestions utiles auprès du CoPI.

Références

- Bankier, M., Lachance, M. et Poirier, P. (1999). A generic implementation of the New Imputation Methodology. **Proceedings of the Survey Research Methods Section**, American Statistical Association, 548-553.
- Fellegi, I.P. et Holt, D. (1976). A systematic approach to automatic edit and imputation. **Journal of the American Statistical Association**, 71, 17-35.
- Kalton, G. et Kasprzyk, D. (1986). Le traitement des données d'enquête manquantes. **Techniques d'enquête**, 12, 1-17.
- Kovar, J.G. et Whitridge, P. (1995). Imputation of business survey data. Dans **Business Survey Methods**, B.G. Cox et coll. (éd.), Wiley, New York, 403-423.
- Kovar, J.G., MacMillan, J. et Whitridge, P. (1988). Overview and strategy for the Generalized Edit and Imputation System. Statistique Canada, document de travail de la Direction de la méthodologie n° BSMD 88-007 E/F.
- Lee, H., Rancourt, E. et Särndal, C.-E. (2002). Variance estimation from survey data under single imputation. In **Survey Nonresponse**, R.M. Groves et coll. (éd.), Wiley, New York, 315-328.
- Statistique Canada (2000a). Description des fonctions du système généralisé de vérification et d'imputation. Rapport technique de Statistique Canada.
- Statistique Canada (2000d). Politique visant à informer les utilisateurs de la qualité des données et la méthodologie. **Manuel des politiques**, 2.3. (Reproduite à l'annexe 2).

2.10 Estimation

Portée et objet

L'*estimation* est le processus qui établit une approximation des paramètres de population inconnus au moyen d'une partie seulement de la population, celle comprise dans un échantillon. On effectue alors des inférences à l'égard de ces paramètres inconnus en utilisant les données échantillonnées et le plan de sondage utilisé. Lorsque les paramètres de population sont des fonctions des totaux de population, leurs estimateurs sont généralement des fonctions correspondantes des totaux de population estimés. Parmi les exemples de paramètres, mentionnons les statistiques descriptives simples telles que les totaux, les moyennes, les ratios et les percentiles, ainsi que les statistiques analytiques plus complexes, comme les coefficients de régression.

Règle générale, on calcule des mesures de *précision* pour évaluer la qualité de l'estimation des paramètres de population et pour obtenir des inférences valides. Bien que la qualité des estimations calculées repose en grande partie sur les étapes d'enquête antérieures, le choix d'une méthode d'estimation s'avère également important. Tout particulièrement, on peut utiliser des données auxiliaires afin de rehausser la précision de ces estimations.

Principes

Un des objectifs propres aux enquêtes consiste à estimer un paramètre descriptif de la population au moyen d'un échantillon. L'*erreur d'enquête totale* de l'estimation est l'étendue par laquelle l'estimation diffère de la valeur réelle de la somme rattachée à la population observée (Thompson, 1997). L'erreur dans l'enquête totale peut être exprimée comme la somme de l'*erreur d'échantillonnage* et de l'*erreur non due à l'échantillonnage*. L'erreur d'échantillonnage représente l'erreur qu'on commet en estimant un paramètre d'intérêt en utilisant des données provenant que de l'échantillon. Les erreurs non dues à l'échantillonnage sous-tendent d'autres raisons qui expliquent la présence d'un estimateur imparfait. Ces dernières erreurs comprennent les erreurs de couverture (base imparfaite), les erreurs de mesure et les erreurs dues à la présence de non-réponse.

La méthode d'estimation et le plan d'échantillonnage déterminent les propriétés de l'erreur d'échantillonnage. Le biais d'échantillonnage et la variance d'échantillonnage sont au nombre des critères qui servent à évaluer l'étendue de l'erreur d'échantillonnage. Il faut donc choisir les méthodes d'estimation qui produisent le plus faible biais et la plus faible variance d'échantillonnage. La convergence par rapport à la distribution d'échantillonnage est une autre propriété souhaitable de l'estimation.

L'*estimateur de Horvitz-Thompson*, qui est convergent par rapport à la distribution d'échantillonnage, est l'estimateur le plus naturel à utiliser lorsqu'aucune information auxiliaire n'est disponible à l'étape de l'estimation. Cet estimateur pondère les données en fonction de l'inverse des probabilités d'inclusion des unités échantillonnées. Ce poids

est appelé *poids d'échantillonnage*. On peut considérer le poids d'échantillonnage comme étant le nombre de fois où l'on doit reproduire chaque unité échantillonnée afin de représenter la population complète.

Il est possible d'améliorer les propriétés de l'estimateur de Horvitz-Thompson lorsque de l'information auxiliaire est disponible. Le *calage aux marges* est une procédure qu'on peut appliquer pour incorporer des données auxiliaires. Cette procédure rajuste les poids d'échantillonnage au moyen de multiplicateurs appelés les *facteurs de calage*, lesquels font correspondre les estimations aux totaux connus. Les poids obtenus sont appelés les *poids de calage* ou les *poids d'estimation finaux*. De façon générale, ces poids de calage donneront des estimations convergentes qui admettront une variance d'échantillonnage plus faible que celle obtenue au moyen de l'estimateur Horvitz-Thompson.

En présence de non-réponse, l'échantillon observé est plus petit que l'échantillon initial sélectionné. Pour contrebalancer les effets de la présence de non-réponse, on peut effectuer de l'imputation (voir la section 2.9) ou une *repondération*. La *repondération* consiste à rajuster les poids d'échantillonnage au moyen des facteurs de correction de la non-réponse avant d'appliquer la technique de calage aux marges. Le principe de base du calcul des facteurs de correction de la non-réponse est d'utiliser l'inverse des probabilités des unités à répondre. Toutefois, les probabilités de réponse sont inconnues et doivent être estimées, contrairement aux probabilités d'inclusion qui sont connues. Le meilleur moyen de réduire le biais et la variance de non-réponse est d'établir un modèle de non-réponse approprié en tirant le plus grand profit de l'information auxiliaire disponible.

Lignes directrices

- Une estimation adéquate est conforme au plan d'échantillonnage. Dans cet esprit, il faut intégrer les poids d'échantillonnage au processus d'estimation. Cela signifie que les aspects du plan d'échantillonnage comme la stratification, les grappes et l'information sur les phases et/ou degrés multiples doivent se refléter dans l'estimation des paramètres et des estimateurs de variance correspondants.
- Utilisez des données auxiliaires aussi souvent que possible pour renforcer la fiabilité des estimations. Évaluer l'utilisation des données auxiliaires. Cela peut se faire par exploration à l'aide, par exemple, du Système généralisé d'estimation (SGE) de Statistique Canada, lequel est fondé sur des techniques d'ajustement basées sur la régression.
- Lorsque des données auxiliaires sont disponibles pour les unités d'échantillonnage, de même que les totaux de population correspondants à ces données, envisagez d'utiliser l'estimation de calage afin que les données auxiliaires pondérées correspondent à ces totaux connus. Cela pourra apporter une plus grande précision et une meilleure uniformité des estimations tirées de diverses sources. Essayer de limiter l'étendue des valeurs que peuvent prendre les poids résultant du calage aux marges. Une grande hétérogénéité des poids peut se traduire par une augmentation de la variance des estimations et, par conséquent,

par une réduction de leur précision. Pour réduire l'étendue des valeurs que peuvent prendre les poids de calage, on peut fixer des bornes à l'intérieur desquelles les poids de calage devront se trouver (Huang et Fuller, 1978; Deville et Särndal, 1992). On peut également utiliser ces méthodes qui imposent des bornes sur les poids de calage pour éviter d'obtenir des poids de calage négatifs ou extrêmement grands. Singh et Mohl (1996), Stukel et coll. (1996) et Fuller (2002) discutent en détail de l'utilisation des données auxiliaires.

- Lorsque la classification initiale des unités d'échantillonnage a changé entre le moment de la sélection de l'échantillon et le moment de l'estimation, il faut envisager l'*estimation du domaine* pour que la nouvelle classification se reflète dans les estimations. L'estimation par domaines se définit comme l'estimation de sous-ensembles précis de la population d'intérêt, les *domaines*. Bien souvent, les unités se trouvant dans ces sous-ensembles n'ont pas été identifiées, ou n'auraient pas pu l'être, au moment de l'échantillonnage. L'estimation des unités qui ne sont plus actives au moment de l'estimation ou hors du champ de l'enquête est un exemple d'estimation du domaine. Ces unités se voient attribuer une valeur de zéro dans le processus d'estimation (Hidiroglou et Laniel, 2001).
- Puisque la qualité des facteurs de correction de la non-réponse des poids est déterminée par les hypothèses du modèle employé, il faut valider le modèle choisi en établissant plusieurs diagnostics et s'assurer de ne pas omettre les variables auxiliaires corrélées avec la propension à répondre. Cela offrira une certaine protection contre le biais dû à la non-réponse. Pour obtenir une certaine protection contre l'inadéquation du modèle, il est suggéré de former des classes homogènes d'ajustement de la non-réponse et estimer les probabilités de réponse des unités en établissant les taux de réponse dans ces classes. Pour former ces classes homogènes, utilisez des données auxiliaires corrélées avec la propension à répondre. Eltinge et Yansaneh (1997) abordent certaines des méthodes servant à former des classes homogènes. On peut appliquer la théorie d'échantillonnage à deux phases pour estimer la variance des divers estimateurs qui intègrent les ajustements pour la non-réponse. Par exemple, ces procédures sont décrites dans le chapitre 15 du livre Särndal, Swensson et Wretman (1992). Knowledge Seeker est un progiciel qu'on peut utiliser pour former des classes homogènes au moyen de la méthode exposée dans Kass (1980).
- S'il y a lieu, utilisez l'échantillonnage double pour améliorer l'estimation en intégrant des données auxiliaires. Il s'agit de données qui sont disponibles pour la population ou pour l'échantillon complet (e.g. l'échantillon de première phase dans le cas de l'échantillonnage à deux phases). On peut utiliser l'échantillonnage double pour a) stratifier l'échantillon de deuxième phase, b) améliorer l'estimation en utilisant un estimateur par différence, par ratio ou par régression et c) former un sous-échantillon d'unités non répondantes. Hidiroglou et Särndal (1998) présente une approche très générale de l'échantillonnage à deux phases lorsque des données auxiliaires sont intégrées au processus d'estimation à l'aide de l'estimateur de régression généralisé (GREG). Dans le cas de l'échantillonnage

double, Hidiroglou (2001) présente une théorie générale lorsque des données auxiliaires sont intégrées au processus d'estimation à l'aide d'estimateurs de régression optimale des totaux.

- L'échantillonnage des unités d'intérêt peut être indirect. Cela signifie que l'échantillon d'une base d'intérêt (qui représente la population observée) ne peut être obtenu que par la sélection d'unités faisant partie d'une autre base. Si des liens peuvent être établis entre les unités des deux bases, pensez à produire l'inférence sur la population observée en calculant les poids d'estimation des unités sondées. On peut calculer ces poids en utilisant la Méthode généralisée de partage des poids décrite par Lavallée (2002).
- Il faut garder à l'esprit que dans les enquêtes longitudinales, deux ensembles de poids d'estimation sont généralement utilisés, soit les poids longitudinaux et les poids transversaux. Les poids longitudinaux se rapportent à la population pour laquelle la sélection initiale de l'échantillon longitudinal a été faite. Ces poids sont habituellement ajustés en fonction de l'attrition de l'échantillon au fil du temps. On utilise les poids longitudinaux lorsqu'on procède à l'analyse des données longitudinales. Les poids transversaux, quant à eux, se rapportent à la population établie au cycle correspondant. Ces poids sont normalement utilisés pour produire des estimations ponctuelles ou des différences d'estimations ponctuelles entre deux périodes. En raison des changements apportés à la population au fil du temps, les poids transversaux diffèrent généralement des poids longitudinaux.
- Dans les enquêtes périodiques où il y a un important chevauchement de l'échantillon entre les cycles, envisagez l'utilisation de méthodes d'estimation exploitant la corrélation au fil du temps (Binder et Hidiroglou, 1988; Singh, Kennedy et Wu, 2001). Une de ces méthodes d'estimation est appelée l'*estimation composite*. Fondamentalement, ces méthodes traitent les données des cycles antérieures comme des variables auxiliaires.
- Intégrez autant que possible les besoins et exigences exprimés pour des petits domaines aux étapes d'élaboration du plan d'échantillonnage et de la répartition de l'échantillon (Singh, Gambino et Mantel, 1994). Si cela n'est pas possible à l'étape de l'échantillonnage ou si les domaines ne sont précisés qu'à une étape ultérieure, envisagez des méthodes spéciales d'estimation (estimateurs de petites régions) à l'étape de l'estimation. Ces méthodes « prennent appui » sur les régions (ou des domaines) connexes pour réduire l'erreur quadratique moyenne de l'estimateur résultant (Platek et coll., 1987; Ghosh et Rao, 1994; Rao, 1999).
- Les valeurs aberrantes entraînent souvent des estimations non fiables des variables continues. Les valeurs aberrantes s'expliquent par des valeurs extrêmes mesurées à l'égard de certaines caractéristiques et/ou par des poids très grands rattachés aux unités sous-jacentes. Envisagez d'utiliser des procédures objectives, telles que les estimateurs à l'épreuve des valeurs aberrantes (robustes) (Hidiroglou

et Srinath, 1981; Fuller, 1991; Lee, 1995; Duchesne 1999; Gwet et Lee, 2000; Chambers, Kocic, Smith et Crudas, 2000). Dans le cas des valeurs aberrantes multivariées, on recommande d'utiliser la distance de Mahalanobis et les estimateurs de Stahel-Donoho, adaptés au plan d'échantillonnage de l'enquête (Patak, 1990; Franklin, Thomas et Brodeur, 2000).

- Lorsque cela est possible, utiliser un logiciel d'estimation généralisé plutôt que des systèmes personnalisés. Parmi les logiciels à envisager, mentionnons le SGE (Estevao, Hidirolou et Särndal, 1995), SUDAAN 8.0 (Shah et coll., 1997), PC CARP (Schnell et coll., 1988), WesVar PC (Brick, et coll., 2000), STATA (1997) et SAS 8.0. En utilisant des systèmes généralisés, on peut espérer un moins grand nombre d'erreurs de programmation et une certaine réduction des coûts et du temps que nécessite la mise au point.

Références

- Binder, D.A. et Hidirolou, M.A. (1988). Sampling in time. **Handbook of Statistics**, P.K. Krishnaish et C.R. Rao (éd.), 187-211.
- Brick, J., Morganstein D. et Valliant R. (2000). Analysis of complex sample data using replication. Voir aussi <http://www.westat.com/wesvar/techpapers>.
- Brogan, D. (1998). Software for sample survey data, misuse of standard packages. Dans **Encyclopedia of Biostatistics**, Volume 5, P. Armitage et T. Colton (éd.), Wiley, New York, 4167-4174. Voir aussi <http://www.rti.org/sudaan/homeabout.cfm?aboutfile=whySUDAAN.cfm>.
- Chambers, R.L., Kocic P., Smith P. et Crudas M. (2000). Winsorization for identifying and treating outliers in business surveys. **Proceedings of the Second International Conference on Establishment Surveys**, juin 17-21, 2000, Buffalo, New York, 717-726.
- Cochran, W.G. (1977). **Sampling Techniques**. Wiley, New York.
- Deville, J.-C. et Särndal, C.E. (1992). Calibration estimators in survey sampling. **Journal of the American Statistical Association**, 87, 376-382.
- Duchesne, P. (1999). Estimateurs de calage robustes. **Techniques d'enquête**, 25, 47-60.
- Eltinge, J.L. et Yansaneh, I.S. (1997). Méthodes diagnostiques pour la construction de cellules de correction pour la non-réponse, avec application à la non-réponse aux questions sur le revenu de la U.S. Consumer Expenditure Survey. **Techniques d'enquête**, 23, 37-45.
- Estevao, V., Hidirolou, M.A. et Särndal, C.E. (1995). Methodological principles for a generalized estimation system at Statistics Canada. **Journal of Official**

Statistics, 11, 181-204. Voir aussi <http://www.fas.harvard.edu/~stats/survey-soft/genest.html>.

- Franklin, Sarah F., Thomas, S. et Brodeur, M. (2000). Robust multivariate outlier detection using Mahalanobis' distance and modified Stahel-Donoho estimators. **Proceedings of the Second International Conference on Establishment Surveys**, juin 17-21, 2000, Buffalo, New York, 697-706.
- Fuller, W.A. (1991). Simple estimators of the mean of skewed populations. **Statistica Sinica**, 1, 137-158.
- Fuller, W.A. (2002). Estimation par régression appliquée à l'échantillonnage. **Techniques d'enquête**, 28, 5-25.
- Fuller, W.A. et Rao, J.N.K. (2001). Un estimateur composite de régression qui s'applique à l'Enquête sur la population active du Canada. **Techniques d'enquête**, 27, 49-56.
- Gambino, J., Kennedy, B. et Singh, M.P. (2001). Estimation composite par régression pour l'Enquête sur la population active du Canada: Évaluation et application. **Techniques d'enquête**, 27, 69-79.
- Ghosh, M. et Rao, J.N.K. (1994). Small area estimation: an appraisal. **Statistical Science**, 9, 55-93.
- Gwet, J.-P. et Lee, H., (2000). An evaluation of outlier-resistant procedures in establishment surveys. **Proceedings of the Second International Conference on Establishment Surveys**, juin 17-21, 2000, Buffalo, New York, 707-716.
- Hidiroglou, M.A. (2001). L'échantillonnage double. **Techniques d'enquête**, 27, 157-169.
- Hidiroglou, M.A. et Laniel, N. (2001). Sampling and estimation issues for annual and sub-annual Canadian business surveys. **International Statistical Review**, 69, 487-504.
- Hidiroglou, M.A. et Särndal, C.E. (1998). Emploi des données auxiliaires dans l'échantillonnage à deux phases. **Techniques d'enquête**, 24, 11-20.
- Hidiroglou, M.A. et Srinath, K.P. (1981). Some estimators of population total containing large units. **Journal of the American Statistical Association**, 47, 663-685.
- Holt, D. et Smith, T.M.F. (1979). Post-stratification. **Journal of the Royal Statistical Society**, A 142, 33-46.
- Huang, E. et Fuller, W.A. (1978). Nonnegative regression estimation for sample survey data. **Proceedings of the Social Statistics Section**, American Statistical Association, 300-303.

- Hulliger, B. (1995). Estimateur Horvitz-Thompson à l'épreuve des valeurs aberrantes. **Techniques d'enquête**, 21, 89-97.
- Kass, G.V. (1980). An exploratory technique for investigating large quantities of categorical data. **Applied Statistics**, 29, 119-127.
- Lavallée, P. (2001). La méthode généralisée du partage des poids et le calage sur marges. Dans **Enquêtes, modèles et applications**, J.-J. Droesbeke and L. Lebart (éd.), Paris: Dunod, 396-403.
- Lavallée, P. (2002). **Le sondage indirect**. Éditions de l'Université de Bruxelles.
- Lavallée, P. et Caron, P. (2001). Estimation par la méthode généralisée du partage des poids : Le cas du couplage d'enregistrements. **Techniques d'enquête**, 27, 171-187.
- Lee, H. (1995). Outliers in business surveys. Dans **Business Survey Methods**, B.G. Cox et coll. (éd.), Wiley, New York, 503-526.
- Lemaître, G. et Dufour, J. (1987). Une méthode intégrée de pondération des personnes et des familles. **Techniques d'enquête**, 13, 211-220.
- Lohr, S. (1999). **Sampling: Design and Analysis**. Duxbury Press.
- Patak, Z. (1990). Robust principal component analysis via projection pursuit. Master's thesis, University of British Columbia, Canada.
- Platek, R., Rao, J.N.K., Särndal, C.E. et Singh, M.P. (éd.) (1987). **Small Area Statistics**. Wiley, New York.
- Rao, J.N.K. (1996). On the estimation with imputed survey data. **Journal of the American Statistical Association**, 91, 499-506.
- Rao, J.N.K. (1999). Quelques progrès récents concernant l'estimation régionale fondée sur un modèle. **Techniques d'enquête**, 25, 199-212.
- Särndal, C.E., Swensson, B. et Wretman, J.H. (1992). **Model Assisted Survey Sampling**. Springer-Verlag, New York.
- Schnell, D., Kennedy, W.J., Sullivan, G, Park, H.J. et Fuller, W.A. (1988). Logiciel d'ordinateur personnel pour l'estimation de la variance dans des enquêtes complexes. **Techniques d'enquête**, 14, 63-73.
- Shah, B.V., Barnwell, B.G. et Bieler, G.S. (1997). **SUDAAN User's Manual Release 7.5**. Research Triangle Institute, Caroline du Nord. Voir aussi <http://www.rti.org/sudaan/home.cfm>.

- Singh, A.C. et Mohl, C. (1996). Comprendre les estimateurs de calage dans les enquêtes par échantillonnage. **Techniques d'enquête**, 22, 107-116.
- Singh, A.C., Kennedy, B. et Wu, S. (2001). Estimation composite par régression pour l'Enquête sur la population active du Canada avec plan de sondage à renouvellement de panel. **Techniques d'enquête**, 23, 35-48.
- Singh, M.P., Gambino, J. et Mantel, H. (1994). Les petites régions: problèmes et solutions. **Techniques d'enquête**, 20, 3-23.
- Singh, M.P., Hidioglou, M.A., Gambino, J. et Kovacevic, M. (2001). Estimation methods and related systems at Statistics Canada. **International Statistical Review**, 69, 461-486.
- STATA Corporation (1997). **STATA User's Guide**. STATA Press, College Station, TX. Voir <http://www.fas.harvard.edu/~stats/survey-soft/iass.html#stata>.
- Stukel, D., Hidioglou, M.A., et Särndal, C.-E. (1996). Estimation de la variance des estimateurs de calage: comparaison des méthodes du jackknife et de la linéarisation de Taylor. **Techniques d'enquête**, 22, 117-126.
- Thompson, M.E. (1997). **Theory of Sample Surveys**. Chapman and Hall.
- Valliant, R., Dorfman, A.H., et Royall, R.M. (2000). **Finite Population Sampling and Inference**. Wiley, New York.
- Yung, W. et Rao, J.N.K. (1996). Linéarisation des estimateurs de variance jackknife dans un échantillonnage stratifié à degrés multiples. **Techniques d'enquête**, 22, 23-31.

2.11 Désaisonnalisation et estimation de la tendance-cycle

Portée et objet

La *désaisonnalisation* consiste à estimer les facteurs saisonniers et à les appliquer à une série chronologique afin d'en supprimer les *variations saisonnières*. Ces variations représentent l'effet composite des facteurs climatiques et institutionnels se reproduisant avec une certaine régularité durant l'année. Une série désaisonnalisée comprend une composante tendance-cycle et une composante irrégulière. La *tendance* représente le mouvement sous-jacent qui s'étale sur de nombreuses années. Le *cycle*, couramment appelé le cycle économique, est une oscillation quasi-périodique qui dure de trois à quatre ans. La *composante irrégulière* représente les variations aléatoires qui sont des mouvements imprévisibles se rapportant à toutes sortes d'événements.

Bon nombre de séries publiées par les organismes de statistique sont désaisonnalisées; ainsi, les mouvements de la tendance-cycle sous-jacente sont plus évidents, ce qui facilite l'analyse des données. Comme elles sont constituées non seulement de la tendance-cycle, mais également de la composante irrégulière, les séries désaisonnalisées ne donnent qu'un aperçu approximatif des mouvements de la tendance-cycle sous-jacente. Il est souvent souhaitable de lisser davantage la série désaisonnalisée afin de supprimer la composante irrégulière et de publier des *estimations de la tendance-cycle* pour compléter la série désaisonnalisée.

Ce chapitre est une copie conforme des lignes directrices pour la désaisonnalisation et estimation de la tendance-cycle de Statistique Canada (2000b).

Principes

Il convient de désaisonnaliser la série chronologique uniquement lorsqu'elle est manifestement influencée par des facteurs saisonniers, et que son caractère saisonnier peut être décelé avec certitude. La saisonnalité est identifiable lorsque le schéma saisonnier n'est pas obscurci par un degré élevé de fluctuations irrégulières et qu'il est facile à circonscrire (Lothian et Morry, 1978).

Une méthode de désaisonnalisation efficace produit une série dénuée de saisonnalité qui est beaucoup plus lisse que la série originale.

Les révisions apportées aux estimations désaisonnalisées devraient être minimales, lorsque d'autres points de données sont ajoutés à la fin de la série. Statistique Canada a adopté les méthodes de désaisonnalisation X11ARMMI (Dagum, 1980) et X11ARMMI/88 (Dagum, 1988) précisément dans le but de réduire l'importance des révisions (Dagum, 1975; Kuiper, 1976; Dagum, 1982).

La fréquence des révisions doit être réduite au minimum. Bien qu'il y ait une révision chaque fois qu'une donnée est ajoutée, les révisions doivent être effectuées uniquement si

elles améliorent les estimations, c'est-à-dire si les estimations révisées sont beaucoup plus près des estimations définitives.

Chaque fois que des données désaisonnalisées se rapportant à la même activité économique sont publiées, les méthodes de désaisonnalisation employées par les Divisions doivent être coordonnées et tout doit être mis en oeuvre afin que les séries connexes soient traitées avec cohérence.

Lorsqu'on produit des estimations de la tendance-cycle afin de compléter la série désaisonnée, il ne faut pas oublier qu'elles visent à rendre compte du stade du cycle économique et, à ce titre, doivent refléter avec exactitude la direction du mouvement, l'amplitude du cycle et, tout particulièrement, les dates des points de retournement.

Les estimations de la tendance-cycle doivent être cohérentes avec les estimations des données désaisonnalisées publiées. Si ces dernières estimations sont immobilisées dans la base de données après trois mois, l'estimateur de la tendance-cycle doit être appliqué aux estimations désaisonnalisées figurant dans la base et dans la publication.

Lignes directrices

- Analysez en profondeur une série que vous désaisonnez pour la première fois, et périodiquement par après à des intervalles de quelques années, pour vérifier que la saisonnalité est identifiable. Ne publiez la série sous sa forme désaisonnée que si la saisonnalité est identifiable.
- Lors de la désaisonnalisation, utilisez les extrapolations ARMMI dans le calcul des facteurs saisonniers pour réduire l'ampleur des révisions. Laissez le programme X11ARMMI choisir automatiquement un modèle ARMMI dans la mesure du possible. Si aucun des modèles incorporés n'est choisi, on recommande que l'utilisateur fournisse un modèle ARMMI.
- Désaisonnez les plus récentes observations en utilisant le *facteur saisonnier actualisé* (Dagum, 1987), obtenu à partir de tous les points de données disponibles. Il est possible de déroger à la présente ligne directrice lorsque les observations les plus récentes ont fait l'objet de révisions historiques importantes. En pareil cas, utilisez les *facteurs saisonniers (de prévision) de l'année suivante* qui peuvent s'avérer plus appropriés (Morry, 1992). Ces facteurs saisonniers reposent sur des données qui se terminent à la fin de l'année précédente.
- Si un facteur saisonnier actualisé est utilisé, il n'est pas nécessaire de réviser les estimations désaisonnées qui remontent à plus d'une période antérieure, lorsque l'observation suivante est ajoutée à la série. Cependant, dans le cas d'observations provisoires, réviser les facteurs saisonniers chaque fois que les observations originales correspondantes sont modifiées substantiellement. Réviser sur une base annuelle les observations désaisonnées des trois dernières années lorsque les observations du premier mois (trimestre de l'année

suivante) deviennent disponibles (Dagum, 1987). Par contre, révisez les quatre dernières années lorsque des facteurs saisonniers prévus ont été utilisés.

- Pour corriger une série des effets jours ouvrables, utilisez les coefficients d'ajustement estimés automatiquement par le programme X11ARMMI. Pendant l'année en cours, gardez ces coefficients constants en les utilisant en tant que poids journaliers *a priori*. Les coefficients d'ajustement seront modifiés lors de la prochaine révision annuelle. Un spécialiste qui connaît très bien ses séries peut cependant fournir au programme des poids journaliers *a priori*.
- Corrigez une série des effets de Pâques en utilisant les facteurs qui sont calculés automatiquement par le programme X11ARMMI.
- Lors de la désaisonnalisation d'une série synthétique qui résulte de l'agrégation de plusieurs séries composantes, ne désaisonnalisez que les composantes qui contiennent une saisonnalité identifiable; les autres ne sont pas désaisonnalisées. La série synthétique est désaisonnalisée soit directement, soit indirectement. Selon la méthode indirecte, combinez les composantes désaisonnalisées et les composantes non désaisonnalisées pour obtenir l'agrégat désaisonnalisé. Selon la méthode directe, désaisonnalisez l'agrégat et rétablissez l'additivité (si nécessaire) en calant les composantes sur le total, sans modifier si possible les composantes non désaisonnalisées. Afin de savoir s'il vaut mieux désaisonnaliser une série synthétique de façon directe ou indirecte, prenez en considération les facteurs suivants : l'agrégat ne doit pas contenir de saisonnalité résiduelle et il doit être relativement lisse (Lothian et Morry, 1977). Choisissez l'approche directe avec calage seulement si le calage ne difforme pas indûment les mouvements d'un mois à l'autre (d'un trimestre à l'autre) des séries composantes désaisonnalisées.
- Coordonnez les méthodes de désaisonnalisation employées par différentes divisions chaque fois que des données désaisonnalisées se rapportant à la même activité économique sont publiées. Il importe notamment de faire preuve de cohérence, dans la mesure du possible, dans le choix entre les méthodes de désaisonnalisation directe et indirecte des séries synthétiques; en outre, les Divisions doivent traiter uniformément les données extrêmes des séries chronologiques qui sont dues à des événements tels que les fermetures d'usine, les grèves, les catastrophes naturelles, et autres.
- Utilisez les moyennes mobiles de Henderson, disponibles dans le programme X11ARMMI, pour produire l'estimation de la composante tendance-cycle. Afin que la composante tendance-cycle se situe dans le nuage de points des observations désaisonnalisées, appliquez les moyennes mobiles de Henderson aux séries désaisonnalisées déjà publiées.
- Avant d'appliquer l'estimateur de la tendance-cycle, extrapolez la série désaisonnalisée sur un horizon d'un an à l'aide d'un modèle ARMMI ajusté à la série désaisonnalisée.

- Appliquez la moyenne mobile de Henderson choisie à la série désaisonnalisée extrapolée et corrigée de ses valeurs extrêmes.
- Laissez le programme X11ARMMI choisir automatiquement la moyenne mobile de Henderson : ce choix est fondé sur la valeur d'un indicateur dit "ratio I/C" (I pour Irrégulier, C pour Tendence-cycle), qui mesure l'importance relative de la composante irrégulière dans la série désaisonnalisée (Shiskin et coll., 1967).
- Informez les utilisateurs du fait que les dernières estimations de la tendance-cycle (notamment la toute dernière) sont sujettes à d'importantes révisions, voire à un revirement du mouvement après l'ajout d'une nouvelle observation (Dagum et Laniel, 1987). On pourrait indiquer, par exemple, le haut degré d'incertitude des estimations en fin de série par une ligne pointillée sur le graphique des tendances ou par un avertissement écrit aux utilisateurs.

Références

- Dagum, E.B. (1975). Seasonal factor forecasts from ARIMA models. **Proceedings of the International Institute of Statistics**, 40^e séance, communications présentées, 3, Varsovie, 206-219.
- Dagum, E.B. (1980). La méthode de désaisonnalisation X-11-ARMMI. Publication n° 12-564F au catalogue, Statistique Canada.
- Dagum, E.B. (1982). The effects of asymmetric filters on seasonal factor revisions. **Journal of the American Statistical Association**, 77, 732-738.
- Dagum, E.B. (1987). Problèmes courants sur la désaisonnalisation. **Techniques d'enquête**, 13, 71-82.
- Dagum, E.B. et Laniel, N. (1987). Revisions of trend-cycle estimators of moving average seasonal adjustment methods. **Journal of Business and Economic Statistics**, 5, 177-189.
- Dagum, E.B. (1988). The X11ARIMA/88 seasonal adjustment method - foundations and user's manual. Time Series Research and Analysis Division, Rapport technique de Statistique Canada.
- Kuiper, J. (1976). A survey and comparative analysis of various methods of seasonal adjustment. **Proceedings of the NBER/Bureau of the Census Conference on Seasonal Analysis of Economic Time Series**, (Arnold Zellner, éditeur), Washington, D.C. 59-76.
- Lothian, J. et Morry, M. (1977). The problem of aggregation: direct or indirect seasonal adjustment. Document de recherche n° 77-08-001E, Seasonal Adjustment and Time Series Staff, Statistique Canada.

- Lothian, J. et Morry, M. (1978). A Test for the Presence of Identifiable Seasonality when using the X-11 Program. Document de recherche n° 78-10-002E, Seasonal Adjustment and Time Series Staff, Statistique Canada.
- Morry, M. (1992). Comparison of revisions in seasonally adjusted external trade series obtained through the use of concurrent versus forecast factors. Statistique Canada, document de travail n°, TSRA-92-004E.
- Shiskin, J., Young, A.H. et Musgrave, J.C. (1967). The X-11 variant of the census method II seasonal adjustment. Technical Paper No. 15, Bureau of the Census, U.S. Department of Commerce.
- Statistique Canada (2000b). Lignes directrices pour la désaisonnalisation et estimation de la tendance-cycle. Document publié par Comité des méthodes et des normes, mars, 2000.

2.12 Évaluation de la qualité des données

Portée et objet

L'évaluation de la qualité des données désigne le processus visant à déterminer si le produit fini satisfait aux objectifs initiaux de l'activité statistique, notamment en ce qui a trait à l'exactitude, à l'actualité ou à la fiabilité de l'information. Elle permet aux utilisateurs de mieux étayer leur interprétation des résultats d'enquête, et à l'organisme statistique d'améliorer les enquêtes.

On distingue deux grands types d'évaluation de la qualité des données :

La *certification* ou la *validation* désigne le processus consistant à analyser les données avant leur diffusion officielle afin d'éviter les erreurs grossières ou de déceler les données de qualité médiocre. Ce processus coïncide souvent avec une analyse interprétative des données. Comme le processus doit généralement être réalisé dans des délais serrés, seules les méthodes pouvant être utilisées rapidement peuvent être employées.

Les *études sur les sources d'erreurs* renferment généralement des renseignements quantitatifs sur des sources précises des erreurs qui se sont glissées dans les données. Si l'actualité demeure importante, il n'est pas rare que les résultats de ces études soient disponibles uniquement après la diffusion officielle des données.

Principes

Il importe que les utilisateurs puissent déterminer dans quelle mesure les erreurs présentes dans les données sont une entrave à leur usage; cependant, rares sont ceux qui sont capables d'évaluer de façon indépendante l'exactitude des données produites par un organisme statistique. Il appartient donc à ce dernier d'effectuer les évaluations nécessaires de la qualité des données et de communiquer rapidement aux utilisateurs les résultats de ces évaluations sous une forme pratique.

Les évaluations de la qualité des données sont également utiles à l'organisme statistique. Dans la mesure où les erreurs peuvent être associées à certaines étapes du processus d'enquête, on peut se servir de ces évaluations pour améliorer la qualité de la prochaine occasion de l'enquête et d'autres enquêtes similaires. Il s'agit ici d'examiner, par exemple, l'efficacité des plans d'enquête, l'importance de la non-réponse, ou les pratiques d'imputation douteuses.

L'actualité des résultats de l'évaluation de la qualité des données est tout aussi importante que l'actualité des données. Idéalement, les résultats de l'évaluation de la qualité sont suffisamment valables et actuels pour améliorer les données qui sont diffusées; par exemple, on peut utiliser les résultats d'une évaluation de la couverture pour compenser pour les différences entre la base de sondage et la population-cible. Lorsque cela s'avère impossible, les résultats de l'évaluation doivent à tout le moins être suffisamment actuels

pour aider les utilisateurs à analyser les données et le personnel à concevoir la prochaine occasion de l'enquête.

Lignes directrices

- Déterminer l'ampleur de l'évaluation de la qualité des données d'un programme ou d'un produit statistique. Les facteurs à considérer incluent les utilisations et les utilisateurs des données; le risque d'erreur et son impact sur l'utilisation des données; la variation de la qualité dans le temps; le coût de l'évaluation par rapport au coût total du programme statistique; les possibilités d'amélioration de la qualité et d'accroissement de l'efficacité et de la productivité; l'utilité des mesures pour les utilisateurs et la facilité avec laquelle elles peuvent être interprétées; le fait que l'enquête sera répétée ou non.
- Concevoir les évaluations de la qualité des données effectuées à Statistique Canada de manière à respecter les exigences obligatoires et minimales de la Politique visant à informer les utilisateurs de la qualité des données et de la méthodologie (Statistique Canada, 2000d; voir l'annexe 2), Dans le cas des données des recensements et des enquêtes, les exigences minimales englobent une mesure ou une évaluation des erreurs de couverture, un taux de réponse ou d'imputation et (s'il s'agit d'une enquête-échantillon) des mesures de l'erreur d'échantillonnage relativement à des caractéristiques clés.
- Fournir une évaluation de la qualité reposant sur un avis d'expert ou une analyse subjective lorsque les évaluations de la qualité des données ne peuvent pas engendrer des mesures quantitatives en raison de la nature du produit statistique, de l'utilisateur, de l'actualité, du coût ou de la faisabilité technique.
- Incrire la planification des évaluations de la qualité des données dans le plan d'enquête puisque l'information que requièrent de telles évaluations doit souvent être recueillie durant le processus d'enquête. Les rapports sur la qualité des données devraient faire partie du calendrier de diffusion de l'enquête.
- Dans le cas d'enquêtes ou d'activités statistiques répétées, il n'est peut-être pas nécessaire ou possible d'effectuer en permanence des évaluations détaillées de la qualité. Cependant, réaliser périodiquement de telles études pour déterminer si l'activité atteint toujours ses objectifs et non pas seulement lorsqu'un problème survient.
- Impliquer les utilisateurs des résultats, qu'ils soient rattachés à l'organisme statistique ou non, dans l'établissement des objectifs du programme d'évaluation de la qualité des données. Lorsque les circonstances le permettent, les impliquer également dans le processus d'évaluation.
- Parmi les méthodes de certification, considérer :
 - les vérifications de la cohérence par rapport aux sources de données

externes, par exemple d'autres enquêtes, des occasions antérieures de l'enquête, ou des données administratives;

- les vérifications de la cohérence interne, par exemple le calcul des ratios qui sont censés se situer à l'intérieur de certaines limites (proportions hommes-femmes, valeur moyenne des biens, etc.);
- les analyses individuelles des plus importantes unités pour ce qui est de leurs contributions aux estimations globales (généralement appliquées aux enquêtes-entreprises);
- le calcul des indicateurs de la qualité des données, par exemple les taux de non-réponse, les taux d'imputation et les coefficients de variation;
- les séances de compte rendu avec le personnel affecté à la collecte et au traitement des données;
- les vérifications de la vraisemblance effectuées par des spécialistes bien informés, ce qui comprend les analyses externes pré-diffusion présentées sous forme de rapports sur les travaux en cours.

■ Évaluer les sources d'erreurs suivantes :

- Les *erreurs de couverture* englobent les éléments qui ont été omis de la base de sondage utilisée dans l'enquête, qui y ont été inclus par erreur et qui y sont répétés. Comme elles influent sur toutes les estimations produites par l'enquête, elles constituent un des plus importants types d'erreur. Les erreurs de couverture peuvent se traduire par un biais positif ou négatif dans les données, et l'impact peut varier selon le sous-groupe de l'univers de l'enquête. On doit aussi s'intéresser aux erreurs de classification, notamment industrielles et géographiques. Par exemple, des limites mal définies ou un codage erroné peuvent se traduire par une omission de parties du territoire.
- Les *erreurs dues à la non-réponse* sont commises lorsqu'on n'obtient pas de réponse à une question ou à l'ensemble des questions de l'enquête. La non-réponse entraîne une augmentation de la variance, en raison de la diminution de la taille réelle de l'échantillon et du recours à l'imputation, et produit un biais si les non-répondants ont des caractéristiques d'intérêt différentes des répondants. Par ailleurs, en traitant les données imputées comme des données observées lors du calcul de l'erreur d'échantillonnage, on risque de sous-estimer celle-ci de façon importante.
- Les *erreurs de mesure* se produisent lorsque la réponse fournie diffère de la vraie valeur; elles peuvent être attribuables au répondant, à l'intervieweur, au questionnaire, à la méthode de collecte ou au système de conservation des dossiers du répondant. De telles erreurs peuvent avoir un caractère aléatoire, ou peuvent causer un biais systématique dans les résultats si ils ne sont pas aléatoires.
- Les *erreurs de traitement* sont commises dans les étapes subséquentes du processus, soit au moment de la vérification des données, du codage, de la saisie, de l'imputation et de la mise en tableaux. À l'instar des erreurs de mesure, les erreurs de traitement peuvent se traduire par une variance et un biais. Il faut aussi étudier l'impact éventuel de ruptures dans les processus

d'enquête : formation inégale des enquêteurs, taux de roulement de personnel anormalement élevé, changements de procédures durant le déroulement des opérations, etc.

- Les *erreurs d'échantillonnage* se produisent lorsque les résultats de l'enquête ont été obtenus à partir d'un échantillon, par opposition à l'ensemble de la population. En pratique, ces erreurs peuvent également englober les *erreurs d'estimation*, qui peuvent être attribuables à l'utilisation d'estimateurs qui créent, délibérément ou non, un biais (p. ex. des estimateurs pour de petites régions).

Une bonne discussion du sujet peut être trouvée dans Lessler et Kalsbeek (1992) et dans Lyberg et coll. (1997).

Références

Biemer, P., Groves, R.M., Mathiowetz, N.A., Lyberg, L. et Sudman, S. (éd.) (1991). **Measurement Errors in Surveys**. Wiley, New York.

Fuller, W. (1987). **Measurement Error Models**. Wiley, New York.

Gosselin, J.-F., Chinnappa, B.N., Ghangurde, P.D. et Tourigny, J. (1978). Couverture. Chapitre 2 dans **Répertoire de méthodes d'évaluation des erreurs dans les recensements et les enquêtes**, n° 13-564F au catalogue, 7-10, Statistique Canada.

Lessler, J.T. et Kalsbeek, W.D. (1992). **Nonsampling Errors in Surveys**. Wiley, New York.

Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N. et Trewin, D. (éd.) (1997). **Survey Measurement and Process Quality**. Wiley, New York.

Statistique Canada (1996b). **Symposium 1996 : Erreurs non dues à l'échantillonnage**.

Statistique Canada (2000d). Politique visant à informer les utilisateurs de la qualité des données et la méthodologie. **Manuel des politiques**, 2.3. (Reproduite à l'annexe 2).

Statistique Canada (2001e). **Symposium 2001 : La qualité des données d'un organisme statistique : une perspective méthodologique**.

Statistique Canada (2002c). Le Cadre d'assurance de la qualité de Statistique Canada - 2002. N° 12-586-XIF au catalogue, Statistique Canada.

Statistique Canada (2003b). Lignes directrices sur l'utilisation de critères d'exactitude des données pour la diffusion de l'information statistique. Document publié par le Comité des méthodes et des normes, juin 2003.

2.13 Contrôle de la divulgation

Portée et objet

En vertu de la loi, Statistique Canada est tenu de protéger la *confidentialité* des renseignements fournis par les répondants. Le *contrôle de la divulgation* désigne les mesures visant à protéger les données de l'organisme, dans le respect des exigences en matière de confidentialité. En général, les activités de contrôle de la divulgation ont une incidence réductrice sur la qualité des données en cela qu'elles peuvent se traduire par la suppression ou la modification d'un niveau de détail. L'objectif consiste à s'assurer que les dispositions régissant la protection de la confidentialité sont respectées tout en préservant le plus possible l'utilité des données produites. Le programme vigilant de contrôle de la divulgation et de protection de la confidentialité de Statistique Canada a grandement contribué à la qualité des données de l'organisme; en effet, les taux de réponse élevés dans les enquêtes du Bureau et la confiance que le public place dans l'organisme en sont tributaires dans une large mesure.

Principes

Les principes qui sous-tendent les activités de contrôle de la divulgation sont presque exclusivement régis par les dispositions de la Loi sur la statistique (1970, S.R.C. 1985, c. S19), plus précisément le paragraphe 17(1) :

aucune personne qui a été assermentée en vertu de l'article 6 ne peut révéler ni sciemment faire révéler, par quelque moyen que ce soit, des renseignements obtenus en vertu de la présente loi de telle manière qu'il soit possible, grâce à ces révélations, de rattacher à un particulier, à une entreprise ou à une organisation identifiables les détails obtenus dans un relevé qui les concerne exclusivement.

Cependant, le paragraphe 17(2) prévoit que certains types de renseignements confidentiels peuvent être diffusés à la discrétion du statisticien en chef et en vertu d'une ordonnance. La diffusion de listes d'entreprises avec adresse et classification industrielle ou la communication de renseignements sur un répondant qui a donné son consentement écrit au préalable constituent les formes les plus courantes de ce type de divulgation. La diffusion d'information qui s'appuie sur le pouvoir discrétionnaire du statisticien en chef est régie par la Politique relative à la révélation discrétionnaire (Statistique Canada, 1993a) et, dans certains cas, par les Lignes directrices relatives à la diffusion de microdonnées non filtrées en vertu d'accords de partage des données prévus par l'article 12 ou en vertu de dispositions de diffusion discrétionnaire des renseignements.

Les dispositions de la Loi sur la statistique en matière de confidentialité sont extrêmement rigoureuses. Par conséquent, leur application dans des cas bien précis représente une tâche ardue quoique extrêmement importante. L'objectif premier consiste à s'assurer qu'aucun résultat personnel identifiable ne puisse être inféré dans une fourchette restreinte. De plus, il est nécessaire de protéger l'information, peu importe si le

sujet est susceptible d'être considéré confidentiel par les répondants; donc, les caractéristiques démographiques de base doivent être protégées au même titre que le revenu. Il est important de souligner que la loi ne précise pas pendant combien de temps l'information est à l'abri de la divulgation. De même, la façon dont le public perçoit la vigilance avec laquelle nous protégeons la confidentialité des statistiques est à tout le moins aussi importante que les mesures réelles que nous prenons pour empêcher la divulgation des données des répondants.

Lignes directrices

- Distinguez les *données tabulaires* des *microdonnées*. Dans le cas des données tabulaires, les données sont diffusées sous forme de tableaux statistiques comportant souvent de nombreuses dimensions, alors que pour les microdonnées, des enregistrements anonymisés sont établis pour les particuliers. Les données tabulaires peuvent être classées en *tableaux de fréquence* ou en *tableaux de données quantitatives*. Les tableaux de fréquence ne renferment que les comptes (ou les comptes estimés) du nombre d'unités correspondant au nombre d'individus classés dans chaque cellule du tableau, alors que les tableaux de données quantitatives font état des valeurs numériques (qui ne sont généralement pas négatives), par exemple la moyenne ou le total des valeurs monétaires ou le nombre d'employés dans chacune des cellules. Les mesures qui assurent la protection de la confidentialité de ces divers produits sont forcément très différentes.
- Ne diffusez pas de tableau de données quantitatives si ce dernier fournit des valeurs rattachées à des cellules jugées confidentielles. Les critères de confidentialité sont habituellement fondés sur des règles simples qui, selon l'opinion générale, protègent contre la divulgation des caractéristiques d'un répondant.
- Déterminez le degré de confidentialité de chaque cellule. Habituellement, on emploie deux critères. Le premier est le nombre de répondants de la cellule et l'autre est fondé sur les mesures de la concentration ou de la prédominance de la distribution des valeurs des répondants dans la cellule. Pour illustrer le premier critère, disons tout simplement que le nombre de répondants d'une cellule doit excéder une valeur minimale. Dans le cadre de nombreuses enquêtes, les tableaux comportant des cellules rattachées à seulement trois répondants peuvent être diffusés. Le nombre de répondants ne peut pas être inférieur à trois, car un des répondants pourrait calculer la valeur correspondant à l'autre répondant au moyen d'une simple soustraction.
- Un grand nombre de règles de suppression des cellules sont fondées sur des mesures de concentration. Déterminez celle qu'il faut utiliser. Les règles les plus faciles à appliquer sont celles qui sont fondées sur des combinaisons linéaires de statistiques d'ordre. Une de ces règles communes est appelée la *règle (n,k)*. Dans l'application de cette règle, une cellule est confidentielle si le plus grand nombre n

de répondants qu'elle comporte représente au moins k % de sa valeur totale. Souvent, on contrôle plus qu'une valeur n , disons $n=1$ ou 2 . Dans certains cas, on utilise différentes valeurs de k selon le nombre de répondants dans la cellule, mais cela n'est pas recommandé, car l'ajout d'un nouveau répondant offrant une contribution négligeable peut transformer une cellule confidentielle en cellule non confidentielle. Intuitivement, cela n'est pas raisonnable en raison de la discontinuité des règles.

- La *règle p-pourcentage* est également fondée sur une mesure de concentration (Subcommittee on Disclosure Limitation Methodology, 1994). Elle sert à assurer qu'une coalition d'unités, généralement l'unité ayant la deuxième valeur d'importance, ne puisse estimer trop précisément la valeur de la plus importante unité. Un exemple de cette règle, où $p = 15$, serait de déclarer une cellule confidentielle si la somme des valeurs de la troisième valeur en importance et de toutes les valeurs plus faibles des répondants représentait moins de 15 % de la valeur la plus importante du répondant. L'extension de la règle *p-pourcentage* est la *règle pq*, où la valeur q ($p < q < 100$) représente l'estimation, par l'organisme, de la capacité des répondants à déterminer avec exactitude les valeurs des autres unités. Une règle *pq* où $p = 15$ et $q = 60$ équivaut à une règle *p-pourcentage* où $p = 25$.
- Déterminez si les cellules à fréquence nulle posent un problème. Les cellules à fréquence nulle peuvent révéler des renseignements confidentiels dans les tableaux de données quantitatives.
- Supprimez les cellules confidentielles du tableau. C'est ce qu'on appelle la *suppression de cellules*. Toutefois, souvent, il ne suffit pas de supprimer uniquement les cellules confidentielles lorsque la distribution marginale est également diffusée, car il est parfois possible de calculer la valeur exacte des cellules supprimées en résolvant un système d'équations linéaires. Même si cela n'est pas possible, on peut calculer une fourchette de valeurs correspondant à la cellule supprimée en utilisant des méthodes de programmation linéaire, et cette fourchette peut être jugée trop restreinte pour protéger suffisamment la valeur supprimée. Par conséquent, on doit trouver des cellules complémentaires à supprimer afin de protéger la cellule confidentielle. Le problème de recherche des cellules complémentaires est davantage compliqué par la présence possible de hiérarchies dans les variables de classification des tableaux (p. ex., les différents niveaux de codage industriel) et les résultats des ensembles de tableaux connexes. Il existe des logiciels de pointe pour déterminer les cellules complémentaires, bien que tous ces logiciels ne permettent pas de résoudre efficacement les problèmes de hiérarchies et de tableaux connexes.
- On peut recourir à d'autres méthodes que la suppression de cellules. On peut notamment modifier les définitions des lignes et des colonnes en procédant au *regroupement* des catégories ou au *regroupement des valeurs extrêmes supérieures* des catégories, de manière qu'il n'y ait pas de cellules confidentielles

(ou qu'il y en ait moins). Parmi les autres méthodes qui peuvent être employées, mentionnons la perturbation des données au terme de l'ajout de bruit aux microdonnées ou aux données tabulaires, comme l'arrondissement. On peut recourir à toute autre méthode qui assure la sécurité de la base de microdonnées sous-jacente afin de protéger les données tabulaires; tous les tableaux seraient ensuite produits à partir de la base de microdonnées « sûre ».

- L'arrondissement des valeurs des cellules peut prendre diverses formes. Il n'est pas rare que l'arrondissement classique ou déterministe n'ajoute pas suffisamment de bruit pour protéger adéquatement la confidentialité. On peut également recourir à l'arrondissement aléatoire.
- Dans les tableaux de fréquence, les cellules à faible fréquence peuvent poser problème. Les individus de ces cellules peuvent facilement être identifiés, de sorte qu'on constate que les autres membres de la population se trouvent dans une autre cellule. Il ne fait pas de doute que si une seule cellule d'une ligne ou d'une colonne n'est pas nulle et qu'on connaît la composition de cette ligne ou colonne, il y a divulgation. S'il y a lieu, mettez en place des contrôles pour empêcher que les distributions d'une ligne ou d'une colonne donnée soit concentrées dans un nombre restreint de catégories. Tout particulièrement, lorsque les colonnes (ou les lignes) définissent les fourchettes d'une variable quantitative, p. ex., le revenu, il faut s'assurer que les valeurs non nulles de chaque ligne (colonne) offrent une fourchette suffisamment large de valeurs possibles pour le revenu.
- Les techniques pour réduire le risque de divulgation par les tableaux de fréquence incluent toutes celles utilisées pour les tableaux de données quantitatives, c'est-à-dire; la suppression de cellules; la redéfinition des lignes, des colonnes, en regroupant des catégories ou encore des valeurs extrêmes d'une même catégorie; la perturbation aléatoire des données via l'ajout de bruit aux microdonnées ou via l'ajout de bruit aux données tabulaires, tel que l'arrondissement aléatoire; toutes autres procédures permettant d'assurer la sécurité au fichier de microdonnées à partir duquel les mises en tableaux sont exécutées.
- Assurez-vous que le Comité de la diffusion des microdonnées (Statistique Canada, 1987) examine tous les fichiers de microdonnées à grande diffusion.
- En ce qui a trait à la diffusion de microdonnées, des enregistrements individuels, par opposition aux données globales, sont publiés. Par conséquent, les critères de divulgation applicables à ces fichiers sont très différents. Bien que les fichiers de microdonnées ne contiennent pas d'information d'identification comme des noms et des numéros de téléphones, ils contiennent un certain nombre de variables, appelées *variables principales*, qui, une fois combinées, peuvent servir à identifier des personnes faisant partie de la population qui peuvent se trouver dans le fichier. L'identification équivaudrait à une divulgation des caractéristiques des microdonnées rattachées à ces personnes. Soulignons que même si les personnes identifiées ne sont pas vraiment uniques ou que si elles ont été identifiées par

erreur, l'apparence d'une divulgation peut parfois être aussi néfaste pour le Bureau qu'un cas réel de divulgation.

- Évaluez le risque de divulgation dans les fichiers de microdonnées. Le nombre et la nature des variables principales peuvent avoir une incidence sur le risque de divulgation. Certaines caractéristiques d'identification, telles que la géographie détaillé ou le revenu exact, sont considérés comme posant un risque élevé de divulgation. D'un autre côté, un faible degré de qualité, comme la présence d'erreurs de mesure ou de valeurs imputées, peut réduire les risques associés à certaines caractéristiques. Les risques de divulgation augmentent de concert avec le taux d'échantillonnage et on ne devrait pas diffuser de microdonnées pour un échantillon de 100 %. De même, les fichiers de microdonnées ne doivent pas contenir d'échantillons de 100 % à l'intérieur des strates ou des sous-groupes identifiables. Les caractéristiques de la population sondée en soi peuvent également influencer sur le risque de divulgation. Les fichiers de microdonnées liés aux entreprises sont rarement diffusés en raison de la nature concentrée des données pour les entreprises. La présence de relations hiérarchiques entre les unités peut également influencer sur le risque de divulgation.
- Il existe deux méthodes générales permettant de contrôler le risque de divulgation des microdonnées. Les *méthodes de réduction des données* englobent l'échantillonnage; dans le cas de certains groupes identifiables, s'assurer que la population est assez grande; élargir les catégories de variables; regrouper les valeurs extrêmes supérieures et inférieures, supprimer certaines variables provenant de certains répondants; supprimer les répondants du fichier. Les *méthodes de modification des données* comprennent l'ajout de bruit aléatoire aux microdonnées, la permutation de données, le remplacement de groupes restreints par des valeurs moyennes ou la suppression de renseignements fournis par certains répondants et leur remplacement par des valeurs imputées.
- Les stratégies de diffusion de fichiers de microdonnées provenant d'enquêtes longitudinales posent un problème encore plus épineux. En pareil cas, il faut déterminer une stratégie convenable avant que l'enquête longitudinale ne soit terminée. Cela suppose que la stratégie doit être élaborée avant que tous les résultats de l'enquête soient disponibles, soit avant la collecte des données pour les prochaines éditions de l'enquête. Comme un des objectifs de la stratégie consiste à définir les variables qui seront diffusées et leur catégorisation, certaines hypothèses doivent être formulées relativement à l'évolution de ces variables dans le temps, notamment à savoir si certaines variables sont susceptibles de devenir des variables clés.
- Les méthodes de réduction des données et de modification des données sont appelées les *méthodes des données à diffusion restreinte*. Comme solution de rechange à la diffusion des fichiers de microdonnées, envisagez d'utiliser des méthodes d'*accès limité* comme l'accès à distance ou les centres de données de recherche. Dans le cas de l'accès à distance, les chercheurs n'ont pas un accès

direct aux données d'enquête du Bureau, mais ils peuvent transmettre par courriel un programme analytique qui s'applique aux microdonnées que détient le Bureau. Les résultats du programme sont contrôlés par le personnel du Bureau et, s'ils ne présentent aucun risque de divulgation, sont transmis par courriel au chercheur. Les centres de données de recherche sont des établissements sûrs où les chercheurs qui présentent des projets approuvés et qui sont « assermentés » comme personnes réputées être employées en vertu de la Loi sur la statistique peuvent avoir accès à des données confidentielles. Les centres sont exploités comme des « extensions » de Statistique Canada et sont dotés d'employés à temps plein de Statistique Canada. Seuls les résultats non confidentiels peuvent être communiqués hors de ces centres.

- Les nombreuses règles établies aux fins de la protection de la confidentialité ne peuvent toutefois pas se substituer au bon sens. Par exemple, il est difficile de formuler des règles afin d'empêcher les divulgations par recoupement lorsque plusieurs produits sont diffusés à partir du même ensemble de données de base, tout particulièrement dans le cas de demandes spéciales; par conséquent, une intervention manuelle est alors nécessaire. Dans ce domaine, de nombreuses questions demeurent sans réponse; par conséquent, il conviendra d'effectuer de la recherche afin que nous puissions diffuser autant de données que possible, tout en respectant les exigences en matière de protection de la confidentialité.
- Lorsque cela est possible, utilisez un logiciel généralisé de contrôle de la divulgation plutôt que des systèmes personnalisés. Parmi les logiciels qu'on peut utiliser, mentionnons le logiciel de suppression des cellules du Bureau, CONFID (Statistique Canada, 2002d), ou τ -ARGUS Software (Hundepool et coll., 2002). En utilisant des systèmes généralisés, on peut espérer un moins grand nombre d'erreurs de programmation et une certaine réduction des coûts et du temps que nécessite la mise au point.
- S'il y a lieu, utilisez les ressources disponibles à Statistique Canada en matière de confidentialité. Consultez la Division des services d'accès et de contrôle des données à propos des politiques liées à la confidentialité de l'information recueillie par Statistique Canada, le Comité de la confidentialité et des mesures législatives et ses sous-comités, le Comité de la révélation discrétionnaire, le Comité de l'examen de la divulgation et le Comité de la diffusion des microdonnées à propos des stratégies et des pratiques de contrôle de la divulgation, ainsi que le Centre de ressources sur le contrôle de la divulgation pour de l'aide technique.

Références

Brackstone, G. et White, P. (2002). Data stewardship at Statistics Canada. **Proceedings of the Social Statistics Section**, American Statistical Association, 284-293.

- Doyle, P., Lane, J.I., Theeuwes, J.J.M. et Zayatz, L.V. (éd.) (2001). **Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies**. North-Holland.
- Eurostat (1996). **Manual on Disclosure Control Methods**. Luxembourg: Office for Official Publications of the European Communities.
- Hundepool, A., van de Wetering, A., de Wolf, P.-P., Giessing, S., Fischetti, M., Salazar, J.-J. et Caprara, A. (2002). τ -ARGUS user manual 2.1. Statistics Netherlands, Voorburg. Voir <http://neon.vb.cbs.nl/casc>.
- Statistique Canada (1970). **Loi sur la Statistique**. Ottawa, Canada.
- Statistique Canada (1987). Politique sur la diffusion des microdonnées. **Manuel des politiques**, 4.2.
- Statistique Canada (1993a). Politique relative à la révélation discrétionnaire. **Manuel des politiques**, 4.3.
- Statistique Canada (2002d). Manuel de l'utilisateur - Système Généralisé d'Analyse et de Protection de la Confidentialité. Division de développement de systèmes.
- Subcommittee on Disclosure Limitation Methodology, Federal Committee on Statistical Methodology (1994). Report on statistical disclosure limitation methodology. Statistical Policy Working Paper 22, Office of Management and Budget, Washington, DC.
- Willenborg, L. et de Waal, T. (1996). **Statistical Disclosure Control in Practice**. Lecture Notes in Statistics, Springer-Verlag, New York.
- Willenborg, L. et de Waal, T. (2001). **Elements of Statistical Disclosure Control**. Lecture Notes in Statistics, Springer-Verlag, New York.

2.14 Diffusion des données

Portée et objet

La *diffusion* consiste à mettre l'information obtenue dans le cadre d'une activité statistique à la disposition des utilisateurs. Divers modes de diffusion peuvent être envisagés, par exemple Le Quotidien de Statistique Canada; CANSIM II ou d'autres supports électroniques accessible sur Internet ; publications sur papier; microfiches; fichiers de microdonnées de réponse d'enquête anonymisées; données transmises par téléphone ou par télécopieur en réponse à une demande spéciale; allocutions prononcées en public; exposés ou entrevues à la télévision ou à la radio. Les données indiquées dans les avis « Disponibilité des données » du Quotidien sont considérées comme étant diffusées et sont régies par les présentes lignes directrices.

Principes

Il convient de s'assurer que l'information communiquée aux utilisateurs est exacte, complète, accessible, compréhensible, utilisable et actuelle; qu'elle satisfait aux exigences en matière de protection de la confidentialité; que son prix est adéquat. En matière de diffusion, on doit également tenir compte des besoins des utilisateurs et tirer parti des progrès technologiques afin de permettre aux utilisateurs de traiter les statistiques de façon rentable et efficace dans leur milieu de travail. Il convient également de prendre en considération les attentes du marché, par exemple en se fondant sur les commentaires fournis par des clients, les essais de produits et les activités de marketing.

Bon nombre des principes et des lignes directrices présentés dans la section précédente, portant sur l'évaluation de la qualité des données, et dans les deux sections suivantes, portant sur l'analyse (et la présentation) des données et la documentation, s'appliquent également à la diffusion des données. Celle-ci est régie par de nombreuses politiques de Statistique Canada (Statistique Canada, 1985a, 1985c, 1986a, 1986c, 1987, 1988, 1992a, 1992b, 1993a, 1993b, 1994a, 1994b); cependant, nous n'analyserons pas ces politiques, sauf si elles ont trait à la qualité. Par exemple, les présentes lignes directrices font abstraction des politiques relatives à l'octroi de licences, aux droits d'auteur et aux communications dans Le Quotidien, bien qu'elles soient importantes dans l'optique de la diffusion en général.

Lignes directrices

- En général, la préparation des données provenant du fichier source d'une activité statistique en vue de leur diffusion s'effectue en de nombreuses étapes. Par conséquent, on doit s'assurer que les données diffusées à la suite de toutes les étapes de traitement correspondent aux données de source obtenues. Dans le cas de données regroupées ou de variables dérivées, cela signifie qu'il faut être en mesure de reproduire les mêmes résultats à partir des données de source.

- Lorsqu'une organisation extérieure doit valider les données et lorsque cela devrait se traduire ou s'est déjà traduit par des avantages importants au chapitre de la qualité des statistiques, les données non confidentielles non diffusées peuvent être fournies à l'organisation extérieure pour validation avant leur diffusion officielle dans Le Quotidien, conformément à la Politique sur Le Quotidien de Statistique Canada (Statistique Canada, 1993b).
- On doit tester les produits électroniques avant leur diffusion pour s'assurer qu'ils fonctionnent comme prévu.
- Les mesures de la qualité des données ou, si possible, les outils nécessaires à leur calcul (p. ex. les tables de validation des coefficients de variation (CV), les programmes de variance d'échantillonnage) doivent accompagner le produit diffusé (Statistique Canada, 2000d; voir l'annexe 2).
- Une documentation doit accompagner le matériel diffusé et doit renfermer, au besoin, des descriptions de la qualité et de la méthodologie utilisée (Statistique Canada, 2000d; voir l'annexe 2). La documentation peut être sur copie papier ou sur support électronique. Lorsqu'elle est sur support électronique, elle doit renfermer des instructions aux fins d'accès à l'information.
- Les éléments de documentation requis doivent être intégrés dans la Base de métadonnées intégrée (Statistique Canada, 2000c) pour chaque période de référence de l'enquête. La BMDI est le principal instrument de mise en œuvre de la Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie (Statistique Canada, 2000d; voir l'annexe 2). La BMDI couvre les enquêtes directes, administratives et dérivées, ainsi que les combinaisons de ces types d'enquêtes.
- Le style et le formatage du produit de diffusion mis au point doivent correspondre aux produits de Statistique Canada, ce qui en facilitera l'utilisation.
- Le nom d'une personne-ressource, un numéro de téléphone et une adresse de courrier électronique doivent être fournis chaque fois que de l'information est diffusée. Un service et un soutien rapides et compétents doivent être offerts durant les heures normales de travail.

Références

Statistique Canada (1985a). Politique d'édition. **Manuel des politiques**, 3.1.

Statistique Canada (1985c). Politique en matière de diffusion de logiciels. **Manuel des politiques**, 3.5.

Statistique Canada (1986a). Politique concernant la distribution à titre gracieux aux médias. **Manuel des politiques**, 3.2.

- Statistique Canada (1986c). Politique sur la distribution de produits aux contacts provinciaux et territoriaux. **Manuel des politiques**, 3.4.
- Statistique Canada (1987). Politique sur la diffusion des microdonnées. **Manuel des politiques**, 4.2.
- Statistique Canada (1988). Politique sur le marketing. **Manuel des politiques**, 3.7.
- Statistique Canada (1992a). Information - politique d'octroi de licences. **Manuel des politiques**, 3.8.
- Statistique Canada (1992b). Politique sur les droits et les frais des produits et services. **Manuel des politiques**, 3.6.
- Statistique Canada (1993a). Politique relative à la révélation discrétionnaire. **Manuel des politiques**, 4.3.
- Statistique Canada (1993b). Politique sur le Quotidien de Statistique Canada (diffusion officielle). **Manuel des politiques**, 3.3.
- Statistique Canada (1994a). Politique de rabais. **Manuel des politiques**, 3.9.
- Statistique Canada (1994b). Politique sur l'enregistrement des produits et des services d'information. **Manuel des politiques**, 3.10.
- Statistique Canada (2000c). Base de métadonnées intégrée:
http://stdsweb/standards/imdb/imdb-menu_f.htm (Site intranet de SC).
- Statistique Canada (2000d). Politique visant à informer les utilisateurs de la qualité des données et la méthodologie. **Manuel des politiques**, 2.3. (Reproduite à l'annexe 2).

2.15 Analyse (et présentation) des données

Portée et objet

L'*analyse des données* désigne le processus de transformation des données brutes en information utilisable, souvent sous forme d'un article d'analyse publiable. Les principales étapes du processus d'analyse consistent à identifier un problème, à poser des questions éloquentes, à étayer les réponses aux questions par l'examen et l'interprétation des données ainsi qu'à communiquer le message au lecteur.

Les résultats analytiques peuvent souligner l'utilité des sources de données en jetant la lumière sur les problèmes. Certains programmes de Statistique Canada dépendent même des résultats analytiques à titre de principal produit de données car, pour des raisons de confidentialité, il est parfois impossible de diffuser les microdonnées. Au cours des dernières années, Statistique Canada a mis l'accent sur l'augmentation de la quantité des analyses pertinentes effectuées à l'aide des données qui y sont produites.

L'analyse des données joue également un rôle important dans le processus d'élaboration et de révision des enquêtes. Elle peut avoir une incidence cruciale sur la qualité des données, car elle permet de cerner les problèmes liés à la qualité des données et elle influence les améliorations apportées au processus d'enquête. L'analyse est essentielle pour comprendre les résultats des éditions antérieures des enquêtes et des études pilotes, pour planifier de nouvelles activités statistiques, pour obtenir des renseignements sur les manques en matière de données, pour concevoir les enquêtes ainsi que pour formuler des objectifs relativement à la qualité.

Principes

L'agence statistique accorde une grande importance à la pertinence et à l'utilité de l'information que contiennent ses données pour les utilisateurs. L'analyse est le principal outil permettant d'obtenir de l'information à partir des données. On peut diviser les résultats d'analyse en deux types généraux : a) les résultats descriptifs, qui sont des résultats liés à la population observée au moment où les données ont été recueillies -par exemple, le revenu médian au cours de l'année où la population a été sondée- et b) les résultats analytiques liés à une population sondée qui s'étendent souvent au-delà de la population observée au moment de la cueillette de données -par exemple, le risque qu'une personne ait une maladie chronique particulière-.

Pour être efficace, l'analyste doit connaître le public cible et les préoccupations (tant celles qui sont actuelles que celles qui sont susceptibles d'émerger dans l'avenir) lorsqu'il détermine les sujets et les façons convenables de présenter les résultats. L'étude du contexte de l'analyse permet à l'analyste de choisir les sources de données et les méthodes statistiques adéquates. Toutes les conclusions présentées dans une étude analytique, dont celles qui peuvent avoir une incidence sur les politiques publiques, doivent être appuyées par les données soumises à l'analyse.

Lignes directrices

- S'assurer que les données conviennent à l'analyse à entreprendre. Cela nécessite l'examen d'un grand nombre de détails : la population observée dans le cadre de l'enquête se rapproche-t-elle suffisamment de la population cible de l'analyse? Les définitions et concepts sous-jacents aux variables sont-ils pertinents dans le cadre de l'étude? La nature longitudinale ou transversale de l'enquête convient-elle à l'analyse? La taille de l'échantillon du domaine de l'étude est-elle suffisante pour dégager des résultats significatifs? Les mesures de qualité des données de l'enquête appuie-t-elle ces résultats?
- Si plus d'une source de données servent à l'analyse, déterminer si les sources sont cohérentes et quelle est la façon de les combiner efficacement.
- Déterminer si les valeurs imputées doivent être incluses dans l'analyse et, le cas échéant, la façon dont elles doivent être traitées (voir la section 2.9).
- Déterminer la façon de traiter la non-réponse totale et/ou partielle dans l'analyse.
- Choisir une méthode analytique qui convient à la question examinée.
- Lorsqu'on compare deux groupes de personnes, d'entreprises ou d'autres unités, il est essentiel de contrôler les facteurs extérieurs. Si les tests statistiques mettent en évidence des écarts importants entre les groupes, envisager alors des explications plausibles pour ces différences.
- Compte tenu que la majorité des analyses sont fondées sur des études par observation plutôt que sur les résultats d'une expérience contrôlée, éviter de tirer des conclusions en ce qui concerne la causalité.
- Utiliser des techniques de diagnostic pour valider le modèle analytique.
- Prendre garde à étudier les tendances à court terme en n'oubliant pas de tenir également compte des tendances à moyen et à long terme. Les tendances à court terme ne représentent souvent que de légères fluctuations d'une tendance plus importante à moyen ou à long terme.
- Lorsque cela est possible, éviter les points de référence arbitraires dans le temps, comme le changement entre l'année précédente et l'année en cours. Privilégier l'utilisation de points de référence comportant une plus grande signification tels que le dernier point tournant pour les données économiques, les différences intergénérationnelles pour les statistiques démographiques et les changements législatifs pour les statistiques sociales.
- Consulter des experts au sujet des domaines spécialisés et des méthodes statistiques.

- Les méthodes analytiques qui ne tiennent pas compte de la conception de l'enquête peuvent être utiles, pourvu que le modèle qu'elles sous-tendent au cours de l'analyse soit correct. Toutefois, les méthodes de rechange qui intègrent les renseignements sur le plan d'échantillonnage, souvent appelées les méthodes fondées sur le plan de sondage, seront généralement efficaces même lorsque certains aspects du modèle sont incorrectement spécifiés. Évaluer si l'information sur la conception de l'enquête peut être incorporée dans l'analyse et, le cas échéant, déterminer la façon de procéder. Après avoir déterminé la méthode analytique adéquate, examiner les choix de logiciel qui permettent d'appliquer la méthode. [Voir Binder et Roberts (2001) pour une définition d'un plan de sondage « ignorable », ainsi que Binder et Roberts (2003) et Skinner, Holt et Smith (1989) pour une discussion sur le fait de ne pas tenir compte du plan de l'enquête. Voir Statistique Canada (2003a), Chambers et Skinner (2003), Korn et Graubard (1999), Lehtonen et Pahkinen (1995), Lohr (1999), Thomas (1993), et Skinner, Holt et Smith (1989) pour avoir de nombreux exemples des avantages des méthodes analytiques fondées sur le plan de sondage.]
- Avant de commencer la rédaction de l'article, établir le plan de l'exposé. Le plan oblige les auteurs à se poser des questions telles que « Quel est le sujet que je veux aborder? Quelles données vais-je utiliser? Puis-je supprimer des données non pertinentes? Quelles sont les méthodes analytiques qui conviennent? Quels sont les résultats que je désire mettre en évidence? Quels sont les résultats intéressants? »
- Mettre l'accent sur les variables et les sujets importants dans les articles. En abordant un sujet trop vaste, l'impact principal du message s'en trouve souvent atténué.
- Structurer les idées de façon logique, en fonction de leur pertinence ou de leur importance. Recourir à des titres, à des sous-titres et à des encadrés afin de renforcer la structure de l'article.
- Rédiger un texte aussi simple que le sujet le permet. Selon le public ciblé de l'article, il est parfois souhaitable de perdre un peu en précision pour rendre le texte plus lisible.
- Insérer des graphiques en complément du texte ou des tableaux pour communiquer le message. Privilégier les titres qui véhiculent un message (p. ex. « Les revenus des femmes demeurent inférieurs à ceux des hommes ») au détriment des titres de graphique classiques (« Revenus selon l'âge et le sexe »). Toujours commenter l'information transmise dans les tableaux et les graphiques afin de faciliter la compréhension des lecteurs.
- Lorsque des tableaux sont insérés, la présentation générale doit contribuer à la clarté des données qu'ils contiennent et prévenir les erreurs d'interprétation. Cela

comprend l'espace, la formulation, l'emplacement et l'apparence des titres, les titres des lignes et des colonnes et les autres identifications.

- Expliquer les pratiques ou les méthodes d'arrondissement. Dans la présentation des données arrondies, le nombre de chiffres significatifs ne doit pas être trop grand de manière à être cohérent avec la précision des données.
- Lorsqu'on présente des détails sur les taux, prendre soin d'établir la distinction entre le taux de variation et la variation en points de pourcentage. Définir la base de calcul des taux.
- S'assurer que toutes les références sont exactes et qu'elles font l'objet de renvois dans le texte.
- S'assurer qu'il n'y a pas d'erreurs dans les articles. Vérifier les détails, par exemple la cohérence des chiffres dans le texte, les tableaux et les graphiques ainsi que l'exactitude des données externes et des calculs arithmétiques simples.
- S'assurer que ce qui est annoncé dans l'introduction est effectivement exprimé dans le reste de l'article. S'assurer que les conclusions sont cohérentes avec les résultats de l'analyse.
- Faire réviser l'article par au moins deux personnes. Le cas échéant, vérifier la qualité de la traduction.
- Comme pratique efficace, envisager de donner un exposé sur les résultats d'analyse obtenus. Il s'agit d'un autre type d'examen par les pairs qui peut contribuer à améliorer l'article. Toujours pratiquer les exposés qui sont destinés à des publics cibles externes.

Références

Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. **International Statistical Review**, 51, 279-292.

Binder, D.A. et Roberts, G. (2001). Can informative designs be ignorable? Newsletter of the Survey Research Methods Section, American Statistical Association, Issue 12.

Binder, D.A. et Roberts, G.R. (2003). Design-based methods for estimating model parameters. Dans **Analysis of Survey Data**, R.L. Chambers et C.J. Skinner (éd.), Wiley, Chichester, 29-48.

Chambers, R.L. et Skinner, C.J. (éd.) (2003). **Analysis of Survey Data**. Wiley, Chichester.

Korn, E.L. et Graubard, B.I. (1999). **Analysis of Health Surveys**. Wiley, New York.

- Lehtonen, R. et Pahkinen, E.J. (1995). **Practical Methods for Design and Analysis of Complex Surveys**. Wiley, Chichester.
- Lohr, S.L. (1999). **Sampling: Design and Analysis**. Duxbury Press.
- Skinner, C.K., Holt, D. et Smith, T.M.F. (1989). **Analysis of Complex Surveys**. Wiley, Chichester.
- Statistique Canada (1995). Politique concernant l'évaluation des produits d'information. **Manuel des politiques, 2.5**.
- Statistique Canada (2001a). Lignes directrices sur la rédaction d'articles d'analyse. Division des communications.
- Statistiques Canada (2003a). **Analysis Handbook**. Préparée par le Centre de ressources en analyse de données, Direction de la méthodologie.
- Statistique Canada. (2003e). Le guide officiel de rédaction des Services de rédaction. Services de rédaction, Division des communications. Voir http://icn-rci.statcan.ca/10/10d/10d_000_f.htm (Site intranet de SC). Publication mise à jour régulièrement.
- Thomas, D.R. (1993). Inference using complex data from surveys and experiments. **Canadian Psychology**, 34, 415-431.

2.16 Documentation

Portée et objet

La *documentation* est un compte rendu de l'activité statistique, notamment des concepts, des définitions et des méthodes utilisés aux fins de la production des données. Elle décrit également les facteurs qui influent sur la comparabilité et la qualité des données.

Les utilisateurs peuvent se reporter à la documentation pour savoir ce qui a été effectué dans le but de fournir un contexte pour l'utilisation efficace et informée des données. Durant la mise en oeuvre, la documentation est un mode de communication assurant l'efficacité de l'élaboration de l'activité statistique. De plus, elle fait état non seulement des décisions qui ont été prises, mais également de leur bien-fondé, et renferme des renseignements qui seront utiles lorsque la même activité statistique, une activité similaire ou une activité remaniée seront élaborées et mises en oeuvre.

Principes

La documentation vise à fournir un compte rendu complet, clair et polyvalent de l'enquête, notamment des données produites. La documentation se rapportant à n'importe quelle étape de l'enquête peut être destinée à un des groupes tels que les gestionnaires, le personnel technique, les planificateurs d'autres enquêtes et les utilisateurs. La documentation fournie doit être actuelle, bien structurée, facilement accessible, concise et précise. Une présentation efficace des résultats constitue une facette importante de la documentation. On peut recourir à l'approche multimédia (copie papier, support électronique, présentation visuelle) et on peut préparer différents documents selon le public-cible et l'objectif à atteindre.

Lignes directrices

- Le niveau de détail de la documentation varie selon le public-cible, la méthode de collecte, les sources de données, l'analyse, le mode de diffusion, l'éventail d'utilisation de l'information et leur impact, ainsi que le budget total du programme statistique. S'assurer que tous les produits statistiques répondent aux exigences de la Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie (Statistique Canada, 2000d; voir l'annexe 2).
- Selon le public-cible et l'objectif à atteindre, la documentation peut renfermer ce qui suit :
 - Objectifs : Inclure des renseignements sur les objectifs et les utilisations des données, l'actualité, la fréquence et les objectifs relatifs à la qualité des données. Les objectifs peuvent changer à mesure que l'enquête progresse (par exemple, en raison de contraintes budgétaires, de la faisabilité anticipée, des résultats de nouvelles études pilotes et de nouvelles technologies). Documenter ces changements puisqu'ils ont un

impact sur la conception du questionnaire et l'analyse des résultats des mises à l'essai.

- Contenu : Inclure le questionnaire utilisé, les concepts et les définitions. Dans le cas d'interviews assistées par ordinateur, fournir les spécifications de développement de l'application informatique. Afin de faciliter l'intégration à d'autres sources, mettre en relief le recours à des concepts, à des questions, à des méthodes, à des classifications types et mettre en évidence les différences, s'il y a lieu. Mentionner le rôle des comités consultatifs et des utilisateurs.
- Tests : Décrire les tests cognitifs, les tests sur le terrain ou les enquêtes pilotes et faire état des résultats et des recommandations par rapport aux spécifications.
- Méthodologie : Décrire des plans d'enquête alternatifs. Traiter des questions telles que la population-cible, la base de sondage, la couverture, la période de référence, le plan d'échantillonnage, la taille et la méthode de sélection de l'échantillon, la méthode de collecte et les méthodes de suivi en cas de non-réponse, l'estimation, la vérification et l'imputation, l'étalonnage et la révision, la désaisonnalisation, la confidentialité et l'évaluation. Accentuer divers aspects à l'intention de différents lecteurs. Grouper les questions techniques dans un même document à l'intention des professionnels. Fournir un survol méthodologique.
- Systèmes : Inclure des renseignements sur les fichiers de données (méthode de saisie, clichés d'article, explication des codes, fréquences de base, méthodes de vérification), les systèmes (construction, algorithmes, utilisations, stockage et extraction) et les rapports de surveillance (temps consacré à des activités spécifiques, sources d'incident, ordonnancement des soumissions pour déterminer si le traitement est effectué dans les délais prévus).
- Opérations : Inclure ou donner en référence les guides de formation, les guides du personnel chargé du traitement et de la saisie des données, les guides des intervieweurs, les rapports d'évaluation et les comptes rendus.
- Mise en oeuvre : Documenter les opérations en spécifiant clairement les données d'entrée et de sortie. Annexer le calendrier de travail de chaque étape de la mise en oeuvre.
- Contrôle de la qualité : Inclure les instructions et/ou guide destinés aux superviseurs et aux vérificateurs.
- Qualité des données : Renseignements d'usage général incluant la couverture, l'erreur d'échantillonnage, l'erreur non due à l'échantillonnage, les taux de réponse, les taux de vérification et d'imputation et l'impact de ces derniers, la comparabilité dans le temps et avec d'autres données, les études de validation, les mesures d'assurance de la qualité et toute autre mesure pertinente qui est propre à l'activité statistique en question. Décrire les événements imprévus qui influent sur la qualité des données (p. ex. inondations, taux de non-réponse élevé). À l'intention des utilisateurs spécialisés, inclure la variance totale ou ses

composantes selon la source, les biais de non-réponse et de réponse, ainsi que l'impact et l'interprétation de la désaisonnalisation.

- Ressources : Établir une liste des ressources qui ont été utilisées en fonction du temps. Fournir un relevé des dépenses salariales et non salariales (montants et temps). Commenter les dépenses par rapport aux budgets.
 - Références : Classer et documenter les références (articles théoriques et généraux et documents qui se rapportent au projet mais qui n'ont pas été produits dans le cadre de celui-ci).
- Fournir les éléments de documentation requis à la Base de métadonnées intégrée (Statistique Canada, 2000c). En tant qu'organe d'archivage de l'information sur les enquêtes et programmes de Statistique Canada, la BMDI contient la majeure partie de l'information sur la méthodologie et l'exactitude des données. Pour les produits électroniques, un lien vers l'enregistrement pertinent de la BMDI sert à accéder à la documentation du produit. Pour les produits imprimés, le texte de la BMDI sert à assurer une documentation adéquate, en accord avec la Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie (Statistique Canada, 2000d; voir l'annexe 2).
- Tenir compte des lecteurs de chaque document. Demander à des gestionnaires, à des représentants des lecteurs-cibles, et à des pairs de procéder à une étude approfondie des documents afin de s'assurer de leur qualité et de leur lisibilité (Statistique Canada, 1995). Réviser méticuleusement les documents. Dater chaque version du document. Vérifier les références.

Références

Nations Unies (1964). Recommendations for the preparation of sample survey reports (édition provisoire). **Statistical Series C**, 1, New York.

Nations Unies, Conférence des statisticiens européens (1983). **Draft guidelines for the preparation of presentations of the scope and quality of statistics for users**. Genève, Suisse.

Statistique Canada (1995). Politique concernant l'évaluation des produits d'information. **Manuel des politiques**, 2.5.

Statistique Canada (1998b). Politique concernant les normes. **Manuel des politiques**, 2.10.

Statistique Canada (2000c). Base de métadonnées intégrée: http://stdsweb/standards/imdb/imdb-menu_f.htm (Site intranet de SC).

Statistique Canada (2000d). Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie. **Manuel des politiques**, 2.3. (Reproduite à l'annexe 2).

Statistique Canada (2001d). Normes et lignes directrices pour la déclaration des taux de non-réponse. Rapport technique de Statistique Canada.

Statistique Canada (2002c). Le Cadre d'assurance de la qualité de Statistique Canada - 2002. N° 12-586-XIF au catalogue, Statistique Canada

Statistique Canada. (2003e). Le guide officiel de rédaction des Services de rédaction. Services de rédaction, Division des communications. Voir http://icn-rci.statcan.ca/10/10d/10d_000_f.htm (Site intranet de SC). Publication mise à jour régulièrement.

2.17 Utilisation des données administratives

Portée et objet

Les *dossiers administratifs* sont constitués aux fins de l'administration de divers programmes non statistiques. Par exemple, on conserve des dossiers administratifs pour régulariser le mouvement transfrontalier des biens et des personnes, pour satisfaire aux exigences légales de l'enregistrement de certains événements, comme les naissances et les décès, et pour administrer les avantages sociaux, comme les pensions, ou les obligations, comme les impôts. Leur raison d'être est liée à la prise de certaines décisions; l'identité de l'unité correspondant à un dossier donné est donc capitale. Par contre, dans le cas des *dossiers statistiques* qui, eux, ne servent pas et ne peuvent pas servir de fondement à une mesure visant un individu, l'identité des individus ne présente aucun intérêt une fois que la base de données est complètement constituée.

Les dossiers administratifs présentent un certain nombre d'avantages pour une agence statistique ou pour des analystes. Les demandes en matière de statistiques liées à tous les aspects de nos vies, de notre société et de notre économie continuent de croître. Ces demandes se présentent souvent dans un contexte de contraintes budgétaires importantes. En outre, les agences statistiques partagent avec de nombreux répondants un souci de plus en plus présent à l'égard de la croissance du fardeau de réponse des enquêtes. Les répondants peuvent également réagir de façon négative s'ils ont l'impression d'avoir déjà communiqué des renseignements semblables (p. ex., revenu) à des enquêtes et des programmes administratifs. Comme ils sont déjà constitués, les dossiers administratifs ne sous-tendent pas le coût de la collecte de données directe ni n'imposent un fardeau supplémentaire aux répondants. Il importe de souligner que l'« explosion » de la technologie a également permis aux agences statistiques de surmonter les obstacles associés au traitement des ensembles de données volumineux. Pour toutes ces raisons, les dossiers administratifs deviennent de plus en plus utiles et sont utilisés à des fins statistiques.

En matière de statistique, les dossiers administratifs servent (i) aux bases de sondage, à titre de base ou de supplément d'une base existante, (ii) au remplacement de la collecte de données (p. ex., utilisation des données fiscales pour les petites entreprises au lieu de chercher à obtenir des données d'enquête à leur sujet), (iii) à la vérification et à l'imputation, (iv) à la tabulation directe, (v) à l'estimation, de façon indirecte (p. ex., comme information auxiliaire dans l'estimation par calage aux marges, de l'étalonnage ou de la calendrialisation) et vi) à l'évaluation de l'enquête, ce qui comprend la *confrontation des données* (p. ex., comparaison des estimations de l'enquête avec les estimations rattachées à un programme administratif connexe).

Principes

Statistique Canada a pour politique d'utiliser des dossiers administratifs chaque fois qu'un tel recours constitue une solution de rechange rentable à la collecte directe de données. Tout comme pour n'importe quel programme d'acquisition de données, il

convient de soupeser les coûts et les avantages de l'utilisation de dossiers administratifs à des fins statistiques; cependant, cette solution jouit d'une supériorité écrasante du fait qu'elle évite les coûts inhérents à la collecte de données et n'augmente pas le fardeau de réponse, pourvu que la couverture et le cadre conceptuel des données administratives sont compatibles avec la population cible. Selon l'objectif à atteindre, il est souvent avantageux de combiner des données administratives avec des données provenant d'une autre source.

Le recours aux dossiers administratifs peut soulever des questions relativement à la protection de la confidentialité de l'information du domaine public. Ces questions sont encore plus importantes lorsque les dossiers administratifs sont couplés avec d'autres sources d'information. La Politique d'information des répondants aux enquêtes (Statistique Canada, 1998a) exige que Statistique Canada communique à tous les répondants des renseignements tels que l'objet de l'enquête, la protection de la confidentialité, les plans de couplage des enregistrements et l'identité des parties à toute entente visant à partager l'information transmise par les répondants. Le couplage des enregistrements doit être conforme à la Politique relative au couplage d'enregistrements du Bureau (Statistique Canada, 1996a). En particulier, toutes les demandes de couplage d'enregistrements doivent être présentées au Comité de la confidentialité et des mesures législatives et approuvées par le Comité des politiques.

Aux fins de l'utilisation de données administratives, l'agence statistique doit parfois réaliser certaines étapes du processus d'enquête, généralement un petit nombre, que nous avons décrites dans les sections précédentes. Cela s'explique du fait que bon nombre des étapes du processus d'enquête (p. ex. la collecte directe et la saisie des données) sont réalisées par l'organisation responsable des données administratives. Par conséquent, il faut ajouter d'autres lignes directrices à celles qui ont été présentées afin de proposer des façons de compenser les différences au chapitre des objectifs de qualité de l'organisme responsable des données (p. ex., pour compenser la qualité de la saisie des données qui, souvent, n'est pas vérifiée).

Il ne faut pas oublier la raison fondamentale qui justifie l'existence de ces dossiers administratifs, qui sont le résultat d'un programme administratif mis en place pour des raisons administratives. Bien souvent, les utilisations statistiques de ces dossiers étaient inconnues lorsque le programme était mis en œuvre et l'agence statistique a invariablement un impact limité sur l'élaboration du programme. Pour cette raison, toute décision relative à l'utilisation des dossiers administratifs doit être précédée d'une évaluation de ces dossiers sur le plan de la couverture, du contenu, des concepts et des définitions, des procédures d'assurance et de contrôle de la qualité mises en place par le programme administratif pour en assurer la qualité, de la fréquence des données, de la rapidité de l'agence statistique à recevoir les données et de la stabilité du programme au fil du temps. Il va de soi que le coût associé à l'obtention des dossiers administratifs est également un facteur clé à prendre en compte dans la décision d'utiliser ou non ces dossiers.

Lignes directrices

- Bon nombre des lignes directrices présentées dans les sections précédentes s'appliquent aux dossiers administratifs. Les lignes directrices relatives à l'échantillonnage et à la saisie des données (voir les sections 2.4 et 2.7) sont pertinentes si les dossiers administratifs existent uniquement sur papier et doivent être codés et saisis. Ces lignes directrices seront également précieuses pour les données administratives accessibles en format électronique, y compris l'ÉÉD et la CÉD. Soulignons que ces données, puisqu'elles sont disponibles en format électronique, peuvent être implicitement moins stables et sujettes à des erreurs additionnelles causées par les processus de traitement et de transmission des données à la source. Les lignes directrices se rapportant à la vérification et à la diffusion (voir les sections 2.8 et 2.14) doivent être respectées lorsqu'on obtient ou crée un fichier de dossiers administratifs d'individus pour analyse et traitement ultérieurs.
- Tenir compte des répercussions sur la protection des renseignements personnels de la publication de données tirées des dossiers administratifs. Bien que, en vertu de la Loi sur la statistique, Statistique Canada a le droit d'accéder aux dossiers administratifs à des fins statistiques, il se peut que ceux qui ont fourni les renseignements à l'origine n'aient pas prévu qu'ils seraient utilisés de la sorte. (Statistique Canada, 1970). Par conséquent, les responsables de programme devraient être prêts à justifier cette utilisation secondaire en expliquant qu'elle est sans gravité et dans l'intérêt public.
- La collaboration avec les concepteurs des systèmes administratifs nouveaux ou remaniés peut faciliter le renforcement des exigences statistiques de ces systèmes dès le départ. De telles possibilités sont rares; cependant, lorsqu'elles se présentent, les avantages éventuels de la participation de l'agence statistique valent largement le temps qu'elle requière.
- Assurer une liaison constante avec le fournisseur des dossiers administratifs. Il faut entrer en contact avec lui dès qu'on commence à les utiliser. Cependant, il est encore plus important de demeurer constamment en contact étroit avec le fournisseur afin que l'agence statistique ne soit pas prise au dépourvu par les changements imminents et puisse même les influencer. Les commentaires sur les données statistiques et leurs lacunes peuvent être utiles au fournisseur, ce qui se traduira par une amélioration de la source administrative.
- Il faut connaître le contexte dans lequel l'organisme administratif a créé le programme administratif (p. ex., législation, objectifs et besoins). Cela a une influence très importante sur i) la population couverte, ii) le contenu, iii) les concepts et les définitions, iv) la fréquence et l'actualité, v) la qualité de l'information emmagasinée et vi) la stabilité au fil du temps.

- Examiner chaque élément de donnée des dossiers administratifs qu'on prévoit utiliser à des fins statistiques. Examiner la qualité des données. Connaître les concepts, les définitions et les procédures qui sous-tendent leur collecte et leur traitement par l'organisme administratif. Certains éléments peuvent être de très mauvaise qualité et donc inutilisables. Par exemple, la qualité du codage de classification (selon la profession, l'activité industrielle, la géographie) peut être inadéquate d'un point de vue statistique ou en limiter l'utilisation.
- Tout comme les données d'enquête, les données administratives ne sont pas à l'abri de la non-réponse, qu'elle soit partielle ou totale. Dans certains cas, le manque de rapidité dans l'obtention de toutes les données administratives donne lieu à un taux de non-réponse plus élevé. Par conséquent, certaines des lignes directrices exposées à la section 2.6 s'appliqueront. À moins de pouvoir effectuer un suivi des non-répondants et obtenir les réponses voulues, il faut élaborer une procédure d'imputation ou de rajustement des poids pour composer avec la non-réponse (voir les sections 2.9 et 2.10). Les sources administratives sont parfois désuètes. Ainsi, dans le cadre du processus d'imputation, accorder une attention spéciale à l'identification des unités actives et/ou inactives. Il peut également être nécessaire de recourir à l'imputation ou à la transformation lorsque certaines unités fournissent les données à une fréquence différente (p. ex., hebdomadaire ou trimestrielle) de la fréquence souhaitée (p. ex., mensuelle).
- Garder à l'esprit que si l'information qu'ils fournissent à la source administrative peut causer des gains ou des pertes à des personnes ou à des entreprises, il est possible que l'information communiquée comporte des biais. Il peut être nécessaire de mener des études spéciales pour évaluer et comprendre ces sources d'erreur.
- Documenter la nature et la qualité des données administratives dès qu'elles sont évaluées. La documentation aide les statisticiens à déterminer les utilisations qui conviennent le mieux aux données administratives. Choisir des méthodes adéquates pour le programme statistique en fonction des données administratives et informer les utilisateurs de la méthodologie et de la qualité des données.
- Garder à l'esprit que la longévité et la portée constante de la source des données administratives sont, de façon générale, entièrement entre les mains de l'organisme administratif. Les éléments administratifs qui ont initialement dicté les concepts, les définitions, la couverture, la fréquence, l'actualité et les autres attributs du programme administratif peuvent, au fil du temps, subir des changements qui distordent les séries chronologiques dérivées de la source administrative. S'informer de ces changements et déterminer leur incidence sur le programme statistique.
- Effectuer une évaluation permanente ou périodique de la qualité des données transmises. L'assurance que la qualité des données est conservée est importante, car l'agence statistique ne contrôle pas le processus de collecte des données. Cette

évaluation peut consister en la mise en œuvre de sauvegardes et de contrôles additionnels (p. ex., l'utilisation de méthodes et de procédures de contrôle de la qualité statistique, les règles de vérification) au moment de la réception des données, de comparaisons avec d'autres sources ou d'études de suivi d'un échantillon.

- Lorsqu'on doit coupler des dossiers administratifs (p. ex. pour le dépistage de répondants, pour remplacer des données d'enquête ou pour analyser des données), on doit respecter la Politique relative au couplage d'enregistrements. L'utilisation d'une seule source de données administratives suscite de l'appréhension au plan de la protection des renseignements personnels; cependant, l'appréhension est multipliée lorsque la source administrative est couplée à d'autres sources. En pareil cas, il se peut que les sujets ne sachent pas que les renseignements fournis en deux occasions distinctes sont combinés. La Politique relative au couplage d'enregistrements vise à s'assurer que l'intérêt public de chaque couplage l'emporte largement sur les atteintes à la vie privée qu'il pourrait occasionner.
- Il n'est pas toujours facile de combiner une source de données administratives à une autre source d'information. Cette tâche est particulièrement ardue lorsqu'il n'y a pas de clé d'appariement commune aux deux sources et que des techniques de couplage doivent être utilisées. En pareil cas, le type de méthode de couplage (c.-à-d. le couplage exact ou statistique) doit être choisi en fonction des objectifs du programme statistique. Lorsque le programme a pour but la création ou la mise à jour d'une base de sondage, la vérification et l'imputation ou la pondération, le couplage exact est approprié. Lorsqu'on couple les sources afin d'effectuer des analyses de données qui, autrement, ne pourraient pas être réalisées, le couplage statistique (c.-à-d. l'appariement d'enregistrements ayant des propriétés statistiques similaires) peut s'avérer un choix judicieux (voir Cox et Boruch, 1988, Scheuren et Winkler, 1993, Kovacevic, 1999).
- Lorsqu'on doit procéder à un couplage d'enregistrements, il convient de faire bon usage des logiciels existants. Il existe un certain nombre de logiciels bien documentés, par exemple le Système généralisé de couplage d'enregistrements de Statistique Canada.
- Au cours de la combinaison de données provenant de plus d'une source administrative, accorder davantage d'attention au rapprochement des différences potentielles dans les concepts, les définitions, les dates de référence, la couverture et les normes de qualité des données appliqués à chaque source de données. Parmi les exemples, mentionnons les sources de données sur l'éducation, les rapports sur la santé et le crime, ainsi que les registres des naissances, des mariages, des immatriculations et des véhicules enregistrés, qui sont fournis par diverses organisations et divers organismes gouvernementaux.
- Certaines données administratives ont un caractère longitudinal (p. ex. impôt sur le revenu et taxe sur les produits et services). Lorsque des enregistrements de périodes de référence différentes sont combinés, ils constituent des mines de

données très riches pour les chercheurs. Il faut demeurer particulièrement vigilant lorsqu'on crée des bases de données longitudinales et axées sur des personnes, car leur utilisation soulève de très importantes questions en matière de protection des renseignements personnels. L'identificateur doit être utilisé avec soin, car une unité peut changer d'identificateur avec le temps. Faire le suivi de tels changements afin que l'analyse temporelle des données soit adéquate. Dans certains cas, la même unité peut avoir deux identificateurs ou plus pour la même période de référence, ce qui suppose un dédoublement dans le dossier administratif. Il convient alors d'élaborer un mécanisme d'élimination du dédoublement.

- On a parfois recours aux données administratives au lieu de poser une série de questions au répondant. En pareil cas, il peut être nécessaire d'obtenir la permission du répondant; à cet égard, la Politique d'information des répondants aux enquêtes doit être respectée. Lorsque le consentement n'a pas été obtenu, mettre en place des mécanismes de collecte afin de poser les questions d'enquête correspondantes aux répondants.
- Les fichiers administratifs sont souvent très volumineux et leur utilisation peut parfois entraîner des coûts de traitement considérables et donner lieu à des problèmes d'actualité. Selon les besoins, utiliser un échantillon aléatoire provenant de fichiers administratifs pour réduire les coûts.

Références

- Brackstone, G.J. (1987). Utilisation des dossiers administratifs à des fins statistiques. **Techniques d'enquête**, 13, 35-51.
- Cox, L.H. et Boruch, R.F. (1988). Record linkage, privacy and statistical policy. **Journal of Official Statistics**, 4, 3-16.
- Hidiroglou, M.A., Latouche, M.J., Armstrong, B. et Gossen, M. (1995). Improving survey information using administrative records: the case of the Canadian Employment Survey. **Proceedings of the Annual Research Conference**, U.S. Bureau of the Census, 171-197.
- Kovacevic, M. (1999). Record linkage and statistical matching - they aren't the same! **SSC Liaison**, Vol. 13, No. 3, 24-29.
- Michaud, S., Dolson, D., Adams, D., et Renaud, M. (1995). Combining administrative and survey data to reduce respondent burden in longitudinal surveys. **Proceedings of the Section on Survey Research Methods**, American Statistical Association, 11-20.
- Monty, A. et Finlay, H. (1994). Strengths and weaknesses of administrative data Sources: experiences of the Canadian Business Register. **Statistical Journal of the United Nations**, ECE 11, 205-210.

- Scheuren, F. et Winkler, W.E. (1993). Analyse de régression de fichiers de données couplés par ordinateur. **Techniques d'enquête**, 19, 45-65.
- Statistique Canada (1970). **Loi sur la Statistique**. Ottawa, Canada.
- Statistique Canada (1996a). Politique relative au couplage d'enregistrements. **Manuel des politiques**, 4.1.
- Statistique Canada (1998a). Politique d'information des répondants aux enquêtes. **Manuel de politiques**, 1.1.
- Wolfson, M., Gribble, S., Bordt, M., Murphy, B. et Rowe, G. (1987). La base de données de simulation de politique sociale : un exemple d'intégration de données d'enquêtes et de données administratives. **Recueil du Symposium international sur les utilisations statistiques des données administratives**, Statistique Canada, 233-266.

3. Contexte de gestion

La présente section décrit le contexte de gestion qui sert de toile de fond à l'application des Lignes directrices concernant la qualité. La structure de gestion, les politiques et les lignes directrices, les mécanismes de consultation, l'approche de réalisation et de gestion des projets et l'environnement du Bureau ont été mis en place pour faciliter et assurer une gestion efficace de la qualité. Les mécanismes fondamentaux de la gestion de la qualité, qui font partie intégrante du Cadre d'assurance de la qualité de Statistique Canada (Statistique Canada, 2002c), sont résumés ci-dessous en fonction des six éléments de qualité décrits à la section 1.

3.1 La qualité à Statistique Canada

L'information est le produit qu'offre Statistique Canada. Si cette information devenait douteuse, la crédibilité du Bureau serait remise en question et sa réputation de source indépendante et objective d'information fiable serait minée. Par conséquent, la gestion de la qualité doit jouer un rôle de premier plan dans la gestion générale du Bureau et faire partie intégrante de la gestion de tous les programmes.

Une des caractéristiques importantes de la gestion de la qualité est l'atteinte d'un équilibre entre les objectifs de qualité et les contraintes des ressources financières et humaines, la volonté des répondants de fournir les données de base et les demandes concurrentes pour une plus grande quantité d'informations. La gestion de la qualité ne signifie pas l'optimisation de la qualité à tout prix, mais l'atteinte d'un juste équilibre entre la quantité et la qualité de l'information produite par le programme du Bureau et les ressources disponibles. Dans chacun des programmes, le défi consiste à faire des compromis judicieux entre les besoins changeants des clients, les coûts, le fardeau des répondants et les divers éléments ou dimensions de la qualité.

Statistique Canada s'applique à intégrer la qualité – *adaptation à l'utilisation* - à tous ses programmes et dans tous ses produits. La qualité de ses statistiques officielles repose sur l'emploi de méthodes scientifiques éprouvées et adaptées progressivement aux besoins changeants des clients, à la réalité en évolution que le Bureau tente de mesurer et à la capacité ou la volonté des répondants de fournir des données fiables et actuelles. Les présentes Lignes directrices concernant la qualité sont un des outils qui aideront le Bureau à intégrer la qualité à la conception de chacun des programmes.

3.2 Cadre de gestion

À Statistique Canada, la gestion de la qualité est assurée dans le contexte d'un cadre de gestion matricielle, c'est-à-dire que la gestion des projets relève d'une organisation fonctionnelle. À cet égard, le Bureau est divisé en sept secteurs. Quatre de ces secteurs sont principalement responsables des programmes statistiques de production et d'analyse des données dans différents domaines spécialisés. Les trois autres secteurs sont avant tout chargés de la prestation d'une infrastructure et de services aux programmes statistiques. Généralement, un programme statistique est géré par une division spécialisée, et ses

responsables font un usage intensif des ressources des secteurs de service, notamment en ce qui concerne les activités de collecte et de traitement des données, le soutien informatique, la méthodologie statistique ainsi que la commercialisation et la diffusion.

Le recours à une approche interdisciplinaire de gestion par équipe de projet pour la conception ou la restructuration d'un programme statistique est importante. Il permet de s'assurer qu'on accorde suffisamment d'attention à la qualité de toutes les composantes et étapes du programme au cours de sa conception, de sa mise en œuvre et de son évaluation. Il incombe aux organisations fonctionnelles de s'assurer que les équipes de projet soient composées de personnes possédant la crédibilité et les compétences requises pour représenter leur secteur fonctionnel. Le personnel spécialisé met à contribution sa connaissance du contenu, des besoins des clients et de la pertinence des données. Pour leur part, les méthodologistes tirent parti de leurs connaissances en regard de l'équilibre à atteindre entre les méthodes statistiques et la qualité des données, particulièrement en ce qui concerne l'exactitude, l'actualité et le coût. Les experts des opérations ont quant à eux l'expérience voulue des méthodes opérationnelles et s'attardent aux questions d'ordre pratique, à l'efficacité, du personnel sur le terrain et aux répondants. Enfin, les spécialistes des systèmes connaissent bien les normes technologiques et les outils à utiliser pour concevoir les projets. Ils apportent aussi la dimension système aux projets.

C'est au sein d'une telle équipe qu'on prend les nombreuses décisions et qu'on fait les compromis nécessaires pour atteindre un juste équilibre entre la qualité, d'une part, et le coût et le fardeau de réponse, d'autre part. Ensemble, les membres de l'équipe doivent trouver un juste équilibre afin de mettre au point une conception optimale. Le fait que chacun des membres de l'équipe fasse partie d'une organisation fonctionnelle spécialisée, à laquelle ils peuvent avoir recours au besoin pour obtenir des ressources de gestion et diverses ressources plus spécialisées, leur permet de résoudre les problèmes techniques et les différends qui peuvent survenir au cours de la réalisation d'un projet.

Les projets sont habituellement orientés par un comité directeur formé de cadres supérieurs qui peut réunir des gestionnaires de chacun des principaux secteurs participants. Ce comité, qui fait partie d'un mécanisme d'approbation officielle de la conception et de la mise en œuvre du programme, fournit une orientation d'ensemble ainsi que les grands paramètres relatifs au budget et à la conception. Il veille aussi à ce que les ressources appropriées soient mises à la disposition des responsables du projet et s'emploie à résoudre les problèmes qui ne peuvent pas être réglés de façon satisfaisante au sein de l'équipe de projet.

Ce cadre de gestion est une composante systémique importante du Cadre d'assurance de la qualité de Statistique Canada.

3.3 Cadre d'assurance de la qualité

Le *Cadre d'assurance de la qualité* est l'ensemble des pratiques, des procédures et des mécanismes de gestion, d'exploitation et de consultation que Statistique Canada utilise pour gérer la qualité de ses produits d'information. Ce cadre a été mis au point et adapté

pendant de nombreuses années et continue d'évoluer. Il établit un lien entre les besoins des utilisateurs et les produits des programmes, et représente un outil de rétroaction, d'examen du rendement, ainsi que de planification et d'élaboration permanentes. Le cadre offre une orientation aux gestionnaires de projet et de programme, puis à leurs équipes, afin de leur permettre d'atteindre une cohérence et un équilibre généraux au sein des programmes, compte tenu des priorités conflictuelles, des contraintes, de la conception et de la qualité. Le Cadre d'assurance de la qualité est résumé ci-dessous dans le contexte des six éléments rattachés à la qualité : la pertinence, l'exactitude, l'actualité, l'accessibilité, l'intelligibilité et la cohérence. On trouvera également, ci-dessous, une brève analyse des principaux piliers de la qualité, sous la rubrique Environnement du Bureau.

3.3.1 Gestion de la pertinence

La gestion de la pertinence englobe les processus qui permettent de déterminer la nature de l'information produite par le Bureau et le niveau des ressources qui doivent être affectées à chaque programme. Cela consiste essentiellement à traduire les besoins des utilisateurs en autorisations de programme et en décisions budgétaires au sein du Bureau. Les processus qui servent à assurer la pertinence permettent aussi d'exercer un contrôle de base sur les autres aspects de la qualité et ainsi, d'évaluer les besoins des utilisateurs par rapport à ces autres dimensions.

Pour que le Bureau puisse réaliser son mandat, il est essentiel que ses programmes et ses résultats reflètent adéquatement et constamment les besoins en information les plus importants pour le pays. Comme ces besoins évoluent au fil du temps, il est primordial de procéder à l'examen continu des programmes en fonction des besoins des clients et d'apporter les rajustements nécessaires.

On établit les besoins des utilisateurs en maintenant une liaison bilatérale et multilatérale avec les principaux utilisateurs, en utilisant l'information et les conseils fournis par les organismes statistiques et les groupes consultatifs, et en consultant les commentaires des utilisateurs sur les produits et les services existants. On procède à des examens réguliers de tous les programmes en préparant des rapports biennaux et quadriennaux de programme, ainsi qu'en effectuant des vérifications et des examens ponctuels.

L'analyse des données offre également une rétroaction sur les lacunes et les limites de l'information de la façon suivante : directement par les analystes, par des articles publiés et par les processus d'examen par les pairs appliqués à ces articles, par une rétroaction ou des commentaires donnant suite à des résultats analytiques et par l'utilisation de cadres analytiques comme le Système de comptabilité nationale, qui intègrent et rapprochent les données provenant des différentes sources de Statistique Canada.

Les décisions et les rajustements concernant les programmes sont généralement établis dans le cadre d'un processus annuel de planification stratégique à long terme qui examine les besoins en information nouveaux et émergents. Outre les besoins des utilisateurs et les coûts, il faut tenir compte du fardeau des répondants, des cordes sensibles de la population et de la capacité et de l'expertise du Bureau. Il faut porter des jugements à la

lumière des priorités stratégiques publiques actuelles afin de déterminer quels sont les programmes statistiques qui nécessitent le plus de remaniement ou un investissement nouveau ou supplémentaire.

Toutefois, les changements ou les rajustements sont assortis de contraintes. On a estimé que plus de 90 % des ressources budgétaires du Bureau sont affectées à des programmes permanents qui ne sont pas discrétionnaires à un moment donné. Ces programmes répondent aux besoins en information d'une vaste clientèle par la diffusion de renseignements de base sur la société et l'économie canadiennes et satisfont aux exigences législatives et réglementaires précisées dans environ une vingtaine de lois du Parlement.

L'interdépendance entre les différents programmes constitue une autre contrainte. Dans de nombreux cas, l'information rattachée à un programme alimente un autre programme (p. ex., l'information sur les ventes au détail sert à calculer le PIB, les statistiques de l'état civil sont utilisées dans les estimations démographiques), de sorte qu'il faut tenir compte de l'incidence que les rajustements apportés à un programme aura sur les autres programmes.

Ainsi, les besoins en information nouveaux ou émergents doivent être financés par des économies réalisées dans les programmes non discrétionnaires sans en compromettre les résultats par la réaffectation des ressources dans la composante discrétionnaire ou par l'acceptation des clients (et particulièrement des clients du gouvernement fédéral) de financer ces ajouts précieux à la base de données nationale.

3.3.2 Gestion de l'exactitude

Les processus décrits sous la rubrique Pertinence déterminent les programmes qui seront mis en œuvre, leurs objectifs généraux et les paramètres de ressources dans lesquels ils doivent être réalisés. Dans le contexte de ces « paramètres de programme », la gestion de l'exactitude nécessite une attention particulière pendant les phases de conception, de mise en œuvre et d'évaluation d'une activité statistique, qui se complètent toutes.

3.3.2.1 Conception et mise en œuvre des programmes

Le niveau d'exactitude atteint - ainsi que le degré d'actualité et de cohérence - sera fonction des méthodes explicites mises en place et des processus d'assurance de la qualité intégrés pour détecter et contrôler les erreurs éventuelles dans les diverses étapes de la mise en œuvre du programme. Il incombe aux responsables des programmes de déterminer et de justifier les décisions relatives à ce qui constitue un degré d'exactitude acceptable selon leur connaissance des besoins des utilisateurs, des circonstances, du budget et autres contraintes, ainsi que selon les possibilités et les objectifs inhérents à leur travail.

Les présentes Lignes directrices concernant la qualité décrivent des pratiques, des méthodes et des questions précises dont il faut tenir compte dans la conception des programmes et indiquent si des normes ou des lignes directrices officielles existent.

Même si la force de la méthode d'enquête sera fonction des jugements de l'équipe de conception de l'enquête, toutes les méthodes appliquées doivent faire partie du domaine des pratiques statistiques généralement acceptées et justifiables dans les circonstances données. On encourage l'utilisation des nouvelles technologies et des innovations pour améliorer la qualité et l'efficacité, mais ces dernières doivent avoir été mises à l'essai adéquatement pour réduire le risque au minimum. Il doit être possible de contrôler la qualité, de réagir efficacement aux problèmes imprévus ainsi que de pouvoir vérifier ou appuyer la crédibilité des résultats et de comprendre les limites des nouvelles technologies et des innovations.

Les résultats de la mise en œuvre ne reposent pas seulement sur la conception particulière et les outils d'enquête (p. ex., le plan d'échantillonnage et le questionnaire), mais également sur les instruments de mise en œuvre. Ceux-ci comprendront les plans de ressources et de matériel, la structure de supervision, les calendriers, les opérations, les procédures et les contrôles, la formation, la publicité, etc. établis et précisés au cours de la phase de conception. Les mécanismes de suivi de la mise en œuvre doivent être intégrés aux processus d'enquête à l'étape de la conception. Il faut examiner deux types d'information. Le premier est l'information servant à superviser et à corriger, en temps réel, les problèmes qui surviennent au cours de la mise en œuvre. Le deuxième est l'information visant à évaluer, après le coup, si le plan a été exécuté comme prévu, si certains aspects du plan ont posé des problèmes et si des enseignements ont été tirés, du point de vue opérationnel, pour faciliter les conceptions futures. L'information directement rattachée à l'exactitude ainsi que l'information liée aux coûts et à l'efficacité des opérations est tout aussi importante sur le plan de l'exactitude des conceptions futures.

3.3.2.2 Évaluation de l'exactitude

L'évaluation de l'exactitude sous-tend la détermination du niveau d'exactitude réellement atteint. Cet aspect doit être pris en compte à l'étape de la conception, car la mesure de l'exactitude nécessite souvent l'enregistrement de l'information en cours d'enquête.

Comme pour la conception, la portée et la complexité des mesures d'évaluation de l'exactitude seront fonction de la taille du programme et de ce que l'on fait des estimations. La Politique de Statistique Canada visant à informer les utilisateurs sur la qualité des données et la méthodologie (Statistique Canada, 2000d; voir l'annexe 2) exige qu'au minimum, l'on tienne compte des quatre principaux domaines d'évaluation de l'exactitude qui suivent dans tous les programmes : l'évaluation de la couverture de l'enquête; l'évaluation de l'erreur d'échantillonnage dans le cas où des échantillons ont été utilisés (les erreurs types ou les coefficients de variation doivent être indiqués pour les principales estimations); les taux de non-réponse et les estimations de l'effet de l'imputation; et les descriptions ou les mesures des autres problèmes graves d'exactitude ou de cohérence dans les résultats de l'enquête. Les mesures de l'exactitude sont également des éléments d'information importants pour l'examen des programmes afin de déterminer si les besoins des utilisateurs sont comblés et de bien utiliser les données pour les besoins de l'analyse. Ces mesures sont aussi des données d'entrée essentielles à la gestion de l'intelligibilité, décrite ci-dessous.

Compte tenu du contenu très technique de nombreux problèmes de conception, on encourage les responsables de programme à intégrer des examens techniques indépendants dans leurs plans de conception, de mise en œuvre et d'évaluation de l'exactitude. Cela peut se faire, par exemple, par le truchement d'un comité interne d'examen technique des principaux programmes, par un renvoi des problèmes liés aux normes techniques, aux méthodes ou aux approches générales au Comité des méthodes et des normes, ou par l'intervention d'un comité consultatif.

3.3.3 Gestion de l'actualité

L'actualité de l'information renvoie à l'intervalle entre le point de référence, ou la fin de la période de référence, auquel se rapporte l'information et la date à laquelle l'information est diffusée aux utilisateurs. L'information qui est accessible aux utilisateurs pendant la période au cours de laquelle elle demeure utile pour ses principaux objectifs est considérée comme étant actuelle.

L'actualité planifiée est une décision liée à la conception, souvent fondée sur des compromis entre l'exactitude et le coût. Par conséquent, l'actualité améliorée n'est pas un objectif inconditionnel. Toutefois, l'actualité est une caractéristique importante dont on doit effectuer un suivi au fil du temps afin de prévenir la détérioration et, dans l'ensemble des programmes, de reconnaître les cas extrêmes de non ponctualité et d'établir les bonnes pratiques. Pour les diffusions majeures, les dates doivent être annoncées longtemps à l'avance. En outre, le respect des dates de diffusion prévues doit faire l'objet d'un suivi et constituer une mesure du rendement de l'actualité, tout comme les changements des dates de diffusion prévues à plus long terme.

Dans le cas de certains programmes, on diffuse des données préliminaires, suivies de chiffres révisés et définitifs, comme stratégie visant à rendre les données plus actuelles. Dans de tels cas, le suivi de la quantité et de l'orientation des révisions peut servir à évaluer la pertinence du compromis actualité-exactitude choisi. Le suivi permet également de reconnaître les biais persistants ou prévisibles qu'on pourrait supprimer dans les données préliminaires à l'étape de l'estimation.

Pour les enquêtes spéciales et les nouvelles enquêtes, ainsi que pour les programmes qui offrent des services personnalisés d'extraction des données, la mesure d'actualité adéquate est le temps écoulé entre la réception de la demande et la transmission de l'information au client. Il faut établir des normes pour ces services et effectuer un suivi de leurs réalisations.

On peut s'attendre à des améliorations au chapitre de l'actualité au fur et à mesure que les nouvelles technologies font leur apparition et que les utilisations des données évoluent. Il peut être nécessaire de continuer à évaluer les pratiques actuelles afin de réaliser et d'améliorer le degré d'actualité de l'information par des évaluations opérationnelles, par l'expérimentation, par les essais et par la mesure des processus. En outre, la capacité d'informer les utilisateurs au sujet des contraintes associées à l'actualité est un aspect important de la gestion de l'actualité.

3.3.4 Gestion de l'accessibilité

L'accessibilité de l'information renvoie à la facilité avec laquelle les utilisateurs peuvent connaître l'existence de cette information, la localiser et l'importer dans leur propre environnement de travail. L'objectif de Statistique Canada en matière de diffusion est d'optimiser l'utilisation de l'information qu'il produit tout en s'assurant que les coûts de diffusion ne réduisent pas sa capacité de recueillir et de traiter les données au départ. Les politiques de diffusion du Bureau et les systèmes de livraison déterminent la plupart des aspects de l'accessibilité.

Les gestionnaires de programme sont responsables de concevoir des produits statistiques, de choisir les systèmes de livraison convenables et de veiller à ce que les produits statistiques soient bien intégrés aux systèmes de catalogage du Bureau. Lorsqu'ils déterminent les produits et les services d'information à offrir, les gestionnaires de programme doivent communiquer avec les clients, effectuer des recherches et tenir compte des demandes des clients, et effectuer un suivi de la rétroaction des clients sur le contenu et le mode de diffusion des produits. (La Division de la commercialisation offre des services pour faciliter la mise au point et l'application de ces processus.) Les gestionnaires de programme doivent également s'assurer que les produits sont conformes aux exigences des politiques et des normes contenues dans *Faits saillants des publications*, dans la Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie, dans la présentation des données et dans l'évaluation des produits d'information (Statistique Canada, 2003d).

À l'échelle du Bureau, les principaux instruments de diffusion sont les suivants : Le Quotidien pour la diffusion initiale de toutes les données; CANSIM II, le répertoire de toutes les données accessibles au public; le site Web de Statistique Canada comme principal point d'entrée pour la recherche de données; et un vaste programme de publications et de rapports analytiques pour des groupes de clients précis.

Les Services conseils offre un point d'accès unique à l'information et aux services de Statistique Canada grâce à un réseau de centres de consultation régionaux répartis à l'échelle du pays. Le programme gouvernemental des bibliothèques de dépôt permet de s'assurer que tous nos produits sont disponibles dans les bibliothèques canadiennes. L'Initiative de libération des données du Bureau permet de veiller à ce que les universités aient accès à un éventail de produits de SC à un coût raisonnable, à des fins d'enseignement et de recherche.

Les gestionnaires de programme disposent de diverses options pour rendre leurs fichiers de données plus accessibles en prévision des activités analytiques. Parmi ces options, on trouve les suivantes : la production de fichiers de microdonnées à grande diffusion qui ont été vérifiés (et approuvés par le Comité de diffusion des microdonnées) afin de protéger la confidentialité; la prestation d'un service d'extraction personnalisé; la passation de marché avec un analyste de l'extérieur en vertu de la Loi sur la statistique; le renvoi au programme des centres de données de recherche administré par le Conseil de recherches en sciences humaines du Canada.

3.3.5. Gestion de l'intelligibilité

Une des responsabilités du Bureau consiste à offrir suffisamment d'information pour permettre aux utilisateurs d'interpréter efficacement les renseignements statistiques. La gestion de l'intelligibilité porte principalement sur la communication de métadonnées ou de l'information sur l'information.

L'information nécessaire à la compréhension des données statistiques se divise en trois catégories générales :

- a) les concepts, les variables et les classifications qui sous-tendent les données;
- b) la méthodologie utilisée pour recueillir et compiler les données;
- c) les indicateurs de précision des données.

Dans le cas des fichiers de microdonnées à grande diffusion, l'information concernant le cliché d'enregistrement et le système de codage et de classification utilisé pour coder les données dans le fichier représente un outil essentiel qui permet aux utilisateurs de comprendre et de se servir des fichiers de données.

Les normes et les lignes directrices de Statistique Canada sur la mise à disposition des métadonnées sont tirées de la Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie (Statistique Canada, 2000d; voir l'annexe 2). Les gestionnaires de programme doivent s'assurer que leurs produits satisfont aux exigences de cette politique et doivent documenter leurs programmes dans la Base de métadonnées intégrée (Statistique Canada, 2000c).

En outre, les clients de Statistique Canada peuvent bénéficier d'un service d'interprétation des données au fur et à mesure qu'elles sont diffusées, grâce à un commentaire dans *Le Quotidien* et à la mise en évidence des principaux résultats dans toutes les publications statistiques, comme le prévoit la Politique sur les points saillants des publications (Statistique Canada, 1985b). La politique permet d'élucider les fausses interprétations graves de la population à l'égard des données (Statistique Canada, 1986b).

3.3.6. Gestion de la cohérence

Par cohérence des données statistiques, on entend la cohérence entre les différentes données élémentaires rattachées à un même point dans le temps, la cohérence entre les mêmes données élémentaires rattachées à différents points dans le temps et la cohérence internationale. À Statistique Canada, on utilise trois méthodes complémentaires pour gérer la cohérence.

La première méthode est l'élaboration et la mise en place de standards (p. ex., le Système de comptabilité nationale), de concepts, de variables et de systèmes de classification normalisés pour toutes les principales variables, ainsi que pour toutes les normes internationales existantes.

La deuxième méthode vise à assurer que le processus de mesure n'entraîne pas d'incohérences entre les sources de données même quand les quantités mesurées sont

définies de façon systématique, par exemple par l'utilisation d'un registre des entreprises commun servant de base à toutes les enquêtes-entreprises; par l'utilisation de questions communément formulées; par l'application de méthodes et de systèmes « harmonisés »; par l'application des Lignes directrices concernant la qualité; par l'utilisation des centres de compétence établis pour certaines méthodes et technologies; et par la consultation des codes internationaux de pratiques exemplaires.

La troisième méthode analyse les données elles-mêmes et est axée sur la comparaison et l'intégration des données tirées de sources différentes ou recueillies au fil du temps (p. ex., l'intégration des données dans les Comptes nationaux, l'étalonnage ou le calage aux marges des estimations sous-annuelles et annuelles). Ce type d'analyse tente de cerner les situations où la variation ou l'incohérence excède les niveaux prévus de la précision souhaitée des données. La rétroaction des utilisateurs externes et des analystes des données qui mettent en évidence les problèmes de cohérence des données actuelles est également une composante importante de l'analyse de la cohérence.

3.3.7 Environnement

Il va de soi que la gestion des six dimensions de la qualité s'effectue dans un environnement organisationnel. Des mesures sont en place en vue de créer un environnement et une culture dans lesquels on reconnaît l'importance de la qualité pour l'efficacité du Bureau et on fait la promotion de la qualité.

Les mesures comprennent un programme de recrutement et de perfectionnement au niveau d'entrée pour les grands groupes professionnels, ainsi qu'un cadre général de formation et de perfectionnement. Ces mesures comprennent divers instruments de communication visant à offrir de l'information aux employés et à obtenir la rétroaction de ces derniers sur la façon d'améliorer les programmes et l'environnement organisationnel. Ces instruments comportent des mesures explicites qui permettent de conclure des partenariats et des ententes avec les fournisseurs du Bureau. Une attention particulière est accordée au suivi des plaintes des répondants. On met les questionnaires à l'essai afin d'assurer qu'ils empiètent le moins possible sur la vie privée, de respecter les cordes sensibles de la population et d'obtenir l'acceptabilité sociale générale. On conclut des accords de collaboration avec les répondants qui fournissent les données en utilisant de nombreux moyens, tels qu'un programme de relations avec les répondants et un programme de gestion du fardeau de réponse.

Parmi les mesures, on trouve également des programmes d'analyse des données et de recherche méthodologique qui encouragent la quête constante de l'amélioration. L'analyse des données favorise la pertinence, l'exactitude et la cohérence des données statistiques du Bureau tout en permettant au personnel d'acquérir un plus grand nombre d'expériences et de contacts. De même, la recherche et le perfectionnement des méthodes et des outils de nature statistique, spécialisée, informatique ou opérationnelle aident à atteindre une plus grande qualité et à créer une culture d'amélioration de la qualité, en plus de réaliser des gains en efficacité.

3.4 Conclusion

Le cadre d'assurance de la qualité de Statistique Canada est constitué d'un large éventail de mécanismes et de processus qui influent sur divers niveaux des programmes du Bureau et à l'échelle de l'organisme. L'efficacité de ce cadre ne dépend pas d'un seul processus ou mécanisme, mais bien de l'effet combiné de nombreuses mesures interdépendantes qui s'appuient sur les intérêts professionnels et la motivation du personnel et qui se renforcent mutuellement avec comme objectif la satisfaction des besoins des clients. Ces mesures mettent l'accent sur le professionnalisme du Bureau et illustrent le souci rattaché à la qualité des données. Un des traits dominants de cette stratégie est la synergie qui résulte de ce que les nombreux intervenants des programmes du Bureau travaillent dans un cadre où l'on privilégie la cohérence des processus et l'uniformité des messages.

Références

- Fellegi, I. (1996). Characteristics of an effective statistical system. **International Statistical Review**, 64, 165-197
- Statistique Canada (1985b). Politique en matière des réponses aux affirmations erronées des médias. **Manuel des politiques**, 1.2.
- Statistique Canada (1986b). Politique sur les faits saillants des publications. **Manuel des politiques**, 2.1.
- Statistique Canada (2000c). Base de métadonnées intégrée: http://stdsweb/standards/imdb/imdb-menu_f.htm (Site intranet de SC).
- Statistique Canada (2000d). Politique visant à informer les utilisateurs de la qualité des données et la méthodologie. **Manuel des politiques**, 2.3. (Reproduite à l'annexe 2).
- Statistique Canada (2002c). Le cadre d'assurance de la qualité de Statistique Canada - 2002. Publication n° 12-586-XIF au catalogue, Statistique Canada.
- Statistique Canada (2003d). **Manuel des politiques de Statistique Canada**. Statistique Canada.
- Trewin, D. (2002). L'importance d'une culture de la qualité. **Techniques d'enquête**, 28, 135-145.

4. Liste d'acronymes

BMDI	Base de métadonnées intégrée
CANSIM	Système canadien d'information socio-économique
CART	Codage automatisé par reconnaissance de texte
CÉD	Collecte électronique des données
CV	Coefficient de variation
DC2	Système généralisé de collecte et de saisie des données
ÉÉD	Échange électronique des données
GQT	Gestion de la qualité totale
GREG	Estimateur de régression généralisé
IPAO	Interview sur place assistée par ordinateur
ITAO	Interview téléphonique assistée par ordinateur
SAA	Système d'appel aléatoire
RIC	Reconnaissance intelligente de caractères
SADCQ	Système d'analyse des données du contrôle qualitatif
SARTe	Système automatisé de regroupement des territoires
SC	Statistique Canada
SCANCIR	Système canadien de contrôle et d'imputation du recensement
SGE	Système généralisé d'estimation
SGÉCH	Système généralisé d'échantillonnage
SGVI	Système général de vérification et d'imputation

5. Bibliographie

- Archer, D. (1995). Maintenance of business registers. Dans **Business Survey Methods**, B.G. Cox et coll. (éd.), Wiley-Interscience, New York, 85-100.
- Bankier, M., Lachance, M. et Poirier, P. (1999). A generic implementation of the New Imputation Methodology. **Proceedings of the Survey Research Methods Section**, American Statistical Association, 548-553.
- Bankier, M.D. (1986). Estimators based on several stratified samples with applications to multiple frame surveys. **Journal of the American Statistical Association**, 81, 1074-1079.
- Bethel, J. (1989). Répartition de l'échantillon dans les enquêtes à plusieurs variables. **Techniques d'enquête**, 15, 49-60.
- Biemer, P., Groves, R.M., Mathiowetz, N.A., Lyberg, L. et Sudman, S. (éd.) (1991). **Measurement Errors in Surveys**. Wiley, New York.
- Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex Surveys. **International Statistical Review**, 51, 279-292.
- Binder, D.A., et Hidioglou, M.A. (1988). Sampling in time. **Handbook of Statistics**, P.K. Krishnaish et C.R. Rao (éd.), 187-211.
- Binder, D.A. et Roberts, G. (2001). Can informative designs be ignorable? Newsletter of the Survey Research Methods Section, American Statistical Association, Issue 12.
- Binder, D.A. et Roberts, G.R. (2003). Design-based methods for estimating model parameters. **Analysis of Survey Data**, R.L. Chambers et C.J. Skinner (éd.), Wiley, Chichester, 29-48.
- Blanc, M., Radermacher, W. et Körner, T. (2001). Quality and users. Session 15.1 of **the International Conference on Quality in Official Statistics**, Stockholm, Sweden. Voir aussi <http://method/English/research/Proceed/Q-2001/15.1.pdf> (Site intranet de SC).
- Brackstone, G. (1999). La gestion de la qualité des données dans un bureau de statistique. **Techniques d'enquête**, 25, 157-171.
- Brackstone, G. et White, P. (2002). Data stewardship at Statistics Canada. **Proceedings of the Social Statistics Section**, American Statistical Association, 284-293.
- Brackstone, G.J. (1987). Utilisation des dossiers administratifs à des fins statistiques. **Techniques d'enquête**, 13, 35-51.

- Brackstone, G.J. (1993). Data relevance: keeping pace with user needs. **Journal of Official Statistics**, 9, 49-56.
- Brick, J., Morganstein D. et Valliant R. (2000). Analysis of complex sample data using replication. Voir aussi <http://www.westat.com/wesvar/techpapers>.
- Brogan, D. (1998). Software for sample survey data, misuse of standard packages. Dans **Encyclopedia of Biostatistics**, Volume 5, P. Armitage and T. Colton (éd.), Wiley, New York, 4167-4174. Voir aussi <http://www.rti.org/sudaan/homeabout.cfm?aboutfile=whySUDAAN.cfm>.
- Burgess, M.J. et Brierly, R. (1995). A self-directed training course for monitors of CATI operations. Division de la recherche et du développement des opérations, Statistique Canada.
- Burgess, R.D. (1988). Évaluation des estimations du sous-dénombrement obtenues par la contre-vérification des dossiers du recensement du Canada. **Techniques d'enquête**, 14, 147-167.
- Chambers, R.L., Kokic P., Smith P. et Crudas M. (2000). Winsorization for identifying and treating outliers in business surveys. **Proceedings of the Second International Conference on Establishment Surveys**, juin 17-21, 2000, Buffalo, New York, 717-726.
- Chambers, R.L. et Skinner, C.J. (éd.) (2003). **Analysis of Survey Data**. Wiley, Chichester.
- Cialdini, R., Couper, M.P. et Groves, R.M. (1992). Understanding the decision to participate in a survey. **Public Opinion Quarterly**, 56, 475-495.
- Cochran, W.G. (1977). **Sampling Techniques**. Wiley, New York.
- Colledge, M.J. (1995). Frames and business registers: an overview. Dans **Business Survey Methods**, B.G. Cox et coll. (éd.), Wiley-Interscience, New York, 21-47.
- Converse, J. M. et Presser, S. (1986). **Survey Questions: Handcrafting the Standardized Questionnaire**. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-063, Sage Publications, Beverley Hills, Californie.
- Couper, M.P. et Groves, R.M. (1992). Le rôle de l'intervieweur dans la participation aux enquêtes. **Techniques d'enquête**, 18, 279-294.
- Couper, M.P., Baker, R.P., Bethlehem, J., Clark, C.Z.F., Martin, J., Nicholls II, W.L. et O'Reilly, J. (éd.) (1998). **Computer Assisted Survey Information Collection**. Wiley, New York.

- Cox, L.H. et Boruch, R.F. (1988). Record linkage, privacy and statistical policy. **Journal of Official Statistics**, 4, 3-16.
- Dagum, E.B. (1975). Seasonal factor forecasts from ARIMA models. **Proceedings of the International Institute of Statistics**, 40^e séance, communications présentées, 3, Varsovie, 206-219.
- Dagum, E.B. (1980). La méthode de désaisonnalisation X-11-ARMMI, Publication n° 12-564F au catalogue, Statistique Canada.
- Dagum, E.B. (1982). The effects of asymmetric filters on seasonal factor revisions. **Journal of the American Statistical Association**, 77, 732-738.
- Dagum, E.B. (1987). Problèmes courants sur la désaisonnalisation. **Techniques d'enquête**, 13, 71-82.
- Dagum, E.B. (1988). The X11ARIMA/88 seasonal adjustment method - foundations and user's manual. Time Series Research and Analysis Division, Rapport technique de Statistique Canada.
- Dagum, E.B. et Laniel, N. (1987). Revisions of trend-cycle estimators of moving average seasonal adjustment methods. **Journal of Business and Economic Statistics**, 5, 177-189.
- DeMaio, T.J. (éd.) (1983). Approaches to developing questionnaires. Statistical Policy Document de travail n° 10, United States Office of Management and Budget, Washington, D.C.
- Deville, J.-C. et Särndal, C.E. (1992). Calibration estimators in survey sampling. **Journal of the American Statistical Association**, 87, 376-382.
- Dielman, L. et Couper, M.P. (1995). Data quality in a CAPI survey: keying errors. **Journal of Official Statistics**, 11, 141-146.
- Doyle, P., Lane, J.I., Theeuwes, J.J.M. et Zayatz, L.V. (éd.) (2001). **Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies**. North-Holland.
- Duchesne, P. (1999). Estimateurs de calage robustes. **Techniques d'enquête**, 25, 47-60.
- Dufour, J. (1996). Qualité des données à l'Enquête sur la population active. Statistique Canada, document de travail de la Direction de la méthodologie n° HSMD-96-002E/F.
- Dufour, J., Kaushal, R., Clark, C. et Bench, J. (1995). Converting the Labour Force Survey to computer-assisted interviewing. Statistique Canada, document de travail de la Direction de la méthodologie n° HSMD-95-009E.

- Eltinge, J.L. et Yansaneh, I.S. (1997). Méthodes diagnostiques pour la construction de cellules de correction pour la non-réponse, avec application à la non-réponse aux questions sur le revenu de la U.S. Consumer Expenditure Survey. **Techniques d'enquête**, 23, 37-45.
- Estevao, V., Hidirolou, M.A. et Särndal, C.E. (1995). Methodological principles for a generalized estimation system at Statistics Canada. **Journal of Official Statistics**, 11, 181-204.
- Eurostat (1996). **Manual on Disclosure Control Methods**. Luxembourg: Office for Official Publications of the European Communities.
- Federal Committee on Statistical Methodology (2001). Measuring and reporting sources of error in surveys. Statistical Policy Working Paper 31. Voir aussi <http://www.fcsm.gov>.
- Fellegi, I. (1996). Characteristics of an effective statistical system. **International Statistical Review**, 64, 165-197
- Fellegi, I.P. et Holt, D. (1976). Systematic approach to automatic edit and imputation. **Journal of the American Statistical Association**, 71, 17-35.
- Fowler, F.J. Jr. (1995). Improving survey questions: design and evaluation. **Applied Social Research Methods Series**, 38, Sage Publications, Thousand Oaks, Californie.
- Franklin, Sarah F., Thomas, S. et Brodeur, M. (2000). Robust multivariate outlier detection using Mahalanobis' distance and modified Stahel-Donoho estimators. **Proceedings of the Second International Conference on Establishment Surveys**, juin 17-21, 2000, Buffalo, New York, 697-706.
- Fuller, W. (1987). **Measurement Error Models**. Wiley, New York.
- Fuller, W.A. (1991). Simple Estimators of the mean of skewed populations. **Statistica Sinica**, 1, 137-158.
- Fuller, W.A. (2002). Estimation par régression appliquée à l'échantillonnage. **Techniques d'enquête**, 28, 5-25.
- Fuller, W.A. et Rao, J.N.K. (2001). Un estimateur composite de régression qui s'applique à l'Enquête sur la population active du Canada. **Techniques d'enquête**, 27, 49-56.
- Gagnon, F., Gough, H. et Yeo, D. (1995). Survey of editing practices in Statistics Canada. Rapport technique de Statistique Canada.

- Gambino, J., Kennedy, B. et Singh, M.P. (2001). Estimation composite par régression pour l'Enquête sur la population active du Canada: Évaluation et application. **Techniques d'enquête**, 27, 69-79.
- Gambino, J.G., Singh, M.P., Dufour, J., Kennedy, B. et Lindeyer, J. (1998). **Méthodologie de l'Enquête sur la population active du Canada**. Publication n° 71-526-XPB au catalogue, Statistique Canada.
- Ghosh, M. et Rao, J.N.K. (1994). Small area estimation: an appraisal. **Statistical Science**, 9, 55-93.
- Gosselin, J.-F., Chinnappa, B.N., Ghangurde, P.D. et Tourigny, J. (1978). Couverture. Chapitre 2 dans **Répertoire de méthodes d'évaluation des erreurs dans les recensements et les enquêtes**, Publication n° 13-564F au catalogue, 7-10. Statistique Canada.
- Gower, A.R. (1994). Conception des questionnaires d'enquêtes-entreprises. **Techniques d'enquête**, 20, 129-141.
- Granquist, L. (1984). On the role of editing. **Statistisk tidskrift**, 2, 105-118.
- Granquist, L. et Kovar, J.G. (1997). Editing of survey data: how much is enough? Dans **Survey Measurement and Process Quality**, Lyberg et coll. (éd.), Wiley, New York, 415-435.
- Groves, R.M. (1989). **Survey Errors and Survey Costs**. Wiley, New York.
- Groves, R.M., Biemer, P., Lyberg, L., Massey, J., Nicholls, W. et Waksberg, J. (éd.) (1988). **Telephone Survey Methodology**. Wiley, New York.
- Groves, R.M., Dillman, D. A., Eltinge, J. L. et Little, R. J. A. (2002). **Survey Nonresponse**. Wiley, New York.
- Gwet, J.-P. et Lee, H., (2000). An evaluation of outlier-resistant procedures in establishment surveys. **Proceedings of the Second International Conference on Establishment Surveys**. June 17-21, 2000, Buffalo, New York, 707-716.
- Hartley, H.O. (1962). Multiple frame surveys. **Proceedings of the Social Statistics Section**, American Statistical Association, 203-206.
- Hidiroglou, M.A. (1994). Sampling and estimation for establishment surveys: stumbling blocks and progress. **Proceedings of the Section on Survey Research Methods**, American Statistical Association, 153- 162.
- Hidiroglou, M.A. (2001). L'échantillonnage double. **Techniques d'enquête**, 27, 157-169.

- Hidiroglou, M. A. et Berthelot, J.-M.(1986). Contrôle statistique et imputation dans les enquête-entreprises périodiques. **Techniques d'enquête**, 12, 79-89.
- Hidiroglou, M.A. et Laniel, N. (2001). Sampling and estimation issues for annual and sub-annual Canadian business surveys. **International Statistical Review**, 69, 487-504.
- Hidiroglou, M.A. et Särndal, C.E. (1998). Emploi des données auxiliaires dans l'échantillonnage à deux phases. **Techniques d'enquête**, 24, 11-20.
- Hidiroglou, M.A., Latouche, M.J., Armstrong, B. et Gossen, M. (1995). Improving survey information using administrative records: the case of the Canadian Employment Survey. **Proceedings of the Annual Research Conference**, U.S. Bureau of the Census, 171-197.
- Hidiroglou, M.A. et Srinath, K.P. (1981). Some estimators of population total containing large units. **Journal of the American Statistical Association**, 47, 663-685.
- Hidiroglou, M.A et Srinath, K.P. (1993). Problems associated with designing sub-annual business surveys. **Journal of Economic Statistics**, 11, 397-405.
- Holt, D. et Smith, T.M.F. (1979). Post-stratification. **Journal of the Royal Statistical Society A142**, 33-46.
- Huang, E. et Fuller, W.A. (1978). Nonnegative regression estimation for sample survey data. **Proceedings of the Social Statistics Section**, American Statistical Association, 300-303.
- Hulliger, B. (1995). Estimateur Horvitz-Thompson à l'épreuve des valeurs aberrantes. **Techniques d'enquête**, 21, 89-97.
- Hundepool, A., van de Wetering, A., de Wolf, P.-P., Giessing, S., Fischetti, M., Salazar, J.-J. et Caprara, A. (2002). τ -ARGUS user manual 2.1. Statistics Netherlands, Voorburg. Voir <http://neon.vb.cbs.nl/casc>.
- Kalton, G. et Citro, C.F. (1993). Enquêtes par panel: ajout d'une quatrième dimension. **Techniques d'enquête**, 19, 217-227.
- Kalton, G. et Kasprzyk, D. (1986). Le traitement des données d'enquête manquantes. **Techniques d'enquête**, 12, 1-17.
- Kass, G.V. (1980). An exploratory technique for investigating large quantities of categorical data. **Applied Statistics**, 29, 119-127.
- Kish, L. (1965). **Survey Sampling**. Wiley, New York.
- Korn, E.L. et Graubard, B.I. (1999). **Analysis of Health Surveys**. Wiley, New York.

- Kott, P.S. et Vogel, F.A. (1995). Multiple-frame business surveys. Dans **Business Survey Methods**, B.G. Cox et coll. (éd.), Wiley-Interscience, New York, 185-203.
- Kovacevic, M. (1999). Record linkage and statistical matching - they aren't the same! **SSC Liaison**, Vol. 13, No. 3, 24-29.
- Kovar, J.G. et Whitridge, P. (1995). Imputation of business survey data. Dans **Business Survey Methods**, B.G. Cox et coll. (éd.), Wiley, New York, 403-423.
- Kovar, J.G., MacMillan, J. et Whitridge, P. (1988). Overview and strategy for the Generalized Edit and Imputation System. Statistique Canada, document de travail de la Direction de la méthodologie n° BSMD 88-007 E/F.
- Kuiper, J. (1976). A survey and comparative analysis of various methods of seasonal adjustment. **Proceedings of the NBER/Bureau of the Census Conference on Seasonal Analysis of Economic Time Series**, A. Zellner (éd.), Washington, D.C. 59-76.
- Laniel, N. et Finlay, H. (1991). Data quality concerns with sub-annual business survey frames. **Proceedings of the Section on Survey Research Methods**, American Statistical Association, 202-207.
- Latouche, M. et Berthelot, J.-M. (1992). Use of a score function to prioritize and limit recontacts in editing business surveys. **Journal of Official Statistics**, 8, 389-400.
- Latouche, M. et Michaud, S. (1997). Concerns pertaining to weighting of longitudinal surveys. **Proceedings of the Section on Government Statistics and Section on Social Statistics**, American Statistical Association, 111-119.
- Lavallée, P. (2001). La méthode généralisée du partage des poids et le calage sur marges. Dans **Enquêtes, modèles et applications**, J.-J. Dreesbeke et L. Lebart (éd.), Dunod, Paris, 396-403.
- Lavallée, P. (2002). **Le sondage indirect**. Éditions de l'Université de Bruxelles.
- Lavallée, P. et Caron, P. (2001). Estimation par la méthode généralisée du partage des poids : Le cas du couplage d'enregistrements. **Techniques d'enquête**, 27, 171-187.
- Lee, H. (1995). Outliers in business surveys. **Business Survey Methods**, B.G. Cox et coll. (éd.), Wiley, New York, 503-526.
- Lee, H., Rancourt, E. et Särndal, C.-E. (2002). Variance estimation from survey data under single imputation. **Survey Nonresponse**, R.M. Groves et coll. (éd.), Wiley, New York, 315-328.

- Lehtonen, R. et Pahkinen, E.J. (1995). **Practical Methods for Design and Analysis of Complex Surveys**. Wiley, Chichester.
- Lemaître, G. et Dufour, J. (1987). Une méthode intégrée de pondération des personnes et des familles. **Techniques d'enquête**, 13, 211-220.
- Lessler, J.T. et Kalsbeek, W.D. (1992). **Nonsampling Errors in Surveys**. Wiley, New York.
- Lévesque, I. et Franklin, S. (2000). Pondérations longitudinale et transversale de l'Enquête sur la dynamique du travail et du revenu, année de référence 1997. Statistique Canada, document de travail n°, 75F0002MIF - 00004.
- Linacre, S. J. et Trewin, D. J. (1989). Evaluation of errors and appropriate resource allocation in economic collections. **Proceedings of the Annual Research Conference**, U.S. Bureau of the Census, 197-209.
- Lohr, S. (1999). **Sampling: Design and Analysis**. Duxbury Press.
- Lothian, J. et Morry, M. (1977). The problem of aggregation: direct or indirect seasonal adjustment. Document de recherche n° 77-08-001E, Seasonal Adjustment and Time Series Staff, Statistique Canada.
- Lothian, J. et Morry, M. (1978). A test for the presence of identifiable seasonality when using the X-11 program. Document de recherche n° 78-10-002E, Seasonal Adjustment and Time Series Staff, Statistique Canada.
- Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N. et Trewin, D. (éd.) (1997). **Survey Measurement and Process Quality**, Wiley, New York.
- Massey, J.T. (1988). An overview of telephone coverage. Dan **Telephone Survey Methodology**, R.M. Groves et coll. (éd.), Wiley, New York, 3-8.
- Michaud, S., Dolson, D., Adams, D. et Renaud, M. (1995). Combining administrative and survey data to reduce respondent burden in longitudinal surveys. **Proceedings of the Section on Survey Research Methods**, American Statistical Association, 11-20.
- Monty, A. et Finlay, H. (1994). Strengths and weaknesses of administrative data sources: experiences of the Canadian Business Register. **Statistical Journal of the United Nations**, ECE 11, 205-210.
- Morry, M. (1992). Comparison of revisions in seasonally adjusted external trade series obtained through the use of concurrent versus forecast factors. Statistique Canada, document de travail n°, TSRA-92-004E.

- Mudryk, W. et Xiao, P. (1996). Quality control methodology for LFS industry and occupation coding operations. Rapport technique de Statistique Canada.
- Mudryk, W. et Xie, H. (2002). Quality control application in ICR data capture for the 2001 Census of Agriculture. **Proceedings of the Section on Quality and Productivity**, American Statistical Association, 2424-2429.
- Mudryk, W., Burgess, M.J. et Xiao, P. (1996). Quality control of CATI operations in Statistics Canada. **Proceedings of the Section on Survey Research Methods**, American Statistical Association, 150-159.
- Mudryk, W., Bougie, B. et Xie, H. (2002). Some guidelines for data analysis in quality control. Rapport technique de Statistique Canada.
- Mudryk, W., Croal, J. et Bougie, B. (1994). Système généralisé de collecte et de saisie des données (DC2): version 2.5.1, Comparaison des échantillons (CÉ). Rapport technique de Statistique Canada.
- Nations Unies (1964). Recommendations for the preparation of sample survey reports (édition provisoire). **Statistical Series C**, 1. New York.
- Nations Unies, Conférence des statisticiens européens (1983). **Draft Guidelines for the Preparation of Presentations of the Scope and Quality of Statistics for Users**. Genève, Suisse.
- Patak, Z. (1990). Robust principal component analysis via projection pursuit. Master's thesis, University of British Columbia, Canada.
- Platek, R., Pierre-Pierre, F.K. et Stevens, P. (1985). **Élaboration et conception des questionnaires d'enquête**. Publication n° 12-519F au catalogue. Statistique Canada.
- Platek, R., Rao, J.N.K., Särndal, C.E. et Singh, M.P. (éd.) (1987). **Small Area Statistics**. Wiley, New York.
- Pollock, K.H., Turner, S.C. et Brown, C.A. (1994). Techniques de saisie-ressaisie pour l'estimation de la taille de la population et de totaux de population lorsqu'on ne dispose pas d'une base de sondage complète. **Techniques d'enquête**, 20, 121-128.
- Rancourt, E., Lee, H. et Särndal, C.E. (1993). Variance estimation under more than one imputation method. **Proceedings of the International Conference on Establishment Surveys**, American Statistical Association, 374-379.
- Rao, J.N.K. (1996). On the estimation with imputed survey data. **Journal of the American Statistical Association**, 91, 499-506

- Rao, J.N.K. (1999). Quelques progrès récents concernant l'estimation régionale fondée sur un modèle. **Techniques d'enquête**, 25, 199-212.
- Särndal, C.E., Swensson, B. et Wretman, J.H. (1992). **Model Assisted Survey Sampling**. Springer-Verlag, New-York.
- Satin, A. et Shastry, W. (1993). Éléments d'un plan de sondage. Chapitre 2, dans **L'échantillonnage : un guide non mathématique - deuxième édition**, 7-10, Publication n° 12-602-XPF au catalogue, Statistique Canada.
- Scheuren, F. et Winkler, W.E. (1993). Analyse de régression de fichiers de données couplés par ordinateur. **Techniques d'enquête**, 19, 45-65.
- Schnell, D., Kennedy, W.J., Sullivan, G, Park, H.J. et Fuller, W.A. (1988). Logiciel d'ordinateur personnel pour l'estimation de la variance dans des enquêtes complexes. **Techniques d'enquête**, 14, 63-73.
- Schuman, H. et Presser, S. (1996). **Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context**. Sage Publications, Thousand Oaks, California.
- Shah, B.V., Barnwell, B.G. et Bieler, G.S. (1997). **SUDAAN User's Manual Release 7.5**. Research Triangle Institute, Caroline du Nord. Voir aussi <http://www.rti.org/sudaan/home.cfm>.
- Shiskin, J., Young, A.H. et Musgrave, J.C. (1967). The X-11 variant of the census method II seasonal adjustment. Technical Paper No. 15, Bureau of the Census, U.S. Department of Commerce.
- Singh, A.C. et Mohl, C. (1996). Comprendre les estimateurs de calage dans les enquêtes par échantillonnage. **Techniques d'enquête**, 22, 107-116.
- Singh, A.C., Kennedy, B. et Wu, S. (2001). Estimation composite par régression pour l'Enquête sur la population active du Canada avec plan de sondage à renouvellement de panel. **Techniques d'enquête**, 27, 35-48.
- Singh, M.P., Gambino, J. et Mantel, H. (1994). Les petites régions: problèmes et solutions. **Techniques d'enquête**, 20, 3-23.
- Singh, M.P., Hidioglou, M.A., Gambino, J. et Kovacevic, M. (2001). Estimation methods and related systems at Statistics Canada. **International Statistical Review**, 69, 461-486.
- Sirken, M.G. et Casady, R.J. (1988). Sampling variance and nonresponse rates in dual frame, mixed mode surveys. Chapitre 11 dans **Telephone Survey Methodology**,

- P. Biemer, R.M. Groves, L. Lyberg, J. Massey, W. Nicholls et J. Waksberg (éd). Wiley, New York, 175-188.
- Skinner, C.K., Holt, D. et Smith, T.M.F. (1989). **Analysis of Complex Surveys**. Wiley, Chichester.
- STATA Corporation. (1997). **STATA User's Guide**. STATA Press, College Station, TX. Voir <http://www.fas.harvard.edu/~stats/survey-soft/iass.html#stata>.
- Statistique Canada (1970). **Loi sur la Statistique**. Ottawa, Canada.
- Statistique Canada (1985a). Politique d'édition. **Manuel des politiques**, 3.1.
- Statistique Canada (1985b). Politique en matière des réponses aux affirmations erronées des médias. **Manuel des politiques**, 1.2.
- Statistique Canada (1985c). Politique en matière de diffusion de logiciels. **Manuel des politiques**, 3.5.
- Statistique Canada (1986a). Politique concernant la distribution à titre gracieux aux médias. **Manuel des politiques**, 3.2.
- Statistique Canada (1986b). Politique sur les faits saillants des publications. **Manuel des politiques**, 2.1.
- Statistique Canada (1986c). Politique sur la distribution de produits aux contacts provinciaux et territoriaux. **Manuel des politiques**, 3.4.
- Statistique Canada (1987). Politique sur la diffusion des microdonnées. **Manuel des politiques**, 4.2.
- Statistique Canada (1988). Politique sur le marketing. **Manuel des politiques**, 3.7.
- Statistique Canada (1989). **Guide de l'utilisateur pour le Système de comptabilité nationale**. Publication n° 13-589 au catalogue, Statistique Canada.
- Statistique Canada (1992a). Information - politique d'octroi de licences. **Manuel des politiques**, 3.8.
- Statistique Canada (1992b). Politique sur les droits et les frais des produits et services. **Manuel des politiques**, 3.6.
- Statistique Canada (1993a). Politique relative à la révélation discrétionnaire. **Manuel des politiques**, 4.3.
- Statistique Canada (1993b). Politique sur le Quotidien de Statistique Canada (diffusion officielle). **Manuel des politiques**, 3.3.

- Statistique Canada (1994a). Politique de rabais. **Manuel des politiques**, 3.9.
- Statistique Canada (1994b). Politique sur l'enregistrement des produits et des services d'information. **Manuel des politiques**, 3.10.
- Statistique Canada (1995). Politique concernant l'évaluation des produits d'information. **Manuel des politiques**, 2.5.
- Statistique Canada (1996a). Politique relative au couplage d'enregistrements. **Manuel des politiques**, 4.1.
- Statistique Canada (1996b). **Symposium 1996 : Erreurs non dues à l'échantillonnage**.
- Statistique Canada (1998a). Politique d'information des répondants aux enquêtes. **Manuel de politiques**, 1.1.
- Statistique Canada (1998b). Politique concernant les normes. **Manuel de politiques**, 2.10.
- Statistique Canada (2000a). Description des fonctions du système généralisé de vérification et d'imputation. Rapport technique de Statistique Canada.
- Statistique Canada (2000b). Lignes directrices pour la désaisonnalisation et estimation de la tendance-cycle. Document publié par Comité des méthodes et des normes, mars, 2000.
- Statistique Canada (2000c). Base de métadonnées intégrée: http://stdsweb/standards/imdb/imdb-menu_f.htm (Site intranet de SC).
- Statistique Canada (2000d). Politique visant à informer les utilisateurs de la qualité des données et la méthodologie. **Manuel des politiques**, 2.3. (Reproduite à l'annexe 2).
- Statistique Canada (2001a). Lignes directrices sur la rédaction d'articles d'analyse. Division des communications.
- Statistique Canada (2001b). **Classification nationale des professions pour statistiques (CNP-S) 2001**. Publication n° 12-583-XPF au catalogue, Statistique Canada.
- Statistique Canada (2001c). **Classification type des biens 2001** (basée sur le Système harmonisé de désignation et de codification des marchandises). Publication n° 12-580 au catalogue, Statistique Canada.
- Statistique Canada (2001d). Normes et lignes directrices pour la déclaration des taux de non-réponse. Rapport technique de Statistique Canada.

- Statistique Canada (2001e). **Symposium 2001 : La qualité des données d'un organisme statistique : une perspective méthodologique.**
- Statistique Canada (2002a). Politique concernant l'examen et la mise à l'essai des questionnaires. **Manuel des politiques**, 2.8. (Reproduite à l'annexe 1).
- Statistique Canada (2002b). **Classification géographique type (CGT) 2001.** Publication n° 12-571-XPB au catalogue, Statistique Canada
- Statistique Canada (2002c). Le cadre d'assurance de la qualité de Statistique Canada - 2002. Publication n° 12-586-XIF au catalogue, Statistique Canada.
- Statistique Canada (2002d). Manuel de l'utilisateur - Système Généralisé d'Analyse et de Protection de la Confidentialité. Division de développement de systèmes.
- Statistiques Canada (2003a). **Analysis Handbook.** Préparée par le Centre de ressources en analyse de données, Direction de la méthodologie.
- Statistique Canada (2003b). Lignes directrices sur l'utilisation de critères d'exactitude des données pour la diffusion de l'information statistique. Document publié par le Comité des méthodes et des normes, juin 2003.
- Statistique Canada (2003c). **Système de classification des industries de l'Amérique du Nord, Canada 2002.** Publication n° 12-501-XPB au catalogue, Statistique Canada.
- Statistique Canada (2003d). **Manuel des politiques de Statistique Canada.** Statistique Canada.
- Statistique Canada. (2003e). Le guide officiel de rédaction des Services de rédaction. Services de rédaction, Division des communications. Voir http://icn-rci.statcan.ca/10/10d/10d_000_f.htm (Site intranet de SC). Publication mise à jour régulièrement.
- Stukel, D., Hidioglou, M.A., et Särndal, C.-E. (1996). Estimation de la variance des estimateurs de calage: comparaison des méthodes du jackknife et de la linéarisation de Taylor. **Techniques d'enquête**, 22, 117-126.
- Subcommittee on Disclosure Limitation Methodology, Federal Committee on Statistical Methodology (1994). Report on statistical disclosure limitation methodology. Statistical Policy. Document de travail n° 22, Office of Management and Budget, Washington, DC.
- Sudman, S. et Bradburn, N.M. (1982). **Asking Questions: A Practical Guide to Questionnaire Design.** Jossey-Bass, San Francisco.

- Swain, L. et Dolson, D. (1997). Current issues in household survey nonresponse at Statistics Canada. **Statistics in Transition**, 3, 439-468.
- Swain, L., Drew, J.D., Lafrance, B et Lance, K. (1992). La création d'un registre des adresses résidentielles pour améliorer la couverture du recensement du Canada de 1991. **Techniques d'enquête**, 18, 139-155.
- Tambay, J. L., Schiopu-Kratina, I., Mayda, J., Stukel, D. et Nadon, S. (1998). Traitement de la non-réponse du cycle deux de l'enquête nationale sur la santé de la population. **Techniques d'enquête**, 24, 159-169.
- Tanur, J.M. (éd.) (1992). **Questions about Questions: Inquiries into the Cognitive Bases of Surveys**. Russell Sage Foundation, New York.
- Thomas, D.R. (1993). Inference using complex data from surveys and experiments. **Canadian Psychology**, 34, 415-431.
- Thompson, M.E. (1997). **Theory of Sample Surveys**. Chapman and Hall.
- Tillé, Y. (2001). **Théorie des sondages – Échantillonnage et estimation en populations finies**. Dunod, Paris.
- Trewin, D. (2002). L'importance d'une culture de la qualité. **Techniques d'enquête**, 28, 135-145.
- Valliant, R., Dorfman, A.H., et Royall, R.M. (2000). **Finite Population Sampling and Inference**. Wiley, New York.
- Willenborg, L. et de Waal, T. (1996). **Statistical Disclosure Control in Practice**. Lecture Notes in Statistics, Springer-Verlag, New York.
- Willenborg, L. et de Waal, T. (2001). **Elements of Statistical Disclosure Control**. Lecture Notes in Statistics, Springer-Verlag, New York.
- Williams, K., Denyes, C., March, M. et Mudryk, W. (1996). Mesure de la qualité dans le traitement des données d'enquête. **Recueil du Symposium 96 : Erreurs non dues à l'échantillonnage**, Statistique Canada, 131-141.
- Wolfson, M., Gribble, S., Bordt, M., Murphy, B. et Rowe, G. (1987). La base de données de simulation de politique sociale : un exemple d'intégration de données d'enquêtes et de données administratives. **Recueil du Symposium international sur les utilisations statistiques des données administratives**, Statistique Canada, 233-266.
- Yung, W. et Rao, J.N.K. (1996). Linéarisation des estimateurs de variance jackknife dans un échantillonnage stratifié à degrés multiples. **Techniques d'enquête**, 22, 23-31.

6. Annexes

6.1 Annexe 1: Politique concernant l'examen et la mise à l'essai des questionnaires

Révisée 25 septembre 2002

Contexte

Les questionnaires jouent un rôle de premier plan dans le processus de collecte des données. Ils ont une incidence considérable sur la qualité des données, notamment sur l'exactitude des réponses, et sur l'image de Statistique Canada auprès du public. Ces effets sont particulièrement importants dans le cas des programmes d'enquêtes permanentes et longitudinales.

Un questionnaire bien conçu devrait permettre de réduire le fardeau de réponse des répondants et de recueillir efficacement les données et ce, avec un minimum d'erreurs de réponse et de non-réponse. En outre, un questionnaire bien conçu devrait faciliter le codage et la saisie des données. Il devrait réduire au minimum les opérations de vérification et d'imputation requises et permettre une diminution globale des coûts et du temps associés à la collecte et au traitement des données.

Statistique Canada cherche à s'assurer que les questionnaires recueillent des renseignements exacts et qu'ils sont adéquatement testés, faciles à remplir, faciles à traiter et conviviaux, tant pour les répondants que pour les intervieweurs. L'examen et la mise à l'essai des questionnaires sont essentiels pour la détection des sources potentielles d'erreurs de réponse et de non-réponse. La présente politique impose donc l'examen et la mise à l'essai des nouveaux questionnaires et des questionnaires révisés. Elle prévoit également l'examen périodique des questionnaires utilisés dans les enquêtes permanentes et longitudinales. La politique vise tous les aspects de la conception des questionnaires susceptibles d'influencer la qualité des données, le comportement des répondants et le rendement des intervieweurs. La présente politique favorise aussi, le cas échéant, le recours à des formulations types pour les questions fréquemment posées dans diverses enquêtes, de façon à utiliser des questions bien testées et à promouvoir l'uniformité.

Définition

Aux fins de la présente politique, le terme « questionnaire » désigne un ensemble de questions destinées à recueillir des renseignements auprès d'un répondant. Le questionnaire peut être rempli par l'intervieweur ou par le répondant lui-même.

Domaine d'application

La politique vise tous les questionnaires que Statistique Canada fait remplir à des répondants extérieurs, notamment les questionnaires utilisés dans les programmes de collecte de statistiques et de renseignements de même que dans d'autres programmes, comme les études de marché et les évaluations de programme. La politique couvre la

collecte de données par les méthodes « papier et crayon » et par d'autres méthodes comme les interviews assistées par ordinateur et la collecte de données par Internet. La politique vise également les révisions apportées à un questionnaire résultant de la modification de la méthode de collecte des données (par exemple, lors du passage d'une méthode « papier et crayon » à l'interview assistée par ordinateur ou à la collecte de données par Internet).

Énoncé de politique

La politique de Statistique Canada est énoncée comme suit :

1. Tous les nouveaux questionnaires et toutes les révisions aux questionnaires existants doivent être testés dans les deux langues officielles avant leur mise en œuvre.
2. Les secteurs de programme doivent consulter le Centre d'information sur la conception des questionnaires quant aux plans de mise à l'essai des nouveaux questionnaires et des questionnaires révisés. Les lignes directrices ci-jointes donnent un aperçu des méthodes de mise à l'essai qu'il est possible d'utiliser.
3. Tous les nouveaux questionnaires et questionnaires révisés doivent être examinés par le Centre d'information sur la conception des questionnaires avant la mise à l'essai et, de nouveau, après les révisions résultant de la mise à l'essai.
4. Tous les questionnaires employés dans les enquêtes permanentes et longitudinales doivent faire l'objet d'une mise à l'essai périodique.

Responsabilité

1. Secteurs de programme

- Veiller à ce que toutes les conditions de mise en œuvre de la présente politique soient réunies, y compris l'obtention d'un budget suffisant et l'affectation d'une période de temps adéquate pour l'examen et la mise à l'essai des nouveaux questionnaires ou des questionnaires révisés.
- Consulter le Centre d'information sur la conception des questionnaires au sujet des plans de mise à l'essai des nouveaux questionnaires et des questionnaires révisés.
- Soumettre tous les nouveaux questionnaires et questionnaires révisés au Centre d'information sur la conception des questionnaires à des fins d'examen avant la mise à l'essai et, de nouveau, après les révisions résultant de la mise à l'essai.

- Consulter la Division des normes pour assurer la conformité aux normes et l'utilisation de formulations communes de questions lorsque celles-ci existent; consulter au besoin la Division des communications pour ce qui est des relations avec les répondants et la Division de la diffusion en ce qui a trait à la conception et à la production des formules.
- Soumettre tous les nouveaux questionnaires et questionnaires révisés à la Division des services d'accès et de contrôle des données.
- Soumettre les versions définitives de tous les nouveaux questionnaires et questionnaires révisés à la Division des normes pour qu'elles soient intégrées à l'« Index des enquêtes et questionnaires de Statistique Canada ».

2. Division des services d'accès et de contrôle des données

- Veiller à ce que les questionnaires soient conformes à la Politique d'information des répondants aux enquêtes.
- Informer, au besoin, les secteurs de programme quant aux exigences de cette politique.
- Veiller à ce que l'autorisation du statisticien en chef relativement à la collecte de renseignements ait été obtenue.

3. Centre d'information sur la conception des questionnaires (Direction de la méthodologie)

- Offrir des services d'examen et de consultation visant la conception des questionnaires.
- Offrir de l'aide et une orientation relativement à l'élaboration et à la mise à l'essai des nouveaux questionnaires et des questionnaires révisés. Ces services comprennent : (a) la mise à l'essai des questionnaires à l'aide de diverses méthodes comme les groupes de discussion, les interviews en profondeur, les interviews cognitives et la mise à l'essai officieuse et (b) la coordination des activités d'élaboration et de mise à l'essai faisant l'objet d'une impartition.
- Examiner les questionnaires et porter attention à la formulation et à l'enchaînement des questions, à la longueur, au format et à la convivialité tant pour le répondant que pour l'intervieweur.
- Fournir de la documentation, des lignes directrices et de la formation pour améliorer la conception des questionnaires.

4. Division de la diffusion

- Fournir de l'aide pour ce qui est de la conception, du format, de la composition et de l'impression des questionnaires et des formules d'enquête.
- Veiller à ce que le questionnaire soit conforme aux dispositions du Programme de coordination de l'image de marque.

5. Comité des méthodes et des normes

- Surveiller la mise en œuvre de la présente politique.
- Approuver toutes les lignes directrices concernant la conception des questionnaires énoncées en vertu de la présente politique.
- Promouvoir, le cas échéant, l'utilisation de formulations communes de questions dans les diverses enquêtes.

6.2 **Annexe 2: Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie**

Approuvée le 31 mars 2000

Introduction

En sa qualité d'organisme professionnel responsable de la production de données officielles, Statistique Canada se doit de faire connaître aux utilisateurs les concepts et la méthodologie utilisés dans la collecte, le traitement et l'analyse de ses données, la précision des données et toutes autres caractéristiques concernant la qualité des données ou l'adéquation des données à leur utilisation.

Les utilisateurs doivent d'abord être en mesure de vérifier si le cadre conceptuel et les définitions qui répondraient à leurs besoins particuliers en matière de données et à l'utilisation qu'ils comptent faire de celles-ci sont identiques ou suffisamment semblables à ceux employés aux fins de la collecte et du traitement des données. Ils doivent alors pouvoir évaluer dans quelle mesure la précision des données et les autres aspects de la qualité correspondent à l'utilisation ou à l'interprétation prévue.

L'évaluation de la qualité des données et de l'adéquation des données à leur utilisation constitue une tâche complexe compte tenu de la nature pluridimensionnelle du concept de la qualité. La gamme complète des utilisations éventuelles des données ne peut pas toujours être prévue et on ne peut évaluer tous les aspects de la qualité dans chaque contexte d'utilisation. Plus particulièrement, les données sont exposées à de nombreuses sources d'erreur et, dans l'état actuel de nos connaissances, une mesure exhaustive de la précision des données est rarement possible. Les mesures de précision fournies aux utilisateurs comporte donc des limites bien nettes et il est impossible d'élaborer une norme stricte relativement à l'évaluation de la qualité des données pour l'ensemble des produits du Bureau. Il faut plutôt insister sur la description et la quantification des principales caractéristiques de qualité des données.

Politique

1. Statistique Canada fournira aux utilisateurs des indicateurs de la qualité des données diffusées et une description de la méthodologie et des concepts sous-jacents.
2. Il faut joindre aux produits statistiques la documentation relative à la qualité et à la méthodologie ou la citer en référence.
3. La documentation relative à la qualité et à la méthodologie doit être conforme aux normes et lignes directrices émises, à intervalles irréguliers, aux termes de la présente politique.

4. Dans des circonstances spéciales, il est possible d'obtenir l'autorisation de se soustraire aux exigences de la présente politique en procédant de la façon dont il est fait état au paragraphe ci-dessous traitant des responsabilités.
5. Il faut faire connaître aux promoteurs d'enquêtes à fonds recouvrables et de projets de consultation statistique à l'issue desquels Statistique Canada ne publiera aucune donnée, les sections pertinentes des normes et lignes directrices émises en vertu de la présente politique et les inciter à les respecter.

Portée

La présente politique s'applique à toutes les données et à tous les résultats analytiques diffusés par Statistique Canada, quelle que soit la méthode de collecte, de calcul ou d'intégration utilisée, abstraction faite du mode de diffusion ou de la source de financement.

Responsabilités

Les directeurs des secteurs de programme ont la responsabilité:

- d'aviser les utilisateurs de la possibilité d'obtenir des renseignements sur la qualité des données et la méthodologie employée;
- de diffuser les mesures ou les descriptions existantes relatives à la qualité des données de même que la documentation sur la méthodologie;
- de s'assurer que des mécanismes ont été ou sont élaborés et mis en œuvre en vue de fournir les renseignements sur la qualité des données nécessaires aux termes de la présente politique;
- de rédiger de la documentation sur la méthodologie qui réponde aux critères de la présente politique;
- de tenir compte des critères de la présente politique en ce qui a trait à la conception, à l'établissement du calendrier et du budget d'activités, de programmes ou de produits nouveaux ou remaniés; et
- de soumettre au Comité des méthodes et des normes les demandes visant à obtenir l'autorisation de se soustraire aux exigences de la présente politique.

Le Comité des méthodes et des normes a la responsabilité :

- de rédiger des rapports périodiques sur l'application de la présente politique;
- d'évaluer périodiquement la mise en œuvre de la présente politique dans certains secteurs de programme et de s'assurer que ces évaluations coïncident avec celles

du programme en question;

- de fournir des normes et des lignes directrices relatives à la mise en œuvre de la politique dans les secteurs de programme;
- de revoir au besoin la politique de même que les normes et les lignes directrices qui s'y rapportent; et
- de réviser et d'approuver les demandes d'autorisation de se soustraire aux exigences de la présente politique.

Demande de renseignements

Toute demande de renseignements ayant trait à l'interprétation de la présente politique doit être adressée au président du Comité des méthodes et des normes.