



12-539-XIF

Statistique Canada Lignes directrices concernant la qualité

Troisième édition - Octobre 1998



Statistique Canada
Statistics Canada

Canada

Des données sous plusieurs formes

Statistique Canada diffuse les données sous formes diverses. Outre les publications, des totalisations habituelles et spéciales sont offertes. Les données sont disponibles sur Internet, disque compact, disquette, imprimé d'ordinateur, microfiche et microfilm, et bande magnétique. Des cartes et d'autres documents de référence géographiques sont disponibles pour certaines sortes de données. L'accès direct à des données agrégées est possible par le truchement de CANSIM, la base de données ordinolingué et le système d'extraction de Statistique Canada.

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 : Direction de la méthodologie, Statistique Canada, Ottawa, Ontario, K1A 0T6 (téléphone : (613) 951-8615), fax : (613) 951-5711, e-mail : kovar@statcan.ca ou à l'un des centres de consultation régionaux de Statistique Canada :

Halifax	(902) 426-5331	Regina	(306) 780-5405
Montréal	(514) 283-5725	Edmonton	(403) 495-3027
Ottawa	(613) 951-8116	Calgary	(403) 292-6717
Toronto	(416) 973-6586	Vancouver	(604) 666-3691
Winnipeg	(204) 983-4020		

Vous pouvez également visiter notre site sur le Web : <http://www.statcan.ca>

Un service d'appel interurbain sans frais est offert **à tous les utilisateurs qui habitent à l'extérieur des zones de communication locale** des centres de consultation régionaux.

Service national de renseignements	1 800 263-1136
Service national d'appareils de télécommunications pour les malentendants	1 800 363-7629
Numéro pour commander seulement (Canada et États-Unis)	1 800 267-6677

Abonnement

Veillez commander par la poste, en écrivant à Statistique Canada, Division de la diffusion, Gestion de la circulation, 120, avenue Parkdale, Ottawa, (Ontario) K1A 0T6; par téléphone en composant le (613) 951-7277 ou le 1 800 770-1033; par télécopieur, en composant le (613) 951-1584 ou le 1800 889-9734; ou par Internet : order@statcan.ca. Lorsque vous signalez un changement d'adresse, veuillez nous fournir l'ancienne et la nouvelle adresses. On peut aussi se procurer les produits de Statistique Canada auprès des agents autorisés, dans les librairies et dans les bureaux régionaux de Statistique Canada.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois et 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 le centre de consultation régional de Statistique Canada le plus près de chez vous.



Statistique Canada
Direction de la méthodologie

Statistique Canada

Lignes directrices concernant la qualité

Troisième édition - Octobre 1998

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 1998

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.

Novembre 1998

N° 12-539-XIF au catalogue

Périodicité : trimestrielle

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

Certaines professions codifient leurs pratiques de façon précise ou les définissent en fonction de codes et de normes professionnels. Cependant, il n'existe pas de code aussi précis qui pourrait régir les techniques d'enquête. Même si les techniques d'enquête sont une collection de pratiques étayées par des notions théoriques qui s'offrent aux praticiens, il appartient à ces derniers de les choisir avec soin, en vue de répondre à leurs besoins particuliers. Ces choix doivent être faits de manière à mettre en équilibre les objectifs de qualité, de pertinence, d'échéance, de coût et de fardeau de réponse qui, souvent, entrent en conflit.

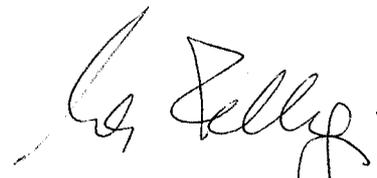
Le présent document regroupe un ensemble d'expériences et de conclusions; elles ont été évaluées dans leur propre contexte pour constituer un ensemble de "bonnes pratiques". Ces "bonnes pratiques" ne s'appliquent pas nécessairement à toutes les activités d'acquisition de données. De plus, même si une ligne directrice spécifique s'applique en théorie à une enquête donnée, des circonstances particulières peuvent être telles qu'un bon jugement professionnel lui donnerait peu d'importance, et donc peu de considération par rapport à d'autres qui, selon le contexte, seraient plus importantes. Ce document doit donc être utilisé avec une grande attention professionnelle et avec jugement.

Même si les lignes directrices présentées dans les pages qui suivent ne remplacent pas l'expertise fondée sur la connaissance et le jugement du personnel responsable de la conception des enquêtes, l'importance accordée à la qualité doit être présente dans toutes nos activités. Il appartient à tous les gestionnaires de programme statistique de voir à ce que l'importance accordée à la qualité à Statistique Canada soit considérée lors de l'élaboration et l'utilisation des méthodes statistiques qui sont de son ressort.

La première édition des Lignes directrices concernant la qualité remonte à 1985. La deuxième édition, qui comprenait de légères modifications par rapport à la précédente, est parue en 1987. Depuis, de nombreux progrès importants ont été réalisés au chapitre des techniques d'enquête; par exemple, dans l'intégration de la technologie informatique et des communications électroniques à la collecte et au traitement des données. Cela s'est traduit par des changements dans l'enchaînement des opérations des enquêtes ainsi que par de nouvelles approches en matière de stockage et de diffusion des données. Ce sont de tels progrès dans les techniques d'enquêtes qui ont incité à mettre à jour, de façon importante, l'édition précédente.

Je tiens à remercier les nombreux experts de Statistique Canada qui ont participé à la rédaction de la troisième édition des Lignes directrices concernant la qualité. Les conseils du Comité des méthodes et des normes, ainsi que les commentaires du Comité consultatif des méthodes statistiques, ont contribué à améliorer le document.

Le statisticien en chef,



Ivan P. Fellegi

TABLE DES MATIÈRES

PRÉFACE	2
TABLE DES MATIÈRES	3
1. INTRODUCTION	4
2. ÉTAPES DE L'ENQUÊTE	8
2.1 Objectifs, utilisations et utilisateurs	10
2.2 Concepts, définitions et classifications	13
2.3 Couverture et bases de sondage	17
2.4 Échantillonnage	21
2.5 Élaboration du questionnaire	25
2.6 Réponse et non-réponse	29
2.7 Opérations de collecte de données	33
2.8 Vérification	37
2.9 Imputation	41
2.10 Estimation	45
2.11 Désaisonnalisation et estimation de la tendance-cycle	50
2.12 Évaluation de la qualité des données	55
2.13 Contrôle de la divulgation	59
2.14 Diffusion des données	64
2.15 Analyse (et présentation) des données	67
2.16 Documentation	71
2.17 Utilisation des données administratives	74
3. CONTEXTE DE GESTION	80
3.1 La qualité à Statistique Canada	80
3.2 Organisation et infrastructure de la gestion de la qualité	81
3.3 Cadre d'assurance de la qualité	82
3.4 Conclusion	86
4. BIBLIOGRAPHIE	87
5. ANNEXES	98
5.1 Annexe 1: Politique sur l'élaboration, l'essai et l'évaluation des questionnaires ..	98
5.2 Annexe 2: Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie	102
6. INDEX DES TERMES IMPORTANTS	104

1. INTRODUCTION

Le présent document réunit les lignes directrices et les questions dont il faut tenir compte afin d'atteindre les objectifs de qualité dans le cadre de la réalisation des activités statistiques. Il tire parti de l'expérience collective de nombreux employés de Statistique Canada. Il devrait être utile au personnel affecté à la planification et à la conception des enquêtes ainsi qu'aux employés chargés d'analyser et d'évaluer les résultats.

Depuis la publication des versions antérieures du présent document, le mot *qualité* a pris un sens plus large, en grande partie dans la foulée de la *gestion de la qualité totale* (GQT). Par « qualité », nous entendons l'adaptation des statistiques à leur utilisation. Cela comprend non seulement les concepts de qualité statistique de la variance et du biais, mais également la pertinence (c'est-à-dire mesurer ce qui doit être mesuré et diffuser les résultats dans des délais acceptables, de manière qu'ils soient encore utiles).

Bien qu'elle soit générale, cette définition est plus restreinte que celle inhérente à la GQT. Afin de concilier ces deux définitions, il faut voir ce que la GQT signifie pour un organisme de statistique. En simplifiant à l'extrême, on peut dire que la GQT semble être constituée de trois éléments : cerner et comprendre les besoins du client (l'utilisateur); amener les employés à participer à la prise des décisions visant à satisfaire ces besoins; revoir constamment les méthodes administratives en vue d'une éventuelle refonte. Ces trois activités permettront à l'organisme de statistique, comme à toute autre organisation, d'améliorer la qualité. Les Lignes directrices concernant la qualité sont liées à ces trois éléments et viennent étayer les efforts que Statistique Canada déploie depuis longtemps afin d'assurer la qualité de ses produits et de ses services. Dans les paragraphes qui suivent, nous approfondirons les diverses dimensions de la qualité auxquelles l'organisme de statistique doit s'intéresser (Statistique Canada, 1997a).

La qualité des données doit être définie et assurée en fonction de l'adaptation des données à leur utilisation. Cette adaptation des données et de l'information statistique à leur utilisation dépend d'une part des utilisations prévues et des caractéristiques intrinsèques de la qualité et, d'autre part, des attentes des utilisateurs relativement à ces caractéristiques.

Les organismes de statistique n'ont pas adopté une définition normalisée de la qualité des statistiques officielles. On peut néanmoins dire que l'adaptation des statistiques à leur utilisation s'appuie sur un certain nombre de critères de qualité qui sont généralement reconnus, même s'ils évoluent. Les critères de qualité ou d'adaptation des données à leur utilisation dont doit se préoccuper Statistique Canada sont principalement la pertinence, l'exactitude, l'actualité, l'accessibilité, l'intelligibilité et la cohérence. Ces *éléments de la qualité* doivent être étudiés et dosés dans la conception et la mise en oeuvre des programmes statistiques du Bureau.

La *pertinence* des données ou de l'information statistique est une appréciation qualitative de

leur valeur. Cette valeur est caractérisée par le degré d'adéquation des données par rapport aux objectifs de ceux qui les ont produites et aux besoins des utilisateurs. La valeur des données est en outre caractérisée par les mérites de ces objectifs du point de vue du mandat du Bureau, des exigences de la législation et du coût d'opportunité de la production des données ou de l'information.

L'*exactitude* des données ou de l'information statistique est le degré de justesse de l'estimation ou de la description des quantités ou des caractéristiques que l'activité statistique visait à mesurer. La question de l'exactitude a de nombreuses facettes, et, concrètement, il n'existe pas de moyen unique de la mesurer. On ne peut donc mesurer ou décrire ces aspects de l'exactitude que du point de vue de l'erreur ou des conséquences éventuelles des erreurs qui sont associées aux principales sources d'erreur, par exemple, la couverture, l'échantillonnage, la non-réponse, la réponse, le traitement et la diffusion.

L'*actualité* de l'information équivaut au temps qui s'écoule entre le moment où les données sont disponibles et celui où s'est produit l'événement ou le phénomène mesuré, dans l'optique de l'utilité et de l'utilisation des données. Généralement, il faut arriver à un équilibre entre l'actualité et la fiabilité des données.

L'*accessibilité* témoigne de la disponibilité de l'information du Bureau. À cet égard, on doit tenir compte de l'adéquation du support de données, des moyens de diffusion, de l'existence de métadonnées et de la possibilité pour l'utilisateur de savoir que les données sont disponibles et comment il peut y avoir accès. L'abordabilité de l'information pour les utilisateurs par rapport à la valeur qu'elle revêt pour eux est également un aspect de cette caractéristique.

L'*intelligibilité* des données et de l'information renvoie à la facilité avec laquelle l'utilisateur peut comprendre et interpréter correctement les données ou l'information. L'intelligibilité des données est dans une large mesure tributaire de l'adéquation des définitions des concepts, des populations cibles, des variables et des termes sous-jacents, ainsi que des renseignements au sujet des éventuelles limites des données.

La *cohérence* des données et de l'information correspond au degré de liaison logique entre des données et de l'information émanant d'un même programme statistique et les données tirées de plusieurs ensembles ou programmes statistiques, ainsi qu'à leur degré d'intégralité. Une cohérence absolue suppose une cohérence logique : intrinsèquement, dans le temps et d'un produit ou d'un programme à l'autre. S'il y a lieu, les concepts et les populations cibles qui sont utilisés ou présentés doivent se distinguer logiquement de concepts et de populations cibles analogues, mais non identiques, d'autres programmes statistiques ou de notions ou de termes d'usage courant.

Ces éléments de la qualité se chevauchent et sont liés les uns aux autres, souvent de façon arbitraire et déconcertante. De même qu'il n'existe pas de mesure unique de l'exactitude, il

n'y a pas de modèle statistique qui permettrait de réunir efficacement toutes ces caractéristiques de la qualité en un seul indicateur. Il n'existe pas de modèle statistique général qui permettrait de déterminer, sauf dans les cas simples ou unidimensionnels, si un ensemble particulier de caractéristiques qualitatives assure globalement une meilleure qualité qu'un autre ensemble.

Pour obtenir un niveau acceptable de qualité, il faut tenir compte des facteurs ou des éléments qui forment la qualité, les gérer et les équilibrer dans le temps, en portant attention aux objectifs du programme, aux principales utilisations des données, aux coûts ainsi qu'aux conditions et aux circonstances qui ont une incidence sur la qualité des données et sur les attentes des utilisateurs. Comme les différents éléments de la qualité sont unis par des relations complexes, toute mesure qui vise à en modifier un aura généralement des effets sur les autres. L'équilibre entre ces facteurs peut ainsi se trouver modifié de bien des façons, sans qu'il soit possible de le modéliser ni de le quantifier adéquatement à l'avance. Les décisions et les mesures qui permettent d'atteindre cet équilibre s'appuient sur la connaissance, l'expérience, les examens, la rétroaction, la consultation et, forcément, sur le jugement.

La majeure partie du présent document (section 2) met l'accent sur les questions liées à la *qualité* dans l'optique de la *conception* des différentes *enquêtes* (voir la définition présentée à la section 2). Toutefois, il importe de se rappeler que le contexte de l'élaboration de chaque enquête impose des contraintes au plan de la conception. Bien qu'elle vise à satisfaire des besoins en information immédiats, chaque nouvelle enquête enrichit également une base de statistiques qui peut avoir de nombreuses utilisations outre celles identifiées lors de la conception. Par conséquent, il est crucial de voir à ce que les données produites dans le cadre d'une enquête puisse, dans la mesure du possible, être intégrées aux données connexes tirées d'autres enquêtes et utilisées avec elles. Cela suppose qu'il est nécessaire de respecter les normes statistiques régissant le contenu ou les domaines spécialisés qui ont été adoptées afin que les données soient cohérentes et soient harmonisées avec le système statistique national. Cela comprend les cadres statistiques (tels que le Système de comptabilité nationale), les systèmes de classification statistiques (p. ex. ceux qui sont utilisés avec les branches d'activité et les régions géographiques), ainsi que d'autres concepts et définitions qui précisent les variables statistiques à mesurer. L'utilité des statistiques nouvelles est accrue dans la mesure où elles peuvent être employées avec des données existantes.

Le processus de conception se déroule également dans un cadre organisationnel. Les présentes lignes directrices ont été rédigées à l'intention d'un organisme central de statistique au sein duquel la conception des enquêtes relève généralement d'une équipe de projet multidisciplinaire. Celle-ci est principalement constituée du gestionnaire de projet et d'un groupe de spécialistes. Ce groupe comprend habituellement un spécialiste du domaine, un méthodologiste, un spécialiste de l'informatique et un spécialiste des opérations. Il arrive qu'un membre de l'équipe joue plusieurs rôles et qu'on doive faire appel à d'autres spécialistes, par exemple un spécialiste des relations publiques.

La section 3 décrit le cadre administratif dans lequel les lignes directrices concernant la qualité sont mises en oeuvre. La description de ce cadre, qu'on appelle le cadre d'assurance de la qualité, réunit les politiques, les processus de gestion, les mécanismes de consultation et les méthodes techniques qui influent sur la gestion de la qualité à Statistique Canada. Si la section 2 porte principalement sur la conception des différentes activités statistiques, la section 3 analyse l'assurance de la qualité dans une perspective plus globale, soit celle de l'organisation.

Références

Statistique Canada (1997a). Aperçu du cadre d'assurance de la qualité de Statistique Canada. Rapport non publié, Comité des méthodes et des normes, Statistique Canada.

2. ÉTAPES DE L'ENQUÊTE

La présente section est divisée en sous-sections qui correspondent aux principales activités qui se déroulent dans le cadre d'une enquête type. Les sous-sections ont toutes la même structure, étant constituées des rubriques suivantes : *Portée et objet*, *Principes* et *Lignes directrices*. La première sous-section traite de l'étape où les objectifs, les utilisations et les utilisateurs sont identifiés (2.1). Les autres étapes de l'enquête sont décrites dans les sous-sections suivantes, sensiblement dans l'ordre chronologique où elles se déroulent habituellement. Cependant, il existe une interdépendance marquée entre certaines étapes, par exemple entre l'étape de la conception du questionnaire (2.5) et les opérations de collecte des données (2.7). C'est pourquoi certaines sous-sections renvoient à d'autres sous-sections. De même, les étapes étudiées dans le présent document ne se déroulent pas toujours de façon séquentielle. Certaines activités peuvent se dérouler simultanément, par exemple, la conception du questionnaire (2.5), la création de la base de sondage (2.3) et le plan de sondage (2.4). D'autres étapes telles que la documentation (2.16) et l'évaluation de la qualité des données (2.12) sont liées à la plupart des autres activités et ne constituent pas des activités en elles-mêmes. Enfin, une section est consacrée à l'utilisation des données administratives (2.17) afin d'analyser séparément les questions qui se rattachent à ce type de source d'information.

Enquête

Nous employons le terme *enquête* de façon générique pour désigner toute activité visant à recueillir ou à acquérir des données. Cela englobe :

- le *recensement*, qui vise à recueillir des données auprès de tous les membres d'une population;
- l'*enquête-échantillon*, qui consiste à recueillir des données auprès d'un échantillon (habituellement tiré au hasard) de membres de la population;
- la collecte de données provenant de *dossiers administratifs*, qui consiste à obtenir des données à partir de dossiers qui, à l'origine, n'ont pas été constitués à des fins statistiques;
- l'*activité statistique subséquente*, qui consiste à estimer, à modéliser ou à calculer autrement des données à partir de sources de données statistiques existantes.

Les lignes directrices ont été rédigées principalement dans l'optique des recensements et des enquêtes-échantillons. Bien que nombre des lignes directrices s'appliqueront également au traitement des dossiers administratifs, une section supplémentaire (2.17) a été ajoutée afin de mettre en lumière les considérations propres à cette activité. Bien entendu, la qualité des activités statistiques subséquentes est tributaire dans une large mesure de la qualité des éléments utilisés; c'est pourquoi le présent document ne met pas directement l'accent sur les activités statistiques subséquentes.

Conception

Nous employons le terme *conception* pour désigner tous les aspects de l'enquête, de la définition d'un besoin en données à la production de résultats finaux (le fichier de microdonnées, les séries statistiques et l'analyse).

Portée et objet

La rubrique *Portée et objet* décrit l'activité et fait état de ses répercussions éventuelles sur la qualité. Fondamentalement, elle définit l'activité et délimite le contexte.

Principes

Les *principes* sont les politiques, les approches et les directives générales qui servent de base à la conception de l'activité en question, tout particulièrement celles qui sont liées à la qualité.

Lignes directrices

Les *lignes directrices* sont des règles de pratique reconnues qui ont été établies dans le cadre de la conception et de la réalisation d'enquêtes statistiques. Ces lignes directrices ne peuvent pas toutes s'appliquer à chaque enquête. Elles renferment des listes de contrôle qui facilitent la conception de l'enquête. Il convient néanmoins de faire preuve de discernement afin d'évaluer les considérations soulevées par ces lignes directrices.

D'autre part, Statistique Canada a adopté des politiques qui influent à bien des égards sur ses activités statistiques et qui peuvent imposer des exigences quant à la façon dont ces activités sont réalisées. Ces politiques sont décrites dans le Manuel des politiques de Statistique Canada. Chaque fois qu'une politique a une incidence sur un sujet visé par les présentes lignes directrices, nous faisons mention de son existence et de sa pertinence.

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 que cet exercice de planification initiale engendre un instrument de collecte véritable, il importe ensuite de préciser les objectifs et les utilisations pour s'assurer que la nouvelle activité ou l'activité refondue réponde aux besoins particuliers des utilisateurs.

Les *objectifs* indiquent pourquoi les données seront recueillies; ils sont formulés dans le contexte du programme, du problème de recherche ou des hypothèses qui font que l'information est nécessaire. Les *utilisations* décrivent les décisions qui seront prises à la lumière des données recueillies et expliquent comment ces dernières serviront de fondement aux décisions. Dans le cas des enquêtes périodiques, les utilisations peuvent évoluer avec le temps. Les *utilisateurs* sont les organisations, les organismes, les groupes ou les particuliers qui sont censés utiliser l'information. Un consensus sur les utilisations et les objectifs spécifiques, permet de prendre des décisions rationnelles relativement à la conception de l'enquête.

Principes

Le fait de préciser les objectifs et les utilisations nous amène à établir un plan détaillé pour la nouvelle activité de concert avec les utilisateurs des données et les participants au projet. Les utilisateurs peuvent être mis à contribution aux fins de la description des objectifs de l'activité. Les participants au projet peuvent quant à eux cerner les problèmes conceptuels, méthodologiques et opérationnels qui doivent être réglés et proposer un calendrier de travail raisonnable.

Il importe que les objectifs soient bien compris et soient énoncés concrètement en ce qui a trait aux hypothèses à vérifier et à la nature précise des données à recueillir; cet énoncé des objectifs doit également faire état du niveau de qualité escompté, des contraintes budgétaires et des dates de livraison prévues.

L'énoncé des objectifs permettra aux utilisateurs subséquents, qui n'auraient pas les mêmes objectifs, de déterminer dans quelle mesure le produit pourrait répondre à leurs besoins. C'est également un outil de communication important afin que les utilisateurs (éventuels) sachent ce qu'ils peuvent attendre des produits de l'activité statistique et dans quelle mesure ils doivent être prudents s'ils font un usage des données qui va au-delà de ce qui était prévu.

Lignes directrices

- Dans le cadre de l'analyse des besoins des utilisateurs, il faut mettre l'accent sur les solutions les plus rentables à court et à long termes. Avant d'entreprendre la conception d'une nouvelle activité statistique (ou la refonte d'une activité existante), il convient d'analyser les données actuellement disponibles dans le domaine, notamment leur provenance, la fréquence de la production, leur qualité, leur actualité, etc. Il faut mettre dans la balance, d'une part l'adéquation des données par rapport aux besoins des clients, et d'autre part, l'argent et le temps qu'il faut consacrer à la réalisation d'une nouvelle activité visant à produire des statistiques qui n'existent pas encore.
- Il convient de formuler les objectifs de l'enquête en collaboration avec les grands utilisateurs et les intervenants. Les gestionnaires de programme doivent demeurer en liaison étroite avec les utilisateurs des secteurs privé et public et du grand public afin d'assurer la pertinence de l'information produite et de commercialiser leurs produits et services. Parmi les grands utilisateurs, mentionnons les représentants des marchés éventuels, les décideurs qui utiliseront l'information ainsi que les mandataires qui en ont besoin à des fins prévues par la loi. Avant d'entreprendre des travaux de conception ou de refonte de grande envergure, de vastes consultations ciblées devront être menées systématiquement; on pourra ainsi déterminer quelles sont les options au plan du contenu et amener le public à appuyer le programme à l'étape de la collecte des données. Lorsqu'il s'agit d'élaborer ou de modifier un programme, il importe d'adopter une approche transparente, car des statistiques pertinentes et exactes ne servent à rien sans la confiance.
- Lorsqu'on détermine dans quelle mesure une enquête répondra aux besoins des utilisateurs, il importe d'arriver à un équilibre raisonnable entre ces besoins et le budget, le fardeau de réponse et la protection des renseignements personnels. Bien que le pouvoir discrétionnaire de l'organisme puisse être limité lorsqu'il y a une prescription d'une loi, dans d'autres cas, on a intérêt à élaborer des approches de rechange en ce qui a trait à la méthodologie, aux méthodes et aux modes de collecte, à la fréquence, aux niveaux de détail géographique, etc. en vue d'en arriver à une solution optimale.
- À intervalles réguliers, il faut passer en revue les activités statistiques en cours. Les programmes statistiques doivent évoluer, s'adapter et innover, afin de continuer à répondre aux exigences des utilisateurs-clients. Le but de l'activité ou l'énoncé de ses objectifs doit être revu périodiquement afin que le produit statistique demeure pertinent compte tenu de l'évolution des besoins de l'utilisateur. Il est parfois essentiel de procéder à la refonte des enquêtes existantes pour que les principales séries statistiques demeurent fiables, tout particulièrement si les sources d'information ont changé ou si le mode de diffusion est modifié en profondeur ou repensé.

- Lorsque des objectifs explicites de qualité des données ont été formulés, il convient de les inscrire dans l'énoncé des objectifs de l'enquête sous forme d'éléments mesurables de la qualité. On peut fixer des objectifs sous forme de mesures (p. ex. taux de réponse, erreurs d'échantillonnage, taux de couverture et actualité). Dans le cas des données administratives et des activités statistiques subséquentes, la qualité du produit est directement liée à celle des données d'entrée.
- Lorsqu'on définit les objectifs et les utilisations d'une activité statistique, 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

Brackstone, G. J. (1991). Shaping Statistical Services to Satisfy User Needs. **Statistical Journal of the United Nations**, ECE 8, 243-257.

Brackstone, G. J. (1993). Data Relevance: Keeping Pace with User Needs. **Journal of Official Statistics**, 9, 49-56.

2.2 Concepts, définitions et classifications

Portée et objet

Les *concepts* sont les sujets d'enquête et d'analyse qui intéressent les utilisateurs. Ils renvoient aux caractéristiques ou aux attributs généraux d'une unité statistique ou d'une population d'unités statistiques similaires. Les *définitions* décrivent non seulement les concepts de façon précise mais également les variables qu'on utilise pour mesurer les concepts. Les définitions des variables spécifiques requises aux fins de la mesure des concepts ne doivent pas être ambiguës et doivent être clairement énoncées dans le contexte de l'analyse qui est l'objet de la cueillette des données. Comme toutes les données doivent être classifiées aux fins de l'analyse, les critères de *classification* qu'on choisit afin de les grouper systématiquement doivent être adéquats dans l'optique de l'analyse.

Principes

Afin de tirer des conclusions à partir d'un ensemble de données, il est essentiel que les utilisateurs connaissent les concepts et les définitions sous-jacentes (c.-à-d. ce que les données sont censées mesurer) et soient consultés à cet égard. Bien que le recours à des concepts, à des définitions et à des classifications harmonisés aide les utilisateurs à comparer et à intégrer les données, il se peut que ceux qui sont en usage doivent être différents pour qu'on puisse satisfaire le besoin. Les concepts, les définitions et les classifications doivent être documentés avec soin, et tout écart par rapport aux normes ou à ceux qu'on emploie pour produire des données connexes doivent être soulignés. Cette documentation est particulièrement importante pour les utilisateurs qui désirent employer les statistiques à d'autres fins que celle prévue.

Il existe parfois plusieurs façons de mesurer un concept. Il se peut que les définitions et les classifications choisies doivent également tenir compte d'autres facteurs, par exemple la facilité avec laquelle on peut obtenir les données requises, le fardeau de réponse, la méthode de collecte, le contexte dans lequel la question ou les questions doivent être posées, le traitement des données (notamment les techniques de vérification, d'imputation et de pondération), la possibilité d'extraire l'information des dossiers administratifs et les coûts associés à la collecte et au traitement. Donc, l'approche choisie aux fins de la mesure permettra plus ou moins d'obtenir l'interprétation désirée du concept. La définition choisie à un moment donné sera peut-être désuète ultérieurement si de nouveaux facteurs font sentir leur influence et, par conséquent, devra être modifiée.

Lignes directrices

- Les concepts et les définitions devraient être clairement précisés; de plus, il faut établir un lien entre eux et l'utilisation prévue. Il serait bon d'utiliser la politique concernant les normes (Statistique Canada, 1998c). Les règles de dénomination devraient tenir

compte de la similarité ou de la dissimilarité avec les normes en vigueur et l'usage courant. Les appellations tirées des normes existantes devraient uniquement être employées pour désigner ce qui est défini dans les normes.

- Afin de maximiser la flexibilité, il convient, dans la mesure du possible, de coder les données et de conserver les fichiers au niveau le plus bas de la classification appropriée. L'agrégation à un niveau supérieur peut être nécessaire aux fins de certaines analyses ou encore en raison de contraintes de confidentialité ou de fiabilité des données. Dans la mesure du possible, ces agrégations devraient être effectuées dans le cadre d'une stratégie de regroupement commune et être définies en fonction des classes ou des agrégations à des niveaux supérieurs de la norme. Les différences devraient être documentées. Les classifications devraient faire état des niveaux les plus détaillés et des niveaux regroupés. Il faut également expliquer clairement aux utilisateurs où ces données s'insèrent dans les classifications à un niveau supérieur (c.-à-d. moins détaillées).
- Grâce à l'utilisation de définitions et de classifications types, il est possible de comparer des données provenant de sources différentes et de les intégrer (Statistique Canada, 1998c). Statistique Canada se sert de classifications types pour les industries, les produits, les professions et les régions géographiques (Statistique Canada, 1980, 1986a, 1992d, 1993d, 1993e). Des politiques se rapportant à ces classifications types ont été publiées aux fins de la collecte et de la diffusion des données (Statistique Canada, 1987b, 1994b, 1997c). À compter de l'année de référence 1997, le Système de classification des industries de l'Amérique du Nord (Statistique Canada, 1998a) remplacera progressivement la Classification type des industries de 1980 et la Classification canadienne type des industries pour compagnies et entreprises de 1980 dans les programmes statistiques de Statistique Canada.
- L'initiative d'harmonisation de la statistique sociale de Statistique Canada a mis en lumière les concepts, les définitions des variables, les classifications éventuelles et la concordance avec les pratiques internationales dans divers domaines spécialisés (Statistique Canada, 1997d). L'état de chaque variable, à savoir si elle constitue une pratique excellente recommandée, une norme recommandée ou une norme, est également précisé. Cette initiative privilégie le recours à des définitions et à des classifications types, mais laisse une certaine latitude aux utilisateurs, en leur permettant de formuler leurs questions en fonction des contraintes de la politique concernant les normes et de leur méthodologie et de leur instrument de collecte dans la mesure où les produits respectent les normes.
- Outre les classifications types de Statistique Canada, il existe des classifications types internationales produites par le Bureau de statistique des Nations Unies, le Bureau international du travail, Eurostat et d'autres organismes internationaux et régionaux. Comme les chargés d'enquête doivent souvent fournir des données à des organismes

internationaux qui utilisent les classifications internationales, la Division des normes a réalisé des concordances officielles avec la Classification internationale type, par industrie, de toutes les branches d'activité économique (CITI Rev2 et 3) et la classification des branches d'activité utilisée par Eurostat (NACE Rev 1), de même qu'avec la Classification internationale type des professions (CITP) afin d'aider les divisions d'enquête à convertir les données de façon uniforme. (U.S. Executive Office of the President et coll., 1997). Des concordances entre les différentes éditions des classifications nationales ont également été réalisées afin d'aider les chargés d'enquête et les analystes à intégrer les séries statistiques.

- De même, l'utilisation d'unités d'observation types facilite la comparaison des données. De plus, les classifications sont généralement conçues dans l'optique de certaines unités d'observation. Par exemple, dans le cas des statistiques de la production, la classification type des industries repose sur l'établissement et l'emplacement; pour ce qui est des statistiques financières, les unités d'observation sont la compagnie et l'entreprise.
- Dans les activités statistiques subséquentes ou les cadres statistiques (p. ex. le Système de comptabilité nationale), on définit un ensemble de concepts et de définitions qui peuvent avoir une incidence considérable sur certaines activités de collecte de données (Statistique Canada, 1989).
- En l'absence d'une norme officielle, il est utile d'analyser les concepts, les définitions et les classifications employés par les programmes statistiques connexes.

Références

Statistique Canada (1980). **Classification type des industries 1980**. Division des normes, Statistique Canada, Ministre des Approvisionnements et Services du Canada, Ottawa.

Statistique Canada (1986a). **Classification canadienne type des industries pour compagnies et entreprises, 1980**. Division des normes, Statistique Canada, Ministre des Approvisionnements et Services du Canada, Ottawa.

Statistique Canada (1987b). Politique sur les classifications types des industries. **Manuel des politiques, 2.7**.

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

Statistique Canada (1992d). **Classification géographique type (CGT) 1991**. Division des normes, Statistique Canada, Ministre de l'Industrie, des Sciences et de la Technologie, Ottawa.

- Statistique Canada (1993d). **Classification type des biens 1992** (basée sur le Système harmonisé de désignation et de codification des marchandises). Division des normes, Statistique Canada, Ministre de l'Industrie, des Sciences et de la Technologie, Ottawa.
- Statistique Canada (1993e). **Classification type des professions 1991**. Division des normes, Statistique Canada, Ministre de l'Industrie, des Sciences et de la Technologie, Ottawa.
- Statistique Canada (1994b). Politique sur les classifications types des professions. **Manuel des politiques**, 2.9.
- Statistique Canada (1997c). Politique sur les unités géographiques types. **Manuel des politiques**, 2.4.
- Statistique Canada (1997d). Initiative d'harmonisation de la statistique sociale. Rapport non publié, version 2.1, Statistique Canada.
- Statistique Canada (1998a). **Système de classification des industries de l'Amérique du Nord, Canada 1997**. Publication n° 12-501 au catalogue, Statistique Canada.
- Statistique Canada (1998b). Rapport d'étape sur l'harmonisation des statistiques sociales. Article présenté à l'occasion de la Réunion de travail sur les systèmes de métainformation, Conférence des statisticiens européens, Commission de statistique et Commission économique pour l'Europe, février 1998.
- Statistique Canada (1998c). Politique concernant les normes. **Manuel de politiques**, 2.10.
- U.S. Executive Office of the President, Office of Management and Budget, U.S. Department of Commerce, Statistique Canada et Office statistique des Communautés européennes (1997). Concordance internationale entre les classifications industrielles des Nations Unies (CITI Rev 3) et du Canada (CTI 1980), de l'Union Européenne (NACE Rev 1) et des États-Unis (1987 SIC).

2.3 Couverture et bases de sondage

Portée et objet

Une *base de sondage* est une liste, un document ou un système qui délimite et définit les éléments d'une population-cible et permet d'y avoir accès. La *population-cible* correspond à l'ensemble d'éléments sur lesquels on désire recueillir des renseignements et produire des estimations; bien que, pour des considérations d'ordre pratique, il peut être nécessaire d'exclure certaines unités (p. ex. personnes vivant en institution, sans-abris). La *couverture* désigne la proportion des éléments de la population-cible qui se retrouvent dans la base de sondage. En général, il existe deux types de base de sondage : la base aréolaire et la base liste.

La présente section porte sur le choix d'une base de sondage ainsi que sur la vérification et la préservation de la qualité de la couverture et des caractéristiques connexes.

Principes

La base de sondage doit correspondre à la population-cible, et le sous-dénombrement et le surdénombrement (y compris les duplicats) doivent être minimes. La création, l'utilisation, la mise à jour et la vérification de la base de sondage doivent respecter les contraintes opérationnelles et financières.

Il importe que les caractéristiques des unités de la base de sondage (p. ex. la classification, la personne-ressource, l'adresse et la taille) soient de qualité supérieure en raison de leur utilisation dans le cadre de la stratification, de la collecte, du suivi, de l'estimation, du couplage des enregistrements, de l'évaluation de la qualité et de l'analyse. Les imperfections de la base de sondage telles que les erreurs de couverture et les caractéristiques désuètes sont susceptibles de biaiser les estimations de l'enquête, d'en diminuer la fiabilité et d'accroître les coûts de la collecte des données.

Lignes directrices

- Vérifier à l'étape de la planification de l'enquête la qualité des bases de sondage qui pourraient être utilisées et déterminer si elles sont adéquates.
- Lorsqu'il existe plusieurs bases de sondage et que certaines d'entre elles sont incomplètes ou que leur utilisation occasionne des coûts prohibitifs, il convient d'envisager le recours à des bases multiples (Hartley, 1962; Sirken et Casady, 1988).
- À Statistique Canada, la base de sondage habituellement utilisée dans les enquêtes menées auprès des entreprises et des institutions est le registre des entreprises. Dans les enquêtes sur l'agriculture, on utilise le registre des fermes. Pour ce qui est des enquêtes-ménages, on peut envisager de recourir au registre des adresses, à la base

de sondage de l'Enquête sur la population active et aux fichiers téléphoniques.

- Dans la mesure du possible, utiliser la même base de sondage dans les enquêtes ayant la même population-cible afin d'éviter les incohérences et de réduire les coûts de la mise à jour et de l'évaluation de la base.
- Afin de rehausser ou de maintenir la qualité de la base de sondage, adopter une marche à suivre qui permet d'éliminer les dédoublements et de mettre à jour la base en fonction des naissances, des décès, des unités hors du champ d'observation et des changements apportés aux caractéristiques.
- Vérifier la qualité de la base de sondage en évaluant périodiquement sa couverture. Pour ce faire, on peut recourir à plusieurs techniques : coupler la base de sondage ou un échantillon de la base avec d'autres sources comparables pour la population-cible ou des sous-ensembles de celle-ci; analyser les déclarations des répondants à l'enquête afin de déceler les dédoublements, les décès, les unités hors du champ de l'enquête et l'évolution des caractéristiques; utiliser certaines questions du questionnaire afin de faciliter le suivi de la couverture; effectuer des vérifications auprès des autorités locales (p. ex. les bureaux régionaux, le personnel affecté aux enquêtes sur le terrain, les unités d'enquête); vérifier la base de sondage ou des sous-ensembles de celle-ci sur le terrain (p. ex. vérifier les unités hors du champ); comparer la base de sondage aux listes dressées à partir d'un échantillon d'unités d'une base aréolaire correspondante; mettre à jour la base de sondage afin de déterminer quels changements y ont été apportés; vérifier la cohérence des chiffres par rapport à d'autres sources ou à des données obtenues au moyen de répétitions spéciales; employer des données d'évaluation produites dans le cadre d'autres enquêtes reposant sur la même base de sondage (Lessler et Kalsbeek, 1992).
- Dans certains cas, il peut être souhaitable de vérifier la base de sondage entre le tirage de l'échantillon et la période de référence de l'enquête.
- Les mises à jour de la base de sondage doivent être effectuées le plus rapidement possible.
- En formant adéquatement le personnel, en faisant ressortir l'importance de la couverture et en adoptant des méthodes efficaces d'assurance de la qualité dans le cadre des activités relatives à la base de sondage, on contribue ainsi à minimiser les erreurs dans la base.
- Dans le cas des bases aréolaires, il convient de procéder à des vérifications de la carte géographique pour s'assurer que les secteurs géographiques utilisés dans le plan d'échantillonnage sont clairement délimités et ne se chevauchent pas (p. ex. au moyen de vérifications sur le terrain ou d'autres cartes-sources).

- Dans le cas des activités statistiques faisant appel aux sources administratives ou des activités statistiques subséquentes, l'évaluation et la vérification de la couverture doivent être effectuées par l'entremise du gestionnaire responsable des sources administratives lorsque les changements apportés à la couverture ne relèvent pas du gestionnaire des activités statistiques. Lorsqu'il est possible d'exercer une influence sur la base de sondage, les modifications requises doivent être négociées avec le gestionnaire de la source.
- Il peut être nécessaire de rajuster les données ou d'utiliser des données supplémentaires provenant d'autres sources afin de compenser les différences existant entre la base de sondage et la population-cible au chapitre de la couverture.
- La documentation de l'enquête doit renfermer une description de la population-cible, de la base de sondage et de la couverture.

Références

- 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. in **Business Survey Methods**, B.G. Cox et coll. (éd.), Wiley-Interscience, New York, 21-47.
- Gosselin, J.-F., Chinnappa, B.N., Ghangurde, P.D. et Tourigny, J. (1978). Couverture. Chapitre 2, dans **Répertoire des méthodes d'évaluation des erreurs dans les recensements et les enquêtes**, 7-10. Publication n° 13-564F au catalogue, Statistique Canada.
- Groves, R.M., Biemer, P., Lyberg, L., Massey, J., Nicholls, W. et Waksberg, J. (éd.) (1988). **Telephone Survey Methodology**. Wiley, New York.
- Hartley, H.O. (1962). Multiple Frame Surveys. **Proceedings of the Social Statistics Section**, American Statistical Association, 203-206.
- 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. In **Telephone Survey Methodology**, R.M. Groves et coll. (éd.), Wiley & Sons, 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° 12-602F au catalogue, Statistique Canada.
- Singh, M.P., Drew, J.D., Gambino, J.G. et Mayda, F. (1990). **Méthodologie de l'Enquête sur la population active du Canada**. Publication n° 71-526 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 in **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* consiste à choisir un ensemble d'unités dans une population-cible. Cet ensemble d'unités est désigné sous le nom d'*échantillon*. Le choix de la méthode d'échantillonnage influe directement sur la qualité des données; il est influencé par de nombreux facteurs, notamment le degré de précision que devront posséder les données produites, la disponibilité de bases de sondage appropriées, la disponibilité de variables de stratification adéquates, les méthodes d'estimation qui seront employées et le budget disponible.

Principes

On a recours à l'*échantillonnage probabiliste* afin de constituer un échantillon représentatif de la population-cible. On cherche ainsi à recueillir des renseignements utiles à partir des unités échantillonnées afin de tirer des inférences sur la population-cible. L'échantillonnage probabiliste suppose une sélection probabiliste à partir de la base de sondage de manière que toutes les unités de la population-cible aient une *probabilité d'inclusion* connue et positive. La taille de l'échantillon est déterminée en fonction de la précision exigée et du budget affecté à l'observation des unités sélectionnées. La distribution des probabilités qui régit le tirage de l'échantillon est un *plan d'échantillonnage*. On choisit une combinaison de plan d'échantillonnage et de méthode d'estimation (voir la section 2.10) de manière à produire des estimations aussi précises que possible compte tenu du budget ou à obtenir un degré de précision déterminé en réduisant les coûts au minimum. Les données recueillies à l'égard des unités échantillonnées peuvent être complétées, à l'étape de l'estimation, par des données (les *données auxiliaires*) provenant d'autres sources que l'enquête, par exemple les dossiers administratifs et les projections du recensement. Le choix du plan d'échantillonnage tient compte de la disponibilité des données auxiliaires. Ces concepts sont analysés dans Särndal, Swensson et Wretman (1992).

Lignes directrices

- La *stratification* consiste à diviser la population en sous-ensembles (qu'on appelle des strates), avant qu'un échantillon n'ait été constitué dans chacun de ces sous-ensembles. D'un point de vue pratique, ces sous-ensembles peuvent être définis en fonction des objectifs de l'enquête, par exemple lorsque des estimations provinciales sont requises. Du point de vue de l'efficacité statistique, la stratification est effectuée de manière que chaque strate renferme des unités relativement homogènes par rapport à l'information qu'on veut recueillir dans le cadre de l'enquête.
- Il arrive que l'information nécessaire à la stratification ne se trouve pas dans la base de sondage. En pareil cas, on pourrait constituer un premier échantillon de grande

envergure afin d'obtenir ces renseignements. Le premier échantillon est ensuite stratifié, et un sous-échantillon est tiré dans chaque strate. Pareil plan d'échantillonnage est désigné sous l'expression d'*échantillonnage à deux phases*. Lorsqu'on opte pour ce type de plan, il importe de tenir compte du coût de l'échantillonnage à chaque phase, de la disponibilité de l'information nécessaire à chaque phase ainsi que de la précision accrue que procure la stratification de l'échantillon de la première phase.

- En pratique, il est parfois difficile de choisir les unités qui déclareront l'information demandée ou peu commode de communiquer avec elles directement. En pareil cas, on doit d'abord choisir les unités (les *unités primaires d'échantillonnage*) qui sont des *grappes* d'unités déclarantes, et ensuite procéder à un sous-échantillonnage dans chacune des unités primaires d'échantillonnage sélectionnées afin de constituer un échantillon des unités déclarantes. C'est ce qu'on appelle l'*échantillonnage à deux degrés*. En raison de contraintes budgétaires ou d'autres contraintes, il peut être nécessaire d'aller au-delà de deux degrés. Il est important de déterminer combien de degrés d'échantillonnage sont nécessaires et quelles unités d'échantillonnage sont appropriées à chaque degré. Pour chaque type d'unité éventuel, on doit tenir compte de la facilité avec laquelle on peut communiquer avec l'unité déclarante et recueillir et mesurer les données, de la qualité des renseignements fournis par l'unité ainsi que du coût de la collecte de données sur cette unité et auprès de celle-ci.
- Lorsqu'on détermine la taille de l'échantillon, on doit prendre en considération les degrés de précision requis pour les estimations d'enquête, le type de plan et d'estimateur qui sera employé, la disponibilité de données auxiliaires ainsi que les facteurs liés à l'échantillonnage (p. ex. la mise en grappes) et les facteurs non liés à l'échantillonnage (p. ex. la non-réponse). Dans le cas des enquêtes périodiques, il faut tenir compte qu'il y aura des naissances et des décès parmi la population-cible.
- Il est important de se rappeler que, dans le cadre de la plupart des enquêtes, on produit des estimations à l'égard de nombreuses variables et que l'optimisation de l'échantillon pour une variable en particulier peut avoir un effet défavorable sur d'autres variables importantes. Pour contourner ce problème, on commence généralement par identifier les variables les plus importantes et ensuite, on utilise ce sous-ensemble de variables pour définir et répartir l'échantillon.
- Lorsque la population présente une asymétrie marquée, l'enquête devrait utiliser une strate de grandes unités qui seront échantillonnées avec certitude. Ces grandes unités représentent généralement une part non négligeable des estimations produites à l'égard de la population.
- Comme la plupart des enquêtes visent à produire des estimations pour divers *domaines* d'intérêt (c.-à-d. des sous-ensembles d'une population), il convient d'en

tenir compte dans le plan en assurant une stratification adéquate (par exemple, selon la province). Si cet élément est omis, il faudra envisager des méthodes spéciales à l'étape de l'estimation (voir la section 2.10).

- Il convient de réaliser des études afin d'évaluer l'efficacité des solutions de rechange relativement aux méthodes d'échantillonnage, à la stratification et à la répartition de l'échantillon. L'utilité de ces études repose sur la disponibilité et la qualité des données des enquêtes et des recensements antérieurs ainsi que des données administratives employées aux fins de la réalisation des études.
- Durant l'échantillonnage, la taille et les caractéristiques de l'échantillon réel devraient être comparées à ce qui était prévu. Le suivi du processus d'échantillonnage peut nous aider à déceler et à corriger les erreurs éventuelles, par exemple, l'utilisation de taux d'échantillonnage erronés ou la perte de questionnaires. De même, il est important de comparer la précision des estimations avec les objectifs.
- La répartition de l'échantillon et la taille des échantillons stratifiés devraient tenir compte des taux prévus de classification erronée des unités dans la base de sondage. Ce problème s'explique du fait que les bases de sondage sont souvent désuètes et que, par conséquent, les données sur la classification utilisées aux fins de la stratification ne sont pas toujours actuelles. Si elles ne sont pas prises en considération comme il se doit à l'étape de l'échantillonnage, les estimations de l'enquête ne seront pas aussi précises que prévu. Il faut également s'attaquer à ce problème à l'étape de l'estimation (voir la section 2.10).
- Dans le cas des enquêtes périodiques conçues de manière que la taille de l'échantillon suit la croissance de la population, il convient souvent d'élaborer une méthode afin que la taille de l'échantillon demeure stable. La *précision* de l'enquête est généralement davantage tributaire de la taille globale de l'échantillon que du *taux d'échantillonnage* (le ratio de la taille de l'échantillon à la taille de la population).
- Il importe que le plan des enquêtes périodiques soit aussi flexible que possible afin qu'on puisse s'adapter aux changements futurs, par exemple une augmentation ou une diminution de la taille de l'échantillon, une nouvelle stratification, un nouvel échantillonnage ou une révision des probabilités de sélection. Si des estimations doivent être produites à l'égard de certains domaines d'intérêt (p. ex. des estimations infraprovinciales), on devrait, dans la mesure du possible, former les strates en fusionnant de petites unités stables associées aux domaines en question (p. ex. de petites régions géographiques). Il sera alors plus facile de s'adapter aux modifications qui seront apportées aux définitions des strates.
- Dans les enquêtes périodiques, si des estimations efficaces du changement sont nécessaires ou si le fardeau de réponse suscite de l'appréhension, on devrait recourir

à l'échantillonnage avec *renouvellement*, qui consiste à remplacer une partie de l'échantillon à chaque période. Le choix de la fréquence de rotation sera le fruit d'un compromis entre la précision exigée pour les estimations du changement et le fardeau de réponse imposé aux unités déclarantes. Moins l'échantillon sera renouvelé souvent, plus les estimations du changement seront précises et plus le taux de réponse sera faible. Un renouvellement peu fréquent présente un autre avantage, soit une réduction des coûts si la première prise de contact est beaucoup plus onéreuse que les prises de contact subséquentes.

- Toujours dans les enquêtes périodiques, il convient d'élaborer des méthodes afin de contrôler la qualité du plan d'échantillonnage dans le temps. On devrait également adopter une stratégie de mise à jour afin de procéder à une refonte sélective des strates qui se sont détériorées considérablement.
- Dans la mesure du possible, on devrait utiliser des logiciels généralisés, par opposition à des systèmes faits sur mesure, aux fins du tirage des échantillons. On peut notamment recourir au Système généralisé d'échantillonnage (SGECH), mis au point par Statistique Canada. L'utilisation de systèmes généralisés devrait réduire le nombre d'erreurs de programmation ainsi que les coûts et les délais du 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**. John Wiley and Sons, New York.

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.

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

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

Singh, M.P., Drew, J.D., Gambino, J.G. et Mayda, F. (1990). **Méthodologie de l'Enquête sur la population active du Canada**. Publication n° 71-526 au catalogue, Statistique Canada.

2.5 Élaboration du questionnaire

Portée et objet

Le *questionnaire* consiste en une série de questions qu'on formule en vue de recueillir des renseignements auprès d'un répondant. Un intervieweur peut poser les questions, ou le répondant peut remplir le questionnaire; la collecte des renseignements peut se faire avec papier et crayon ou à l'aide d'un ordinateur. Le questionnaire joue un rôle clé dans le cadre du processus de collecte de renseignements. Son influence sur la qualité des données, le comportement des répondants, la prestation de l'intervieweur et les relations avec les répondants s'avère déterminante.

Principes

Le questionnaire est conçu en fonction des exigences des utilisateurs au plan statistique, des exigences administratives de l'organisme chargé de la réalisation de l'enquête, des exigences relatives au traitement de l'information ainsi que de la nature et des caractéristiques des répondants. Un questionnaire bien conçu impose un faible fardeau de réponse tout en étant convivial pour le répondant comme pour l'intervieweur. Il permet de recueillir les données avec efficacité en occasionnant un minimum d'erreurs; de plus, il facilite le codage et la saisie des renseignements et minimise la vérification et l'imputation qui doivent être effectuées.

Les questionnaires des enquêtes permanentes doivent être évalués périodiquement. La Politique sur l'élaboration, l'essai et l'évaluation des questionnaires (Statistique Canada, 1994c; voir l'annexe 1) exige que tous les questionnaires nouveaux ou modifiés qui sont élaborés à Statistique Canada soient testés avant d'être utilisés, et ce dans leurs versions française et anglaise.

Lignes directrices

- Dans le questionnaire, utilisez une formulation et des concepts ayant la même signification pour les répondants et les concepteurs; dans les enquêtes-entreprises, choisissez des questions, des périodes de référence et des catégories de réponse qui soient compatibles avec les pratiques de l'établissement en matière de conservation de dossiers. Dans la mesure du possible, harmonisez les concepts et la formulation avec ce qui existe déjà, et réutilisez les questions provenant d'autres enquêtes.
- La formulation des questions et les termes employés doivent encourager les répondants à remplir le questionnaire. À cette fin, il faut que le questionnaire mette l'accent sur le sujet de l'enquête, que ses questions s'enchaînent en douceur et qu'il suive la logique du répondant pour l'aider à se souvenir des faits et l'aiguiller vers la source d'information appropriée.

- Veillez à ce que l'introduction du questionnaire fasse état du titre ou du sujet de l'enquête et de son commanditaire, explique l'objet de l'enquête, fasse appel à la collaboration du répondant et précise en vertu de quels pouvoirs l'enquête est réalisée. De plus, faites-y mention de la protection de la confidentialité, des projets de couplage d'enregistrements et des ententes de partage des données (Statistique Canada, 1996a). Les premières questions doivent s'adresser à tous les répondants; il doit être facile de répondre à ces questions et celles-ci doivent être intéressantes et démontrer que le répondant fait partie de la population-cible.
- Assurez-vous que l'utilité de fournir les renseignements ne fasse aucun doute pour les répondants et expliquez pourquoi il est important de remplir le questionnaire et comment les données d'enquête seront utilisées.
- Dans le cas d'un questionnaire qui doit être rempli par le répondant, faites en sorte qu'il soit attrayant et facile à remplir. À cette fin, la lettre d'accompagnement et la page couverture doivent faire bonne impression au premier coup d'oeil. Donnez au questionnaire une apparence professionnelle et sérieuse. Si on pose les questions en personne ou au téléphone, assurez-vous que le questionnaire facilite la tâche de l'intervieweur.
- Afin de réduire au minimum le risque d'erreurs de déclaration, assurez-vous que les instructions aux répondants ou aux intervieweurs soient brèves, claires et faciles à trouver. Au besoin, fournissez des définitions au début du questionnaire ou dans certaines questions. Assurez-vous que le répondant sache exactement quelles sont la période de référence et les unités de mesure à utiliser; imprimez en caractères gras les éléments importants; insérez le verbe « inclure » ou « exclure » dans les questions (et non dans les instructions); assurez-vous que les catégories de réponse soient exhaustives et sans chevauchement.
- En ce qui a trait à la présentation du questionnaire, mettez un titre ou un en-tête à chaque section; de plus, assurez-vous que les instructions et l'espace prévu pour les réponses favorisent l'exactitude des réponses. Faites appel à la couleur, à l'ombrage, aux illustrations et aux symboles pour attirer l'attention du répondant ou de l'intervieweur, pour l'aiguiller vers les sections qui doivent être lues et pour indiquer où les réponses doivent être inscrites. À la dernière page ou à la fin du questionnaire, prévoyez un espace afin que le répondant puisse inscrire ses commentaires et ajoutez une formule de remerciement.
- On peut employer un large éventail de méthodes afin de tester et d'évaluer le questionnaire. Leur efficacité et la fréquence de leur utilisation sont tributaires de divers facteurs et circonstances. Mentionnons notamment le type et l'envergure de l'enquête, son contenu, l'utilisation des questions de la dernière enquête, s'il s'agit d'une enquête permanente, la méthode de collecte, le calendrier du projet, le budget

et la disponibilité des ressources (Statistique Canada, 1994c; voir l'annexe 1).

- Afin de savoir comment les répondants réagissent à un questionnaire, effectuez des tests qualitatifs. Parmi les méthodes qu'on peut employer, mentionnons les groupes de discussion et les interviews en profondeur, les approches cognitives telles que la reformulation et les interviews où on amène le répondant à penser à voix haute, ainsi que le codage du comportement. On a recours aux groupes de discussion et aux interviews individuelles en profondeur pour tester et évaluer le libellé, l'ordre et la présentation des questions. Les approches cognitives permettent d'analyser le processus mental des répondants lorsqu'ils répondent aux questions de l'enquête et de vérifier s'ils les comprennent et peuvent fournir des réponses exactes. En analysant l'interaction entre l'intervieweur et le répondant, le codage du comportement constitue un moyen systématique et objectif d'évaluer l'efficacité du questionnaire. On peut également procéder à des tests qualitatifs pour définir le contenu du questionnaire en évaluant et en explorant des concepts clés.
- On peut soumettre le questionnaire à un test préliminaire officieux afin de déceler les lacunes dans la formulation et l'ordre des questions, les erreurs dans la présentation du questionnaire ou dans les instructions, ainsi que les problèmes attribuables à l'incapacité ou au refus de répondre. On peut également effectuer un test préliminaire pour proposer des catégories de réponse supplémentaires qui peuvent être précodées sur le questionnaire, et pour donner un aperçu de la durée de l'interview et des problèmes de refus. Le test préliminaire peut être suivi de séances de compte-rendu avec les intervieweurs afin de tenir compte de leurs commentaires dans le cadre du processus de conception ou de refonte.
- Effectuez un test avec échantillon fractionné lorsqu'il existe au moins deux versions du questionnaire et qu'on veut déterminer laquelle est la « meilleure ». Cela consiste à intégrer un concept expérimental au processus de collecte afin d'étudier des problèmes tels que la formulation et l'agencement des questions ainsi que les méthodes de collecte.
- Réalisez un essai-pilote après un test préliminaire approfondi afin de voir comment l'ensemble des opérations de l'enquête, notamment celle qui consiste à remplir le questionnaire, s'agencent en pratique. L'étude-pilote est la « générale ». Elle reproduit sur une petite échelle le plan final de l'enquête dans toutes ses étapes, sans oublier le traitement et l'analyse des données. L'essai-pilote permet de mettre la dernière main au questionnaire avant qu'il soit utilisé dans le cadre de l'enquête principale.
- Assurez-vous que les versions française et anglaise du questionnaire concordent.

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, Beverley Hills, Californie.
- 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.
- Fowler, F.J. Jr. (1995). Improving Survey Questions: Design and Evaluation. **Applied Social Research Methods Series**, 38, Sage Publications, Thousand Oaks, Californie.
- Gower, A.R. (1994). Conception des questionnaires d'enquêtes-entreprises. **Techniques d'enquête**, 20, 129-141.
- Nelson, D.D. (1985). Informal Testing as a Means of Questionnaire Development. **Journal of Official Statistics**, 1, 179-188.
- 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
- Statistique Canada (1994c). Politique sur l'élaboration, l'essai et l'évaluation des questionnaires. **Manuel des politiques**, 2.8. (Reproduite à l'annexe 1).
- Statistique Canada (1996a). Politique d'information des répondants aux enquêtes. **Manuel des politiques**, 1.1.

2.6 Réponse et non-réponse

Portée et objet

Même si les chargés d'enquête et le personnel des opérations font tout ce qu'ils peuvent pour maximiser le taux de réponse, il est presque assuré qu'il y aura une certaine proportion de non-réponse. Pour qu'on considère que l'unité a répondu à la question, le degré de *réponse partielle* (le répondant a fourni une réponse suffisamment précise uniquement pour certains des éléments d'information demandés) doit être égal ou supérieur à un certain seuil en-deçà duquel la réponse serait rejetée et considérée comme une *non-réponse partielle* (on juge alors que la personne, le ménage, l'entreprise, l'institution, l'exploitation agricole ou une autre unité de l'échantillon n'a pas répondu).

La non-réponse a deux répercussions sur les données : elle contribue à augmenter la variance d'échantillonnage des estimations du fait que la taille réelle de l'échantillon est réduite par rapport à ce qu'on recherchait au départ; elle accroît le *biais* des estimations lorsque les non-répondants se distinguent des répondants dans les caractéristiques mesurées.

Principes

Les efforts déployés aux fins de l'obtention des réponses varient en fonction des contraintes budgétaires et des contraintes de temps, de l'ajustement pour des besoins d'utilisation ainsi que du risque de biais de non-réponse. Par la suite, on rajuste les données afin de compenser la non-réponse (p. ex. ajustement des poids ou imputation). Les décisions prises relativement au degré approprié de recherche effectuée aux fins de la mise au point des techniques de rajustement sont également influencées par le budget, le temps, l'utilisation des données et le risque de biais. Dans le cadre des mesures générales prises pour contrôler la qualité des données, on assure un suivi de la non-réponse afin que le personnel d'enquête puisse réagir immédiatement et ultérieurement, et on communique le taux de non-réponse aux utilisateurs des données de l'enquête.

Un programme efficace de relations avec les répondants ainsi qu'un questionnaire bien conçu s'avèrent essentiels si l'on veut maximiser le taux de réponse (voir la section 2.5).

Lignes directrices

- Un bon *taux de réponse* s'obtient en partie en assurant une qualité suffisante et en satisfaisant les objectifs à toutes les étapes de la conception et de la mise en oeuvre d'une enquête. Pour obtenir le taux de réponse désiré, tenez compte de facteurs tels que :
 - la qualité de la base de sondage (relativement à la couverture de la population, à la facilité à établir un contact avec le répondant);
 - la population-cible;

- la méthode de collecte (par exemple, par la poste, interview sur place, interview téléphonique, interview assistée par ordinateur);
 - la méthode d'échantillonnage;
 - la période de l'année et la durée de la période de collecte;
 - le fardeau de réponse (durée de l'entrevue, difficulté du sujet, enquête répétée);
 - la nature du sujet (sujets délicats);
 - la longueur et la complexité du questionnaire;
 - les méthodes de suivi (le cas échéant);
 - les difficultés qu'on prévoit rencontrer afin de retrouver les répondants qui ont déménagé (le cas échéant);
 - les résultats obtenus par le passé avec ce genre d'enquête;
 - l'expérience et les aptitudes manifestes du personnel affecté à la collecte;
 - la charge de travail du personnel affecté à la collecte;
 - les relations établies avec le répondant;
 - la stratégie de communications;
 - le budget total;
 - la répartition 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 relatifs aux intervieweurs, p. ex. la formation, l'expérience, l'entregent, l'établissement de relations et le roulement du personnel;
 - le recours à des mesures pour inciter les répondants à participer et leur efficacité.
- Le test préliminaire peut être efficace aux fins de la vérification, entre autres, du taux de réponse escompté.
 - Dans la mesure où les contraintes opérationnelles le permettent, suivez les non-répondants (suivi de l'ensemble des non-répondants ou d'un sous-échantillon). Un tel suivi permet d'augmenter le taux de réponse et peut nous aider à vérifier dans une certaine mesure si les répondants et les non-répondants sont similaires dans les caractéristiques mesurées. Le suivi des non-répondants est particulièrement important au moment de la mise en oeuvre d'une enquête longitudinale; en pareil cas, on investit manifestement à plus long terme, et les activités de suivi sont particulièrement importantes.
 - Établissez la priorité des activités de suivi. Par exemple, dans les enquêtes-entreprises, assurez d'abord un suivi des grandes unités ou des unités influentes, au risque d'oublier des unités plus petites (voir la section 2.8). De même, donnez la priorité aux unités non répondantes qui appartiennent à des domaines associés à un risque élevé de biais de non-réponse.

- Vérifiez et inscrivez la cause de la non-réponse (p. ex. refus, aucun contact, absence temporaire, problème technique).
- Parce que les différences entre les répondants et les non-répondants peuvent se traduire par un biais dans les estimations, il est important de tenter de déterminer si de telles différences existent. Bien que cela soit difficile, on peut y arriver en partie en effectuant un appariement avec des sources de données externes (p. ex. des dossiers administratifs), en partie en examinant les réponses des non-répondants qui ont été convertis au terme d'un suivi. En général, il est cependant plus facile de comparer les caractéristiques des répondants et des non-répondants pour voir dans quelle mesure il existe des différences au niveau des caractéristiques connues. L'information ainsi obtenue peut avoir un impact sur les méthodes de compensation de la non-réponse.
- Il existe généralement deux façons de compenser la non-réponse (Kalton et Kasprzyk, 1986) : en rajustant les poids d'échantillonnage ou en se servant d'une technique d'imputation. Dans la mesure du possible, évaluez dans quelle mesure les méthodes de compensation de la non-réponse corrigent les biais éventuels.
- Rapportez les taux de réponse et de non-réponse. À Statistique Canada, des normes et des lignes directrices ont été établies aux fins de la déclaration de la non-réponse (Statistique Canada, 1992c, 1993c). Tentez de vous conformer à ces normes afin d'accroître la comparabilité des enquêtes. Dans les lignes directrices, l'unité doit être classée comme répondante ou non répondante. Indiquez clairement s'il y a des « non-répondants partiels » et comment ils sont classés. Les normes existantes devraient cependant être modifiées afin d'être mieux adaptées aux enquêtes longitudinales et à certaines enquêtes-entreprises.

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.
- Kalton, G. et Kasprzyk, D. (1986). Le traitement des données d'enquête manquantes. **Techniques d'enquête**, **12**, 1-17.
- Statistique Canada (1992c). 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 (1993c). Standards and Guidelines for Reporting of Nonresponse Rates. 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.

2.7 Opérations de collecte de données

Portée et objet

La *collecte de données* désigne tout processus visant à acquérir ou à faciliter l'acquisition de données. On procède à la collecte en demandant à des particuliers ou à des organisations de fournir des renseignements pertinents au moyen d'un instrument approprié (voir la section 2.5). Si on n'obtient aucun renseignement à la première demande ou si les données s'avèrent inadéquates au moment de la vérification préliminaire, on peut relancer les répondants dans le cadre de la collecte des données (voir la section 2.8). L'information est validée et convertie dans un format électronique qui convient aux étapes subséquentes. Souvent, à l'étape de la conversion, on procède au *codage* des données, et parfois, on les transmet à un autre endroit. L'impact des opérations de collecte sur la qualité des données est à la fois direct et crucial du fait que les données recueillies constituent la matière première de l'organisme d'enquête. Par conséquent, la qualité du produit final est tributaire de celle des données.

Principes

Les répondants ou les fournisseurs de données, tout notamment les particuliers et les organisations qui remplissent les questionnaires sans être rémunérés, représentent la ressource la plus précieuse pour l'organisme d'enquête. Si l'on veut qu'ils continuent à collaborer, il est essentiel de minimiser le fardeau de réponse. Le meilleur moyen de corriger les lacunes ou les incohérences statistiques consiste à consulter les répondants durant la collecte ou peu après. Compte tenu de l'impact considérable de la collecte sur la qualité du produit final, il convient d'utiliser des outils appropriés d'évaluation de la qualité et des résultats afin de gérer le processus et de fournir des mesures objectives au client (Mudryk et coll., 1996). Tout au long du processus, il faut prendre les mesures qui s'imposent pour préserver la confidentialité des renseignements recueillis (voir la section 2.13).

Lignes directrices

- Le rôle des intervieweurs exerce une influence déterminante sur le succès de la plupart des opérations de collecte de données. S'assurer que leur formation et leurs outils sont adéquats.
- Tirer parti de la technologie disponible afin d'accroître l'efficacité et d'améliorer la qualité du processus. La collecte est souvent l'étape la plus coûteuse du processus d'enquête et a une grande incidence sur la qualité des données. Tant que la technologie des communications et la technologie informatique progresseront à grands pas, on peut s'attendre à ce que de nouvelles occasions se présentent afin de réduire considérablement les coûts et les risques associés à la collecte ou d'accroître le taux de réponse, l'exactitude et l'actualité. L'interview assistée par ordinateur

(IAO), le système de déclaration électronique des données (SDED), la saisie des données (DC2, CASES), la saisie automatisée de données (SAD) et le codage automatisé par reconnaissance de texte (CART) sont des exemples de nouvelles approches qui mettent à profit les technologies disponibles.

- Dans le cas des activités plus complexes accomplies par les humains, par exemple les interviews, les vérifications manuelles, le codage et la saisie des données, on recommande des méthodes statistiques de contrôle de la qualité afin d'assurer la qualité des opérations de collecte. Ces opérations doivent faire l'objet de comptes rendus destinés aux gestionnaires, aux participants, aux spécialistes et aux méthodologistes. Ces rapports doivent renfermer des renseignements sur la fréquence et les causes des erreurs.
- Pour toutes les opérations de collecte de données, il faut mettre en place des mécanismes appropriés de contrôle de l'échantillon. De tels mécanismes permettent de suivre l'état des unités de l'échantillon du début jusqu'à la fin de la collecte de manière que les gestionnaires responsables de la collecte et les intervieweurs puissent faire le point à tout moment. Les mécanismes de contrôle de l'échantillon et les comptes rendus de ces activités permettent également de s'assurer que chaque unité de l'échantillon franchit toutes les étapes de la collecte et que leur état définitif est consigné.
- Mettre en place des systèmes efficaces de contrôle afin d'assurer la sécurité de la transmission et de la manipulation des données. Il doit être impossible à un intrus d'accéder à l'information; de plus, il faut prévenir les pertes de renseignements (et la baisse de qualité qui en découle) attribuables à des défaillances du système ou à des erreurs humaines durant la transmission.
- Afin d'optimiser les taux de réponse et la qualité des renseignements recueillis auprès des entreprises et des institutions, s'assurer qu'on communique avec la bonne personne au sein de l'organisation, au moment adéquat, et d'avoir une méthode et un format de données qui soient pratiques pour le répondant, afin que l'information puisse être obtenue facilement.
- Lorsqu'on conçoit les opérations de collecte de données, tout particulièrement la vérification et le codage, il faut employer les mêmes méthodes avec l'ensemble des unités à l'étude et, dans la mesure du possible, s'assurer qu'elles sont exemptes d'erreur. L'automatisation est souhaitable. Il peut également être utile de permettre au personnel ou aux systèmes de soumettre les dossiers épineux à un nombre restreint d'experts compétents. De même, il est avantageux de centraliser le traitement, car cela réduit les coûts et simplifie le recours aux experts disponibles. Comme il peut y avoir des résultats imprévus dans les données recueillies, les opérations doivent être adaptables de manière qu'on puisse apporter les modifications nécessaires au nom de

l'efficacité.

- Le chargé d'enquête a besoin des données sur les dépenses et d'autres mesures du rendement et de la qualité recueillies durant le processus pour décider si le processus doit être remanié ou modifié. Les coûts réels de l'affranchissement, des appels téléphoniques, de la production des instruments de collecte, du traitement informatique et de l'utilisation des jours-personnes doivent faire l'objet d'un suivi. Parmi les principales mesures de la qualité, mentionnons le taux de réponse, le taux de suivi et la répartition de la non-réponse selon la cause. Ces mesures doivent être disponibles à tous les niveaux pour lesquels des estimations sont produites et aux diverses étapes du processus et peuvent être utilisés aux fins de la mesure du rendement et de la qualité des données (voir la section 2.6). La fréquence des rejets à l'étape de la vérification ainsi que le nombre et le type de corrections apportées aux données selon la strate, la méthode de collecte, le type de traitement, l'élément d'information et la langue de l'instrument de collecte peuvent être comparés au nombre prévu de rejets et de corrections aux fins du suivi de la qualité des données et de l'efficacité de la fonction de vérification. On peut recourir aux mesures de la qualité et de la productivité afin de donner des commentaires au personnel affecté au traitement ou aux intervieweurs ainsi que pour cerner les éléments à l'origine des erreurs dans la conception de l'instrument de collecte ou les méthodes de traitement.

- Les étapes subséquentes de l'enquête peuvent fournir des renseignements utiles sur la qualité, lesquels peuvent révéler que les méthodes ou les outils de collecte doivent être modifiés dans les prochains cycles de l'enquête. Par exemple, les étapes de la vérification ou de l'analyse des données (voir les sections 2.8 et 2.15) peuvent laisser entendre qu'il existe un risque de biais de réponse ou que d'autres problèmes liés à la collecte peuvent se poser. Par conséquent, il convient d'obtenir des commentaires auprès des personnes affectées aux étapes subséquentes de l'enquête.

Références

- 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**. John Wiley & Sons, 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., Biemer, P., Lyberg, L., Massey, J., Nicholls, W. et Waksberg, J., (éd.) (1988). **Telephone Survey Methodology**. Wiley, New York.
- Groves, R.M. (1989). **Survey Errors and Survey Costs**. New York: John Wiley.
- 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., Burgess, M.J. et Xiao, P. (1996). Quality Control of CATI Operations in Statistique Canada. **Proceedings of the Section on Survey Research Methods**, American Statistical Association, 150-159.
- Mudryk, W., Croal, J. et Bougie, B. (1994). Generalized Data Collection and Capture (DC2): Release 2.5.1, Sample Verification (SV). Rapport technique de Statistique Canada.
- Mudryk, W. et Xiao, P. (1996). Quality Control Methodology for LFS Industry and Occupation Coding Operations. 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 englobe les mesures visant à repérer les inscriptions manquantes, incorrectes ou incohérentes ainsi que les enregistrements susceptibles de renfermer des erreurs. Certaines de ces mesures mettent en relief des relations logiques qui découlent directement des concepts et des définitions. D'autres sont plus empiriques ou sont le fruit du recours à des méthodes ou à des tests statistiques (p. ex. les techniques d'analyse des valeurs aberrantes). Les données recueillies antérieurement dans le cadre de la même enquête ou provenant d'autres sources peuvent également être utilisées.

La vérification englobe un large éventail d'activités allant des vérifications sur le terrain effectuées par les intervieweurs, des avertissements produits par ordinateur au moment de la collecte ou de la saisie des données, en passant par l'identification des unités qui feront l'objet d'un suivi, jusqu'aux vérifications des relations complexes, à la localisation des erreurs pour fins d'imputation et à la validation des données. Les deux dernières activités seront abordées dans les sections 2.9 et 2.12.

Principes

La vérification comporte trois objectifs (Granquist, 1984) : jeter les bases des améliorations qui seront apportées à l'instrument d'enquête; fournir des renseignements sur la qualité des données d'enquête et épurer ces dernières. Il y a tout lieu de croire que les ressources affectées à la réalisation du troisième objectif sont disproportionnées. Par conséquent, l'apprentissage émanant du processus de vérification joue souvent un rôle indûment secondaire.

Bien qu'on reconnaisse que les *erreurs fatales* (p. ex. inscriptions erronées ou incohérentes) doivent être supprimées des fichiers de données afin de préserver la crédibilité du Bureau et faciliter l'automatisation du traitement et l'analyse de l'information, il faut éviter d'utiliser à outrance les *vérifications exploratoires* (qui sont axées sur les enregistrements susceptibles d'être erronés). La vérification des données constitue vraisemblablement l'activité la plus onéreuse du cycle de l'enquête-échantillon ou du recensement : d'après les estimations, elle représente jusqu'à 40 % du budget total des enquêtes-entreprises (Gagnon, Gough et Yeo, 1994). La survérification se traduit non seulement par un accroissement des ressources financières, de la période de traitement et du fardeau de réponse, mais elle peut également entraîner de graves biais lorsque les données sont rajustées en fonction de modèles implicites imposés par les vérifications.

Lignes directrices

- Assurer la cohérence interne des règles de vérifications (c.-à-d. voir à ce qu'elles ne

soient pas contradictoires).

- Vérifier de nouveau les unités qui ont fait l'objet de corrections pour s'assurer que d'autres erreurs n'ont pas été introduites directement ou indirectement.
- La vérification convient tout à fait aux fins de la détection des erreurs fatales (Granquist et Kovar, 1997), comme cette tâche peut facilement être automatisée. Cette activité doit être accomplie aussi rapidement que possible. Bien qu'elle puisse nécessiter un traitement manuel, un logiciel généralisé et réutilisable est particulièrement utile à cette fin.
- Ce sont généralement les vérifications exploratoires (Granquist et Kovar, 1997) qui font grimper les coûts de la vérification. Avec cette forme de vérification, il faut exécuter certaines tâches manuellement et assurer un suivi des répondants, et la plupart du temps, les données ne sont guère modifiées ou ne le sont pas du tout. Ce type de vérification doit être rationalisé en cela qu'on doit mettre en équilibre la détection des erreurs et les coûts.
- Les *taux de succès* des vérifications (la proportion d'avertissements qui mettent en lumière des erreurs réelles) se sont avérés faibles, se situant souvent dans les 20 à 30 % (Linacre et Trewin, 1989). En outre, on a constaté que l'impact des erreurs variait beaucoup, tout particulièrement dans les enquêtes qui recueillent des données numériques. En d'autres termes, il n'est pas rare que quelques erreurs soient responsables de la plupart des modifications. Par conséquent, une vérification sélective pourrait accroître l'efficacité (Granquist et Kovar, 1997), sans que la qualité des données n'en souffre. On pourrait établir les priorités en fonction du type ou de la gravité des erreurs, ou encore de l'importance de la variable ou de l'unité déclarante.
- La vérification est d'une utilité restreinte, et, en fait, peut aller à l'encontre du but recherché (voir, par exemple, Linacre et Trewin, 1989). À tort, on considère souvent les modifications apportées aux données comme des corrections. On peut soutenir que, à un certain moment du processus de vérification, on commet autant d'erreurs qu'on en corrige. Il faut conjuguer les efforts afin de repérer et de respecter cette fin logique du processus.
- L'automatisation permet aux chargés d'enquête d'accroître la portée et le volume des vérifications qui peuvent être effectuées. Il faut résister à cette tentation. En d'autres termes, certains enregistrements doivent satisfaire à toutes les règles de vérifications. Au lieu d'intensifier l'effort de vérification, il faut rediriger les ressources vers des activités plus rentables (p. ex. analyse des données, analyse des erreurs de réponse, etc.).

- Il faut s'efforcer de moins compter sur la vérification pour corriger les problèmes après coup, notamment dans les enquêtes répétées. Sa contribution à la réduction des erreurs est limitée. La vérification est essentielle; toutefois, il convient d'en réduire la portée et de s'en servir à une autre fin. De plus, il convient de donner plus d'importance aux leçons qu'on peut tirer du processus de vérification. De manière à réduire les erreurs, il faut regarder en amont plutôt que d'apporter des corrections à la fin. Il faut s'employer à prévenir les erreurs et non à les corriger. Pour atteindre cet objectif, la vérification doit être effectuée dans les premières étapes du processus d'enquête, de préférence lorsqu'on peut encore rejoindre le répondant, par exemple, en ayant recours à l'interview assistée par ordinateur, à l'interview sur place ou au questionnaire d'auto-dénombrement.
- Les vérifications ne permettent pas de déceler les petites erreurs systématiques qui sont constamment signalées dans les enquêtes répétées, des erreurs qui peuvent se traduire par d'importants biais dans les estimations. Une vérification plus stricte n'est pas la solution. On doit plutôt recourir à d'autres méthodes, par exemple, les méthodes classiques de contrôle de la qualité, l'analyse minutieuse des concepts et des définitions, les études post-interviews, etc.
- Ne pas surestimer la capacité des répondants de fournir des renseignements au moment du suivi. Il se peut que leur agrégation soit différente, que leur mémoire ait des failles et que leur « récompense » soit négligeable. On recommande donc de restreindre les activités de suivi auprès des répondants.
- Ne pas sous-estimer la capacité des vérificateurs de rajuster les données déclarées en fonction des modèles imposés par les règles de vérifications. Il existe un danger réel d'effectuer à tort des modifications uniquement pour s'assurer que les données ne soient pas rejetées à la vérification. Il faut contrôler le processus!
- Le processus de vérification est souvent très complexe. Lorsque ce processus relève de l'organisme de statistique, le personnel affecté à cette tâche doit être formé adéquatement et utiliser des méthodes détaillées et actuelles; de plus, le travail doit faire l'objet d'un suivi approprié. Des mécanismes de contrôle de la qualité en bonne et due forme doivent être envisagés.
- La vérification peut s'avérer utile en épurant une partie des données. Cependant, elle est encore plus utile en fournissant des renseignements sur le processus d'enquête, que ce soit pour mesurer la qualité de l'enquête ou proposer des améliorations en vue des enquêtes subséquentes. La vérification doit être considérée comme un élément qui fait partie intégrante de la collecte de données et qui permet de recueillir de l'information sur le processus. À ce titre, la vérification peut être d'une aide inestimable en précisant les définitions, en améliorant l'instrument d'enquête, en évaluant la qualité des données, en détectant les sources d'erreurs non dues à l'échantillonnage, en jetant les bases des améliorations qui seront apportées au

processus d'enquête et en alimentant le cycle d'apprentissage continu. Pour atteindre cet objectif, assurer un suivi du processus et créer des pistes de vérification, des diagnostics et des mesures du rendement de manière à déterminer les meilleures approches.

Références

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? In **Survey Measurement and Process Quality**, Lyberg et coll. (éd.), Wiley & Sons, New York, 415-435.

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.

2.9 Imputation

Portée et objet

L'*imputation* est le processus visant à remédier aux réponses manquantes, non valides ou incohérentes détectées lors de la vérification. Cela consiste à modifier certaines des réponses ou des valeurs manquantes de l'enregistrement vérifié afin d'en assurer la plausibilité et la cohérence interne. On arrive à régler certains problèmes plus tôt en communiquant avec le répondant ou en procédant à une étude manuelle du questionnaire. En général, il est impossible de résoudre tous les problèmes dans les premiers stades de l'enquête en raison des préoccupations relatives au fardeau de réponse, aux coûts et à l'actualité des données. On a ensuite recours à l'imputation pour traiter le reste des rejets à la vérification comme on souhaite produire un fichier complet et cohérent renfermant des données imputées. Il est préférable que l'imputation soit effectuée par les personnes ayant un accès complet aux microdonnées et disposant de données auxiliaires de qualité.

Bien que l'imputation permette d'améliorer la qualité des données finales en compensant les réponses manquantes, non valides ou incohérentes, veiller à choisir la méthode d'imputation appropriée. Certaines de ces méthodes ne préservent pas le lien existant entre les variables ou, en fait, peuvent fausser les distributions sous-jacentes.

Principes

L'imputation peut être automatisée et/ou manuelle. Une imputation adéquate restreint le biais attribuable au fait que les valeurs désirées n'ont pas toutes été observées; de plus, grâce à l'imputation, on dispose d'une analyse rétrospective aux fins de l'évaluation et on s'assure de la cohérence interne des enregistrements imputés. Un processus d'imputation adéquat est automatisé, objectif, reproductible et efficient. Suivant les principes énoncés par Fellegi-Holt (1976), on modifie le moins de champs possible pour s'assurer que l'enregistrement terminé franchisse avec succès toutes les vérifications.

Les méthodes d'imputation sont déterministes ou stochastiques, selon que les données imputées aient ou non un caractère aléatoire (Kalton et Kasprzyk, 1986; Kovar et Whitridge, 1995). Parmi les méthodes d'*imputation déterministe*, on retrouve l'imputation logique, l'imputation historique, l'imputation par la moyenne, l'imputation par quotient et par régression et (suivant une certaine interprétation) l'imputation par voisin le plus proche. Ces méthodes peuvent également être subdivisées, selon qu'elles consistent uniquement à inférer la valeur imputée à partir des données dont on dispose sur le non-répondant et d'autres données auxiliaires (logiques et historiques) ou qu'elles utilisent les observations se rapportant à d'autres unités déclarantes dans le cadre d'une enquête donnée. On peut utiliser directement les observations courantes en transférant les données d'un enregistrement d'un donneur choisi (méthode hot deck et imputation par voisin le plus proche) ou au moyen de modèles (imputation par quotient et par régression). Parmi les méthodes d'*imputation stochastique*,

mentionnons la méthode hot deck, la régression avec résidus aléatoires et toute autre méthode déterministe employant des résidus aléatoires.

Lignes directrices

- Le dossier imputé devrait ressembler beaucoup au dossier rejeté à la vérification. À cette fin, réduire au minimum le nombre de variables imputées dans une certaine mesure de manière à préserver autant que possible les données du répondant. Cela repose sur l'hypothèse voulant que le répondant soit plus susceptible de ne commettre qu'une ou deux erreurs que plusieurs bien qu'il n'en soit pas toujours ainsi dans la pratique.
- Il faudrait réduire au minimum le nombre de donneurs lorsqu'on impute des données à l'égard d'un enregistrement. De plus, selon la disponibilité des variables prédictives, l'enregistrement imputé devrait ressembler beaucoup aux donneurs choisis. D'un point de vue opérationnel, cela peut signifier un donneur par section du questionnaire, comme il est presque impossible de traiter simultanément toutes les variables d'un long questionnaire.
- Afin de réduire les coûts de calcul inhérents aux enquêtes de grande envergure, deux passages ou plus peuvent être nécessaires pour traiter les variables. Ce faisant, il est plus difficile de respecter à la lettre les lignes directrices : on peut utiliser plusieurs donneurs et le nombre de variables imputées peut être supérieur au nombre minimal. Si un enregistrement renferme de nombreuses erreurs de réponse, il peut être nécessaire d'excéder le nombre minimal de variables imputées, que l'imputation soit automatisée ou manuelle.
- Il faut que, d'après les donneurs disponibles, les imputations équivalentes aient une chance appropriée d'être choisies pour ne pas exagérer faussement les groupes importants de la population.
- Il faudrait s'assurer la cohérence interne des enregistrements imputés.
- Signaler les valeurs imputées et identifier clairement les méthodes et les sources. Conserver pour évaluation les valeurs non imputées et imputées des champs de l'enregistrement. Évaluer le degré et les effets de l'imputation. Envisager le recours à des techniques afin de mesurer la variance attribuable à l'imputation (Rancourt et coll., 1993). Cette information est exigée en vertu de la politique de Statistique Canada visant à informer les utilisateurs de la qualité des données et la méthodologie (Statistique Canada, 1992).
- Tenir compte du degré et de l'impact de l'imputation au moment de l'analyse des données. Les méthodes d'imputation employées peuvent influencer considérablement sur

les distributions des données. Par exemple, il se peut qu'il n'y ait pas eu beaucoup de changements dans l'ensemble, mais que les valeurs d'un domaine aient augmenté systématiquement, et que cette augmentation ait été annulée par la diminution des valeurs d'un autre domaine. Par ailleurs, même lorsque le degré d'imputation est faible, les modifications apportées aux différents enregistrements peuvent avoir une incidence considérable (p. ex. si les déclarations des grandes unités sont modifiées ou si d'importantes modifications sont apportées aux déclarations de quelques unités). En général, plus le degré et l'impact de l'imputation sont importants, plus l'analyste doit utiliser les données de façon judicieuse. En pareil cas, les analyses peuvent être faussées si les valeurs imputées sont traitées comme des valeurs observées.

- En ce qui a trait à l'élaboration des méthodes d'imputation, il convient de souligner qu'il existe un certain nombre de systèmes généralisés qui exécutent divers algorithmes pour des données continues ou catégoriques. Ces systèmes sont habituellement faciles à utiliser une fois que les vérifications ont été précisées; ils renferment des algorithmes qui déterminent quels champs seront imputés. Ils sont bien documentés et conservent des analyses rétrospectives aux fins de l'évaluation du processus d'imputation. Deux systèmes peuvent actuellement être utilisés à Statistique Canada : le Système généralisé de vérification et d'imputation (SGVI) (Kovar et coll., 1988, Cotton, 1993), pour les variables économiques quantitatives; la méthode d'imputation numérique (MIN) (Bankier et coll., 1994), pour les données catégoriques et certaines données continues du recensement.

Références

- Bankier, M., Fillion, J.-M., Luc, M. et Nadeau, C. (1994). Imputing Numeric and Qualitative Variables Simultaneously. **Proceedings of the Section on Survey Research Methods**, American Statistical Association, 242-247.
- Cotton, C. (1993). Description des fonctions du système généralisé de vérification et d'imputation. Rapport technique de Statistique Canada.
- 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., MacMillan, J. et Whitridge, P. (1988). Système généralisé de vérification et d'imputation - aperçu et stratégie, Statistique Canada, document de travail de la Direction de la méthodologie n° BSMD 88-007 E/F.

Kovar, J.G. et Whitridge, P. (1995). Imputation of Business Survey Data. in **Business Survey Methods**, B.G. Cox et al. (éd.), Wiley, New York, 403-423.

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.

Statistique Canada (1992). Politique visant à informer les utilisateurs de la qualité des données et la méthodologie. **Manuel des politiques**, 2.3.

2.10 Estimation

Portée et objet

L'*estimation* est un processus qui consiste à attribuer des valeurs à des paramètres inconnus qui se rapportent à une population à partir des données d'un ensemble de données. Les paramètres devant être estimés peuvent la plupart du temps être exprimés sous forme de fonctions des effectifs totaux. Mentionnons notamment les statistiques descriptives simples telles que les totaux, les moyennes et les ratios, ainsi que les statistiques analytiques plus complexes, comme par exemple les coefficients de régression. Des mesures de *précision* sont habituellement associées à ces paramètres estimés. Si la qualité des estimations calculées est en grande partie reliée aux étapes précédentes de l'enquête, le choix de la méthode d'estimation peut également être déterminant, tout particulièrement lorsqu'on peut faire un usage judicieux des données auxiliaires.

Principes

Des méthodes d'estimation adéquates produiront des *estimations sans biais par rapport au plan* lorsqu'il n'y a pas de non-réponse. L'*estimateur de Horvitz-Thompson* est un estimateur sans biais fondamental. Il pondère les données au moyen de coefficients de pondération équivalant à l'inverse de la probabilité d'inclusion produite par le plan d'échantillonnage. Un tel coefficient de pondération est appelé *poids d'échantillonnage* et peut être considéré comme le nombre de fois que chaque unité de l'échantillon doit être reproduite pour représenter la population. Il est parfois nécessaire de rajuster les poids d'échantillonnage, notamment afin de compenser pour la non-réponse ou pour intégrer les données auxiliaires afin d'accroître la précision de l'estimateur. L'intégration des données auxiliaires au processus d'estimation est désignée sous le terme de *calibrage*, et les facteurs qui en découlent sont appelés *facteurs de calibrage*. L'utilisation des facteurs de calibrage (rajustés en fonction de la non-réponse) avec les poids d'échantillonnage engendre les *poids de calibrage*. Ces poids assurent l'intégration du plan d'échantillonnage et des données auxiliaires. En général, de tels rajustements se traduisent par des estimations qui sont presque sans biais par rapport au plan; cependant, le biais est habituellement négligeable. La variance estimée des estimateurs ainsi obtenus doit témoigner du plan d'échantillonnage et de l'utilisation des données auxiliaires.

Lignes directrices

- L'estimation doit être conforme au plan d'échantillonnage. À cette fin, les poids d'échantillonnage doivent être intégrés au processus d'estimation. Cela suppose que l'estimation fasse état des aspects du plan d'échantillonnage tels que la stratification, la répartition en grappes, les données liées aux degrés ou aux phases multiples.
- Dans le cas des enquêtes longitudinales, deux ensembles de poids d'estimation sont

habituellement fournis : les poids longitudinaux et les poids transversaux. Les poids longitudinaux sont associés à la population lors du tirage initial de l'échantillon longitudinal. En général, ces poids sont rajustés en fonction de l'attrition de l'échantillon dans le temps. On se sert des poids longitudinaux dans le cadre de l'analyse des données longitudinales. Les poids transversaux sont liés à la population en date de chaque cycle d'enquête. On a habituellement recours à ces poids afin de produire des estimations ponctuelles ou de déterminer les variations des estimations ponctuelles entre deux périodes. Comme la population change dans le temps, les poids transversaux sont généralement différents des poids longitudinaux.

- Les rajustements visant à compenser la non-réponse peuvent biaiser les estimations si les unités déclarantes et les unités non déclarantes appartenant à la même classe de rajustement n'ont pas le même comportement. Même si les unités déclarantes et les unités non déclarantes se comportent de la même façon, les méthodes de correction de la non-réponse comme telles (repondération et imputation) peuvent également entraîner un biais. S'efforcer de réduire l'impact de ces biais.
- Lorsque la classification originale des unités d'échantillonnage a changé entre le tirage de l'échantillon et l'estimation, il faut envisager de recourir à l'*estimation par domaine* afin que les estimations témoignent de la nouvelle classification. L'estimation par domaine consiste à produire des estimations pour certains sous-ensembles de la population (ou *domaines* d'intérêt). Souvent, les éléments qui font partie de ces sous-ensembles n'ont pas été ou n'auraient pas pu être identifiés avant l'échantillonnage. En général, l'estimation par domaine consiste à attribuer une valeur nulle aux caractéristiques des unités échantillonnées qui ne font pas partie du domaine d'intérêt en question; les autres caractéristiques conservent la valeur mesurée qui leur a été attribuée au départ.
- Se fonder sur le plan d'échantillonnage et la répartition de l'échantillon afin de satisfaire aux exigences des petits domaines d'intérêt (Singh, Gambino et Mantel, 1992). Si cela s'avère impossible à l'étape de la conception, il faut envisager des méthodes d'estimation spéciales (des *estimateurs pour petits domaines*) à l'étape de l'estimation. Ces méthodes « puisent » dans les secteurs (ou domaines) connexes afin de minimiser l'erreur quadratique moyenne de l'estimateur (Ghosh et Rao, 1994).
- Fournir les erreurs-types ou les coefficients de variation estimés pour les estimations ponctuelles comme mesures de la précision. Au besoin, les intervalles de confiance peuvent être fournis. Les mesures de la précision doivent témoigner du plan d'échantillonnage et de toute correction apportée aux données, telle les ajustements afin de compenser pour la non-réponse (Rao, 1996; Gagnon et coll., 1996). Au besoin, il faut envisager de recourir à l'estimation par domaine.

- On attribue une valeur nulle aux unités disparues de l'échantillon durant le processus d'estimation. Les variances estimées seront plus élevées du fait que des valeurs nulles entrent dans le calcul.
- Dans la mesure du possible, utiliser les données auxiliaires afin d'accroître la fiabilité des estimations. Évaluer l'utilisation de ces données. À cette fin, on peut recourir à l'exploration, par exemple au moyen du Système généralisé d'estimation de Statistique Canada, qui repose sur des techniques de rajustement par régression. Les données auxiliaires peuvent être employées de diverses façons dans un même ensemble de données, c'est-à-dire que différentes variables peuvent être utilisées dans un modèle de régression, ce qui se traduit par des estimateurs distincts.
- Chaque fois qu'on dispose de données auxiliaires et d'effectifs totaux connus, il peut être utile de recourir à l'estimation par calibrage de manière à ce que les données auxiliaires pondérées correspondent à ces effectifs totaux connus. Cela peut accroître la précision et la cohérence des estimations produites à partir de diverses sources. Tenter de restreindre l'étendue des coefficients de pondération issus du calibrage. Une grande hétérogénéité des coefficients de pondération peut accroître la variance des estimations produites, et, par conséquent, diminuer leur précision. On peut réduire l'étendue des coefficients de pondération en les délimitant comme dans Huang et Fuller (1978) ou dans Deville et Särndal (1992). On peut également utiliser ces méthodes de délimitation pour éviter que les coefficients de pondération soient négatifs ou excessivement élevés.
- Lorsqu'il y a des valeurs aberrantes, il n'est pas rare que les estimations produites pour des variables continues ne sont pas fiables. Les valeurs aberrantes peuvent s'expliquer par des valeurs extrêmes mesurées à l'égard de certaines caractéristiques et/ou par des coefficients de pondération élevés liés aux éléments aberrants. Lorsque les valeurs aberrantes sont attribuables à des valeurs extrêmes mesurées pour certaines caractéristiques, on recommande de recourir à des méthodes objectives telles que les estimateurs (robustes) à l'épreuve des valeurs aberrantes (Chambers, 1986). Lorsque des coefficients de pondération élevés sont à l'origine des valeurs aberrantes, il convient de réduire ces coefficients en utilisant des méthodes objectives qui influent sur les coefficients (Hidiroglou et Srinath, 1981; Lee, 1995).
- Dans les enquêtes périodiques où les échantillons se chevauchent dans une large mesure d'une édition à l'autre, on doit envisager de recourir à des méthodes d'estimation qui tirent parti de la corrélation dans le temps (Binder et Hidiroglou, 1988; Singh, 1996). L'*estimation composite* compte parmi ces méthodes d'estimation. Fondamentalement, ces méthodes considèrent les données des éditions antérieures comme des variables auxiliaires.

- Dans la mesure du possible, utiliser des logiciels d'estimation généralisés par opposition aux systèmes faits sur mesure. Parmi les progiciels qui peuvent être utilisés, mentionnons le Système généralisé d'estimation (SGE) mis au point par Statistique Canada (Estevao et coll. 1995), SUDAAN (Shah, et coll., 1989), PC CARP (Schnell et coll., 1988) ou WesVar PC (Brick et coll., 1997). En utilisant des systèmes généralisés, les erreurs de programmation devraient être moins nombreuses, et les coûts et délais de développement plus faibles.

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.M., Broene, P., James, P. et Severynse, J. (1997). **A User's Guide to WesVar PC**. Westat Inc., Maryland.
- Chambers, R.L. (1986). Outlier Robust Finite Population Estimation. **Journal of the American Statistical Association**, **81**, 1063-1069.
- Deville, J.-C. et Särndal, C.E. (1992). Calibration Estimators in Survey Sampling. **Journal of the American Statistical Association**, **87**, 376-382.
- 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.
- Gagnon, F., Lee, H., Rancourt, E. et Särndal, C.E. (1996). Estimating the Variance of the Generalized Regression Estimator in the Presence of Imputation for the Generalized Estimation System. **Proceedings of the Survey Methods Section**, Société statistique du Canada, 151-156.
- Ghosh, M. et Rao, J.N.K. (1994). Small Area Estimation: An Appraisal. **Statistical Science**, **9**, 55-93.
- Hidirolou, M.A. et Srinath, K.P. (1981). Some Estimators of Population Total Containing Large Units. **Journal of the American Statistical Association**, **47**, 663-685.
- Huang, E. et Fuller, W.A. (1978). Nonnegative Regression Estimation for Sample Survey Data. **Proceedings of the Social Statistics Section**, American Statistical Association, 330-305.
- Holt, D. et Smith, T.M.F. (1979). Post-Stratification. **Journal of the Royal Statistical Society**, A 142, 33-46.

- Lee, H. (1995). Outliers In Business Surveys. In **Business Survey Methods**, B.G. Cox et coll. (éd.), New York: John Wiley and Sons, 503-526.
- Lemaître, G. et Dufour, J. (1987). Une méthode intégrée de pondération des personnes et des familles. **Techniques d'enquête**, 13, 211-220.
- Platek, R., Rao, J.N.K., Särndal, C.E. et Singh, M.P. (éd.) (1987). **Small Area Statistics**. John Wiley and Sons, New York.
- Rao, J.N.K. (1996). On the Estimation with Imputed Survey Data. **Journal of the American Statistical Association**, 91, 499-506
- 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.
- Singh, A.C. (1996). Combining Information in Survey Sampling by Modified Regression. **Proceedings of the Section on Survey Research Methods**, American Statistical Association, 120-129.
- Singh, M.P., Gambino, J. et Mantel, H. (1992). Issues and Options in the Provision of Small Area Data. **Small Area Statistics and Survey Designs**, G. Kalton, J. Kordos et R. Platek, (éd.). Bureau central de la statistique, Varsovie, 31-75.

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. On obtient ainsi une série désaisonnalisée constituée des composantes tendance-cycle et 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 sous-jacents de la tendance-cycle. 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.

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, c'est-à-dire lorsqu'il n'est pas obscurci par un degré élevé de fluctuations irrégulières qu'on retrouve dans la série (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 estimations produites doivent faire l'objet de révisions minimales du fait que d'autres données sont ajoutées à la fin de la série. 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. 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).

Lorsqu'on produit des estimations de la tendance-cycle afin de compléter la série désaisonnalisée, 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.

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.

Lignes directrices

- Analysez en profondeur une série que vous désaisonnalez pour la première fois afin de vérifier que la saisonnalité est identifiable.
- Utilisez les extrapolations ARMMI dans le calcul des facteurs saisonniers en fin de série pour réduire l'ampleur des révisions. Laissez le programme X11ARMMI choisir automatiquement un modèle ARMMI dans la mesure du possible. Si aucun modèle n'est choisi, on recommande à l'utilisateur d'en fournir un.
- Une série doit être redésaisonnalisée chaque fois qu'une nouvelle observation s'ajoute. Désaisonnalez la dernière observation en utilisant le facteur saisonnier actualisé obtenu à partir de toutes les observations disponibles (Dagum, 1987). Il est possible de déroger à la présente ligne directrice lorsque les observations les plus récentes ont fait l'objet de révisions historiquement importantes. En pareil cas, utilisez les facteurs saisonniers prévus reposant sur des données se terminant à la fin de l'année précédente. Ceux-ci peuvent être plus appropriés (Morry, 1992).
- Ne réviser l'observation désaisonnalisée à l'aide d'un facteur saisonnier actualisé qu'une seule fois au cours de l'année lorsque l'observation suivante est ajoutée à la série. Cependant, dans le cas d'observations préliminaires, réviser toutes les observations désaisonnalisées chaque fois que les observations originales correspondantes sont modifiées substantiellement. Réviser annuellement les observations désaisonnalisées des trois dernières années lorsque la première observation de l'année suivante est disponible (Dagum, 1987). Par contre, réviser 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 fournis automatiquement par le programme X11ARMMI. Pendant l'année en cours, gardez ces coefficients constants en les utilisant en tant que poids journaliers *a priori*. Réviser les coefficients d'ajustement 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. Désaisonnalisez l'agrégat indirectement en combinant les composantes désaisonnalisées et les composantes qui ne contiennent aucune saisonnalité. Désaisonnalisez l'agrégat directement en agrégeant d'abord les séries composantes, puis en désaisonnalisant la série synthétique ainsi obtenue. Par la suite, modifiez les séries composantes désaisonnalisées de telle sorte que la somme de toutes les composantes désaisonnalisées et non-désaisonnalisées soit égale, pour chaque mois ou chaque trimestre, aux valeurs des observations désaisonnalisées correspondantes de la série synthétique ajustée directement. Afin de savoir s'il vaut mieux désaisonnaliser une série synthétique de façon directe ou indirecte; prenez en considération la saisonnalité résiduelle et le degré de lissage produits des deux façons (Lothian et Morry, 1977).
- 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 le choix entre les méthodes de désaisonnalisation directe et indirecte, ainsi qu'entre les facteurs saisonniers actualisés et prévus.
- Utilisez les moyennes mobiles de Henderson, disponibles dans le programme X11ARMMI, pour produire l'estimation de la composante tendance-cycle. Lissez les observations désaisonnalisées déjà publiées afin que la composante tendance-cycle se situe dans le nuage de points des observations désaisonnalisées.
- Extrapolez la série désaisonnalisée sur un horizon d'un an à l'aide d'un modèle ARMMI avant d'estimer la composante tendance-cycle.
- 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" mesurant l'importance relative de la composante irrégulière dans la série (Shiskin et coll., 1967).

- Ne publiez pas l'estimation de la composante tendance-cycle associée à l'observation la plus récente: elle est sujette à d'importantes révisions, voire à un revirement du mouvement lorsque la série est redésaisonnalisée suite à l'ajout d'une nouvelle observation (Dagum et Laniel, 1987).

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 (1993f). Guidelines for Seasonal Adjustment of Current Observations and their Revisions. Document publié par Comité des méthodes et des normes, 10 février.

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 ou à la fiabilité de l'information. Elle permet aux utilisateurs de mieux étayer leur interprétation des résultats d'enquête, et à l'organisme de 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 de 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 de 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.

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 annuler 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

- Il appartient aux gestionnaires de 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.
- Les évaluations de la qualité des données effectuées à Statistique Canada doivent être conçues 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, 1992c), qui figure à 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.
- Il convient de 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 comme l'information que requièrent de telles évaluations doit souvent être recueillie durant le processus d'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, on peut 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 de 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 qui devraient être envisagées, mentionnons :
 - les vérifications de la cohérence par rapport aux sources de données externes, par exemple d'autres enquêtes ou des occasions antérieures de l'enquête;

- 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 leur 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.
- Les sources d'erreurs suivantes peuvent faire l'objet d'une évaluation.
- 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.
 - 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 peut occasionner un biais si les non-répondants et les répondants se distinguent au chapitre de la caractéristique de l'intérêt.
 - Les *erreurs de mesure* se produisent lorsque la réponse fournie diffère de la valeur vraie; 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 introduire un biais dans les résultats.
 - 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 ou un biais.
 - 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). Cette question a été analysée en profondeur 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 in **Répertoire de méthodes d'évaluation des erreurs dans les recensements et les enquêtes**, Publication 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 (1992c). 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 (1995a). Data Quality Criteria in the Dissemination of Statistical Information. Version provisoire de la politique rédigée par le Comité des méthodes et des normes, Statistique Canada.

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 défavorable 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. Un programme vigilant de contrôle de la divulgation et de protection de la confidentialité influe indirectement sur la qualité des données de l'organisme; en effet, les taux de réponse élevés 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) :

- b) 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'aucune donnée personnelle identifiable ne puisse être inférée 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 à cette fin.

Lignes directrices

- Lorsqu'il est question de diffusion, il est nécessaire de distinguer les *données tabulaires* des *microdonnées*. Dans le cas des microdonnées, des enregistrements anonymisés sont établis pour les particuliers alors que pour les données tabulaires, les données sont diffusées sous forme de tableaux statistiques comportant souvent de nombreuses dimensions. Les données tabulaires peuvent également 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) 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. Les mesures qui assurent la protection de la confidentialité de ces divers produits sont forcément très différentes.
- En ce qui a trait aux données quantitatives, les lignes et les colonnes renvoient habituellement à des variables de classification (régions géographiques, branches d'activité, taille, etc.). Elles sont souvent désignées sous le nom de *variables clés* ou de *quasi-identificateurs*, comme elles correspondent à des variables qui peuvent faciliter l'identification d'une unité.
- Un tableau de données quantitatives est jugé confidentiel (c.-à-d. qu'il ne peut pas être diffusé) si une des cellules est confidentielle. Les critères employés pour déterminer si les cellules sont confidentielles reposent sur des règles simples qui sont généralement censées protéger la divulgation des caractéristiques d'un individu.
- Habituellement, on emploie deux critères pour déterminer dans quelle mesure une cellule est confidentielle. 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 ne comprenant que 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.

- De nombreuses règles reposent sur des mesures de la concentration. Les plus faciles à utiliser sont celles qui sont basées sur des combinaisons linéaires des statistiques d'ordre. Une de ces règles couramment utilisées est la *règle (n,k)*. La cellule est confidentielle si les n répondants parmi les plus importants représentent au moins k % de la valeur globale de la cellule. Souvent, plusieurs valeurs de n sont contrôlées (n=1 ou 2). Dans certains cas, on utilise différentes valeurs de k selon le nombre de répondants dans la cellule; cependant, cela n'est pas souhaitable, étant donné que, au terme de l'ajout d'un nouveau répondant ayant un apport négligeable, une cellule confidentielle pourrait perdre son caractère confidentiel, ce qui, intuitivement, n'est pas raisonnable. Cela s'explique par la discontinuité des règles.
- La *règle C-times* mesure également la concentration. Un exemple de cette règle consiste à contrôler le ratio de la valeur de la plus importante unité à la somme des valeurs de la troisième unité en importance et des valeurs des unités de rang inférieur. On cherche ainsi à s'assurer que le deuxième répondant en importance ne puisse pas obtenir une estimation juste de la valeur de la plus importante unité.
- Les cellules à fréquence nulle des tableaux de données quantitatives peuvent également révéler des renseignements confidentiels. Par exemple, il y a divulgation d'information si la cellule révèle que les profits n'excèdent jamais une certaine valeur dans une certaine classe d'entreprise.
- Les cellules confidentielles sont généralement supprimées 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 supprimer des cellules complémentaires afin de protéger la cellule confidentielle. Il existe des logiciels de pointe qui peuvent effectuer cette tâche.
- 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 (p. ex. arrondissement). Récemment, des méthodes qui perturbent uniquement les unités les plus importantes ont été étudiées. Outre l'ajout de bruit aux microdonnées, on peut recourir à toute autre méthode qui assure la sécurité de la base de microdonnées 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. En général, des contrôles doivent être effectués lorsque la distribution d'une ligne ou d'une colonne est concentrée dans un nombre restreint de catégories.
- Par ailleurs, d'autres règles spéciales peuvent s'appliquer aux tableaux de fréquence, tout particulièrement lorsqu'une des lignes ou des colonnes est un groupement d'une variable quantitative (p. ex. un palier de revenus). Par exemple, on peut contrôler la gamme de valeurs éventuelle d'une ligne ou d'une colonne de manière qu'elle soit suffisamment grande.
- Les techniques utilisées aux fins de la réduction du risque de divulgation inhérent aux tableaux de fréquence englobent celles employées avec les tableaux de données quantitatives : la suppression de cellules, la modification des définitions des lignes et des colonnes au terme du regroupement des catégories ou des valeurs extrêmes supérieures de la catégorie; la perturbation des données au terme de l'ajout de bruit aux microdonnées ou aux données tabulaires (p. ex. arrondissement); les autres méthodes qui contribuent à la sécurité de la base de microdonnées servant à la production des tableaux.
- 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. Fondamentalement, aucun individu figurant dans un fichier de microdonnées ne peut être identifié en vertu de la loi. Pour respecter cette exigence, nous créons généralement un ensemble de *variables clés*. Ces variables sont les identificateurs qu'un intrus peut détenir relativement à un individu. Si une combinaison de ces variables mène à un individu (ou à un nombre restreint d'individus) de la population, le risque de divulgation est élevé.
- 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; rendre les catégories de variables plus grossières; 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 est souvent nécessaire de concevoir la stratégie 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 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.

Références

Schackis, D. (1993). **Manual for Disclosure Control**. Eurostat, Luxembourg.

Statistique Canada (1970). **Loi sur la statistique**. Statistique Canada, Ottawa, Canada.

Statistique Canada (1993a). Politique relative à la révélation discrétionnaire. **Manuel des politiques**, 4.3.

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.

Willenborg, L. et de Waal, T. (1996). **Statistical Disclosure Control in Practice**. Notes de cours de statistique, 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 ou d'autres supports électroniques tels qu'Internet (p. ex. les données sur le commerce); publications sur papier; microfiches; fichiers de microdonnées mis à la disposition des utilisateurs autorisés ou pour grande diffusion; 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.

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 efficiente 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 (l'évaluation de la qualité des données) et les deux sections suivantes (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, 1985b, 1986b, 1986c, 1987a, 1988, 1992a, 1992b, 1993a, 1993b, 1994a, 1994d); 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 qui seront diffusées et les données du fichier source sont cohérentes. Il convient de préciser que des données cohérentes ne sont pas forcément identiques. Par exemple, dans le cas de données regroupées ou de variables calculées, il se pourrait que les données diffusées ne soient pas dans le fichier source; ces données et les données de ce fichier doivent toutefois être cohérentes.

- 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 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, les programmes de variance d'échantillonnage) doivent accompagner le produit diffusé (Statistique Canada, 1992c; voir l'annexe 2).
- La documentation doit accompagner le matériel diffusé et doit renfermer, au besoin, des descriptions de la qualité et de la méthodologie utilisée (Statistique Canada, 1992c). La documentation peut être sur copie papier ou sur support électronique. Lorsqu'elle est sur support électronique, elle doit renfermer des instructions aux fins de l'accès à l'information.
- Un produit semblable aux autres produits de Statistique Canada au chapitre de l'apparence et de l'utilisation sera plus convivial.
- 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 (1985b). Politique en matière de diffusion de logiciels. **Manuel des politiques**, 3.5.

Statistique Canada (1986b). 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 (1987a). 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 (1992c). 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 (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 (1994d). Politique sur l'enregistrement des produits et des services d'information. **Manuel des politiques**, 3.10.

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 dans le but d'accroître la valeur du produit statistique. Fondamentalement, le processus d'analyse consiste à étudier un problème, à poser des questions sensées, à étayer les réponses et à communiquer l'information au lecteur. L'analyse des données est une des étapes de l'enquête qui a le plus d'impact sur la qualité des données, car la qualité de l'analyse et de la communication peut accroître ou diminuer considérablement la valeur des étapes précédentes du processus d'enquête. Une analyse judicieuse de données de piètre qualité vaut souvent mieux qu'une piètre analyse de données de bonne qualité.

Par ailleurs, l'analyse des données nous aide également à cerner les problèmes liés à la qualité des données et influe sur les améliorations qui seront apportées au processus d'enquête. L'analyse peut justifier l'existence du programme et mettre en relief son utilité ou, à l'opposé, identifier les programmes qui ne peuvent pas faire la lumière sur les questions ou qui produisent des données inutilisables. L'analyse est essentielle pour intégrer à la planification de nouvelles activités statistiques les commentaires fournis par le personnel affecté aux éditions antérieures des enquêtes et aux études pilotes, pour obtenir des renseignements clés sur les manques en matière de données, de même qu'aux fins de la conception des enquêtes et de la formulation d'objectifs réalistes relativement à la qualité.

Principes

L'agence statistique doit se demander dans quelle mesure les données produites éclairent l'utilisateur relativement à une question ou à une décision à prendre. Dans cette optique, connaître les utilisateurs visés et les questions qui les préoccupent permet de déterminer les techniques d'analyse et les sujets pertinents. Il peut y avoir des différences très marquées entre les utilisateurs au chapitre de la connaissance des affaires sociales et économiques et des méthodes d'analyse des données; il est donc impératif que les auteurs des articles d'analyse cernent le type d'utilisateurs dès le départ.

On s'attend souvent à ce que l'analyse des données pertinentes effectuée par les agences statistique fassent la lumière sur des questions d'intérêt public. Bien que les analystes ne devraient pas s'abstenir de souligner les répercussions des données à cet égard, il n'appartient pas à l'agence statistique de recommander à l'État des politiques ou des programmes. On peut mettre les données et les analyses dans un contexte réel sans promouvoir une politique ou un programme.

L'analyse devrait mettre l'accent sur les idées, et non sur les données. Une analyse efficace est axée sur les résultats, les questions et les thèmes. Selon les utilisateurs visés, il peut être

indiqué de faire abstraction des détails moins pertinents afin de s'assurer que le message principal est communiqué clairement.

Lignes directrices

- L'analyse doit être axée sur les questions plutôt que sur les données. Il faut éviter de tomber dans le piège classique qui consiste à rédiger l'article en se posant la question « Que puis-je dire sur le sujet x à partir de l'ensemble de données y? ». Dans la mesure du possible, utiliser plusieurs sources de données à la fois de l'agence statistique et de l'extérieur.
- Il n'est pas nécessaire que l'article fasse une analyse exhaustive de la question. Si on ne peut pas toujours exposer la question dans ses grandes lignes, on peut souvent étudier un sujet plus restreint dans un cadre global. Il n'est pas rare que plusieurs articles axés sur une question soient rédigés à partir du même ensemble de données.
- Les articles axés sur une question sont généralement préférables aux « profils », qui portent exclusivement sur les données. Ces derniers n'ont pas d'idée principale et ne font qu'un survol des résultats de l'enquête, passant en revue les variables disponibles et soulignant les points saillants. Si cela s'avère adéquat dans certains cas, par exemple pour une première diffusion, de tels articles ne sont pas particulièrement intéressants ou informatifs et n'augmentent que légèrement la valeur des données.
- Lorsqu'on compare deux groupes de personnes, d'entreprises ou d'autres variables, il est essentiel de contrôler les facteurs extérieurs. Si les tests statistiques mettent en évidence des écarts importants entre les groupes, différentes explications doivent être envisagées.
- Si les données ont été recueillies dans le cadre d'une enquête ayant un plan complexe, l'analyse doit en tenir compte. Cela s'explique du fait que les observations ne peuvent pas être considérées comme un échantillon aléatoire constitué de variables aléatoires distribuées de façon indépendante et identique. L'analyse doit intégrer les coefficients de pondération (calibrage) de l'enquête ainsi que la structure variance / covariance des estimations.
- Consulter le Centre de ressources en analyse de données (CRAD), car il assure des services spécialisés dans l'analyse de statistiques, notamment de données provenant d'enquêtes complexes, de données du recensement et de données administratives.
- La prudence est de mise lorsqu'on tire des conclusions relatives à la causalité. À moins d'être certains qu'une seule cause concorde avec les faits, les auteurs de l'article doivent mentionner toutes les explications possibles et non une seule.

- Prendre garde d'étudier les tendances à court terme sans tenir 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.
- Éviter les points de référence arbitraires, par exemple la variation par rapport à l'année dernière. Utiliser plutôt des points de référence significatifs tels que le dernier tournant important pour les données économiques, les variations entre les générations pour les statistiques démographiques et les modifications législatives pour les statistiques sociales.
- Avant de commencer la rédaction de l'article, établir un plan. Le plan oblige les auteurs à se poser des questions telles que « Quel est le sujet que je veux aborder? Quelle information sera mise en relief et comment? Puis-je supprimer des données non pertinentes? Le message est-il intéressant? » Une structure employée couramment et souvent efficace comporte trois volets : une question intéressante (l'introduction), un agencement logique des preuves utilisées afin de répondre à la question (le développement) et la réponse (la conclusion).
- Mettre l'accent sur les variables et les sujets importants dans les articles. En abordant un sujet trop vaste, on atténue souvent l'impact du message principal.
- 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. Un texte plus compréhensible est souhaitable même si la précision en souffre.
- Insérer des graphiques à la place ou en complément du texte ou des tableaux pour communiquer le message. Privilégier les titres qui véhiculent le 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"). Commenter dans le texte l'information transmise dans les graphiques afin de faciliter la compréhension.
- Lorsque des tableaux sont insérés, la présentation générale, l'espacement, la formulation, l'emplacement et l'apparence des titres, des titres de ligne et de colonne et autres identifications doivent contribuer à la clarté des données qu'ils contiennent et prévenir les erreurs d'interprétation.
- Documenter les pratiques ou les méthodes d'arrondissement. Dans la présentation des données arrondies, le nombre de chiffres significatifs doit être le nombre maximal compte tenu de l'utilité des données. Définir la base de calcul des taux.

- 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; l'exactitude des données externes; les calculs arithmétiques simples. S'assurer que ce qui est annoncé dans l'introduction soit effectivement exprimé dans le reste de l'article. S'assurer que les conclusions sont cohérentes à la lumière des preuves fournies. Demander à au moins deux personnes de réviser l'article. Le cas échéant, vérifier la qualité de la traduction.

Références

- Binder, D.A. (1983). On the Variances of Asymptotically Normal Estimators from Complex Surveys. **International Statistical Review**, **51**, 279-292.
- Binder, D.A., Gratton, M., Hidiroglou, M.A., Kumar, S. et Rao, J.N.K. (1984). Analyses de données qualitatives d'enquêtes complexes : quelques expériences canadiennes. **Techniques d'enquête**, **10**, 155-170.
- Fellegi, I.P. (1980). Approximate Tests of Independence and Goodness of Fit Based on Stratified Multistage Samples. **Journal of the American Statistical Association**, **75**, 261-268.
- Fuller, W.A. (1987). Estimators of the Factor Model for Survey Data in **Applied Probability, Statistics and Sampling Theory**, I.B. MacNeill et G.J. Humphrey (éd.). Reidel, Boston, 265-284.
- Hidiroglou, M.A. et Rao, J.N.K. (1987). Chi-squared Tests With Categorical Data from Complex Surveys, I and II. **Journal of Official Statistics**, **3**, 117-140.
- Lehtonen, R. et Pahkinen, E.J. (1995). **Practical Methods for Design and Analysis of Complex Surveys**. Wiley, Chichester.
- Rao, J.N.K. et Scott, A.J. (1987). On Simple Adjustments to Chi-square Tests with Sample Survey Data. **Annals of Statistics**, **15**, 385-397.
- Skinner, C.K., Holt, D. et Smith, T.M.F. (1989). **Analysis of Complex Surveys**. Wiley, Chichester.
- Statistique Canada (1997b). Guidelines on Writing Analytical Articles. Division des communications.
- Statistique Canada (non daté). Le guide de rédaction officielle des Services de rédaction. Services de rédaction, Division des communications, site Intranet de Statistique Canada. Publication mise à jour régulièrement.

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. Fournir une documentation actuelle qui soit 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. Les renseignements sur la qualité des données sont tributaires des utilisations des statistiques, du risque d'erreur et de leur importance aux fins de l'emploi des données, de la variation de la qualité dans le temps, du coût de l'évaluation de la qualité par rapport au coût total du programme statistique, des possibilités d'amélioration et de l'utilité des mesures de la qualité des données pour les utilisateurs (Statistique Canada, 1992c); voir l'annexe 2.
- Selon le public-cible et l'objectif à atteindre, la documentation peut renfermer ce qui suit :
 - Objectifs : Renseignements sur les objectifs et les utilisations des données, l'actualité, la fréquence et les objectifs relatifs à la qualité des statistiques; les

objectifs peuvent changer à mesure que l'enquête progresse (par exemple, en raison de contraintes budgétaires, de la faisabilité perçue, des résultats des nouvelles études pilotes et des nouvelles technologies), et ces changements doivent être documentés comme ils se répercutent sur l'élaboration du questionnaire et l'analyse des résultats des essais.

- Contenu : Questionnaire utilisé, concepts et définitions. Afin de faciliter l'intégration à d'autres sources, mettre en relief le recours à des concepts, à des questions, à des méthodes et à des classifications types. 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 par rapport aux spécifications.
- Méthodologie : Établir des plans d'enquête de rechange. Traiter des questions telles que la population-cible, la base de sondage, la couverture, la période de référence, la stratification, le plan d'échantillonnage, la taille et la sélection de l'échantillon, la méthode de collecte et les méthodes de suivi en cas de non-réponse, l'estimation, 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, disposition, explication des codes, fréquences de base, méthodes de vérification), les systèmes (construction, algorithmes, utilisations, stockage et extraction) et les rapports de surveillance (temps consacré et activités, sources d'incident, ordonnancement des passages pour déterminer si le traitement est effectué dans les délais prévus).
- Opérations : Inclure ou donner en référence les guides de formation, guides du personnel chargé du traitement et des intervieweurs, rapports d'évaluation et 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: inclure 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 dernières, la comparabilité dans le temps et avec d'autres données, les études de validation 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 : Liste des ressources qui ont été utilisées en fonction du temps. Fournir un relevé des dépenses (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).
- 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, 1995b). 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 (1992c). Politique visant à informer les utilisateurs de la qualité des données et la méthodologie. **Manuel des politiques, 2.3**.

Statistique Canada (1993c). Standards and Guidelines for Reporting of Nonresponse Rates. Rapport technique de Statistique Canada.

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

Statistique Canada (non daté). Le guide de rédaction officielle des Services de rédaction. Services de rédaction, Division des communications, site Intranet de Statistique Canada. 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 tels la perception de l'impôt sur le revenu, par exemple. 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 a été créée.

Les dossiers administratifs présentent un certain nombre d'avantages pour une agence statistique ou pour des analystes. Comme ils existent déjà, on évite d'engager des coûts pour recueillir directement les données, ainsi que d'accroître le fardeau de réponse. En général, des données administratives sur l'ensemble de l'univers sont disponibles; par conséquent, les dossiers administratifs n'imposent pas, la plupart du temps, de contraintes inhérentes aux erreurs d'échantillonnage. Surtout, ils peuvent être utilisés de bien des façons aux fins de la production de statistiques, par exemple (i) la création et la mise à jour de bases de sondage; (ii) le remplacement intégral ou partiel (par couplage d'enregistrements) de la collecte de statistiques; (iii) la vérification, l'imputation et la pondération de données tirées de la collecte de statistiques; (iv) l'évaluation de produits statistiques.

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 n'occasionne pas les coûts inhérents à la collecte de données et n'augmente pas le fardeau de réponse. 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.

Aux fins de l'utilisation de données administratives, l'agence statistique doit parfois réaliser certaines étapes du processus d'enquête, que nous avons décrites dans les sections précédentes; cependant, la plupart du temps, le nombre d'étapes est très restreint. 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 (p. ex. pour compenser la qualité de la saisie des données qui, souvent, n'est pas vérifiée).

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 sont pertinentes si les dossiers administratifs existent uniquement sur papier et doivent être codés et saisis. Les lignes directrices se rapportant à la vérification et à la diffusion doivent être respectées lorsqu'on obtient un fichier de dossiers administratifs individuels pour traitement et analyse.
- 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. Par conséquent, les programmes devraient être prêts à justifier cette utilisation secondaire en expliquant qu'elle est sans gravité et dans l'intérêt public.
- Comprendre les limites des dossiers administratifs. L'organisation qui a créé les dossiers administratifs a des préoccupations qui ont une influence profonde sur : (i) les populations couvertes, (ii) les concepts et définitions, (iii) la fréquence et l'actualité et (iv) la qualité de l'information emmagasinée. La longévité de la source et sa portée sont également du ressort de l'organisation responsable des données administratives. Par conséquent, l'utilisation de données administratives dans les programmes statistiques doit être étudiée attentivement afin de minimiser les répercussions défavorables sur la qualité des produits.
- Avec le temps, les préoccupations de l'organisation responsable des données administratives qui définit les concepts, les définitions, la couverture, la fréquence, l'actualité et d'autres attributs du programme administratif, peuvent faire l'objet de changements qui pourraient déformer les séries chronologiques produites, à partir de la source administrative. Ceci peut nécessiter des études spéciales afin de mieux comprendre et contrôler ces sources d'erreur.
- Chaque fois que des particuliers ou des entreprises jouissent d'avantages ou subissent des pertes à la lumière des renseignements qu'ils ont fournis à la source administrative, il se peut que l'information soit biaisée. Il peut être nécessaire de

réaliser des études spéciales afin de mieux comprendre ces sources d'erreur et de les évaluer.

- 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 et puisse même les influencer. Les commentaires sur les statistiques et leurs lacunes peuvent être utiles au fournisseur, ce qui se traduira par une amélioration de la source administrative.
- La collaboration avec les concepteurs des systèmes administratifs nouveaux ou remaniés peut faciliter l'intégration des exigences statistiques à ces systèmes dès le départ. De telles possibilités ne se présentent pas tous les jours; cependant, le cas échéant, les avantages éventuels de la participation de l'agence statistique valent largement le temps qu'elle requiert.
- Comprendre et documenter les définitions, les méthodes et les concepts inhérents à la collecte des données administratives. Ces renseignements, tout particulièrement la définition de la population couverte, sont essentiels si l'on veut déterminer dans quelle mesure les données produites à partir des dossiers administratifs sont appropriées. Ils devraient faciliter la vérification ou les ajustements nécessaires pour intégrer les données administratives au sein des cadres statistiques, ainsi qu'informer les utilisateurs. Accorder une attention spéciale aux définitions et aux méthodes utilisées aux fins du codage géographique des dossiers si ceux-ci sont utilisés pour produire des données pour des petites régions.
- Implémenter des contrôles permanents ou périodiques de la qualité des données d'entrée. L'assurance du maintien de la qualité des données est importante, car l'agence statistique ne contrôle pas leur production. À cette fin, on peut recourir à des dispositifs de protection et de contrôle supplémentaires (p. ex. les méthodes de contrôle statistique de la qualité et les règles de vérification) au moment de la réception des données, effectuer des comparaisons avec d'autres sources ou réaliser des études de suivi des échantillons.
- La première fois qu'on utilise une source de données administratives à des fins statistiques, vérifier la qualité de chaque élément d'information figurant sur le cliché d'enregistrement. Certains éléments peuvent être de très mauvaise qualité et par conséquent inutilisables. Par exemple, la qualité du codage de classification (p. ex. selon la profession ou la branche d'activité) peut être inadéquate d'un point de vue statistique. En pareil cas, l'agence statistique peut décider d'améliorer les codes de classification, ou utiliser les codes disponibles de façon plus limitée (p. ex. seulement le niveau supérieur de la classification).

- Tout comme les données d'enquête, les données administratives ne sont pas à l'abri de la non-réponse, qu'elle soit partielle ou non. Il est donc important de mettre en place un mécanisme de vérification et d'imputation ou de rajustement des coefficients de pondération afin de traiter la non-réponse. Dans certains cas, le manque d'actualité se traduit par des lacunes supplémentaires qui peuvent être corrigées par imputation ou par un rajustement des coefficients de pondération. Lorsqu'on met en place un mécanisme de vérification ou d'imputation, respecter les lignes directrices décrites dans les sections 2.8 et 2.9. De plus, dans le cadre du processus de vérification et d'imputation, accorder une attention spéciale à l'identification des unités actives ou inactives. Ne pas imputer les unités inactives. Il peut également être nécessaire de recourir à l'imputation lorsque certaines unités ont une fréquence de déclaration différente (p. ex. hebdomadaire ou trimestrielle) de la plupart des unités (p. ex. mensuelle).
- Certaines données administratives ont un caractère longitudinal (p. ex. impôt sur le revenu et taxe sur les produits et services); on attribue aux unités un identificateur qui permet le couplage des données provenant de différentes périodes de référence. 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 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. Certaines administrations fournissent des identificateurs communs à différents fichiers administratifs. Là encore, la qualité de ces couplages doit être évaluée.
- 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 de couplage 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 et Scheuren et Winkler, 1993).
- Lorsqu'on doit coupler des enregistrements et 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 (Statistique Canada, 1986). 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 regroupés. 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.

- 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.
- 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 (Statistique Canada, 1996) 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.

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.
- 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 (1996a). Politique d'information des répondants aux enquêtes. **Manuel des politiques**, 1.1.

Statistique Canada (1996b). Politique relative au couplage d'enregistrements. **Manuel des politiques**, 4.1.

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. Elle fait la synthèse des diverses politiques, des processus de gestion, des mécanismes de consultation et des méthodes techniques qui influent sur la gestion de la qualité à Statistique Canada. Alors que la section 2 portait principalement sur la conception des différentes activités statistiques, la section 3 aborde l'assurance de la qualité dans une perspective plus globale, soit celle de la gestion et de l'organisme de statistique, tout en évoquant bon nombre des politiques et des méthodes décrites à la section précédente. Elle inscrit dans leur contexte les lignes directrices relatives à la qualité présentées dans les pages qui précèdent et repose sur le cadre d'assurance de la qualité de Statistique Canada (1997a).

3.1 La qualité à Statistique Canada

Statistique Canada produit de l'information. L'efficacité de l'organisation dépend de la pertinence et de la crédibilité de cette information. À une époque où les besoins des utilisateurs changent et leurs attentes grandissent, nous devons maintenir et accroître la pertinence de notre information. La crédibilité de l'information doit s'appuyer sur la réputation de l'organisation relativement à la fiabilité des données, et tous nos produits doivent y contribuer. Comme peu d'utilisateurs peuvent évaluer indépendamment la fiabilité de l'information que nous produisons, leur confiance doit reposer sur notre « nom ». Il importe donc que le Bureau se soucie de la qualité dans toutes ses activités.

La qualité de l'information statistique est multidimensionnelle. La qualité englobe de nombreuses caractéristiques, non seulement la pertinence, mais également l'exactitude, l'actualité, l'accessibilité, l'intelligibilité et la cohérence. Au niveau des programmes et des produits statistiques, la qualité englobe trois autres caractéristiques : l'uniformité, la compatibilité et l'intégralité. Chaque produit statistique est également multidimensionnel en cela qu'il renferme un éventail d'informations dont la qualité peut varier et qui sont susceptibles d'avoir de nombreuses utilisations. Par ailleurs, comme les programmes statistiques diffèrent au chapitre des objectifs, des priorités, des contraintes et des possibilités, ils se distinguent également au chapitre des caractéristiques de la qualité.

La qualité ne doit pas et ne peut pas être optimisée à tout prix. Le défi posé par la définition du programme statistique général du Bureau consiste à trouver et à maintenir au fil des ans un juste équilibre entre la quantité et la qualité de l'information produite par le programme, compte tenu des ressources disponibles. Le défi que pose la conception de chacun des programmes statistiques consiste à faire des compromis judicieux entre les besoins changeants des clients, les coûts et le fardeau de réponse et les diverses dimensions de la qualité.

Compte tenu de l'importance capitale de la qualité pour le Bureau, la gestion de la qualité doit faire partie intégrante des pratiques administratives et techniques du Bureau. Elle doit être intégrée à la conception de chacun des programmes et être gérée par ceux-ci.

Statistique Canada s'applique à intégrer la qualité à tous ses programmes et à tous ses produits. La qualité des statistiques officielles du Bureau s'appuie sur l'emploi de méthodes scientifiques éprouvées et adaptées progressivement à l'évolution des besoins des clients, à la situation budgétaire et à l'évolution des phénomènes mesurés. Pour tous les employés et gestionnaires de Statistique Canada, la qualité doit être une source de préoccupation et de fierté. Il n'est pas possible de parvenir à un niveau acceptable de qualité en comptant uniquement sur des mesures, des décrets et des contrôles administratifs. La qualité passe par la compréhension des besoins des clients et par l'utilisation judicieuse de connaissances et de compétences à bien des niveaux. Elle requiert un personnel motivé et compétent et suppose un perfectionnement et une gestion efficaces du personnel et de son travail. À cette fin, Statistique Canada a mis en oeuvre une série de programmes en matière de ressources humaines et a défini parallèlement des politiques, des méthodes et des lignes directrices qui visent à encourager, à faciliter et à renforcer des activités et des comportements qui cadrent avec l'importance capitale que le Bureau accorde à la qualité.

3.2 Organisation et infrastructure de la gestion de la qualité

Statistique Canada est divisé en trois secteurs de programmes statistiques comprenant chacun des groupes spécialisés, un secteur de services de gestion, un secteur d'opérations statistiques et un secteur d'infrastructure technique (géographie, classifications techniques, informatique et méthodes statistiques). Ces secteurs « fonctionnels » sont appuyés par des comités de gestion interne, des comités de spécialistes, des comités consultatifs et des comités de conseils professionnels composés principalement de personnes de l'extérieur.

Chaque *comité de gestion* a la responsabilité de fournir des conseils ou de proposer des orientations à l'égard d'une importante fonction ou responsabilité du Bureau ou d'en assurer la gestion. Les comités de gestion comprennent par exemple le Comité de planification générale, le Comité des méthodes et des normes, le Comité de la formation et du perfectionnement, le Comité du recrutement et du perfectionnement ainsi que le Comité de la diffusion. Le principal comité de gestion qui supervise les activités du Bureau est le Comité des politiques.

Les *comités de spécialistes du sujet* assurent la coordination des programmes statistiques afin d'en promouvoir la qualité dans divers domaines de spécialisation (p. ex. démographie, statistique du travail, données sur les autochtones, statistique de l'environnement, revenu et richesse). Ils cherchent en particulier à assurer la pertinence et la cohérence des données, notamment en favorisant l'emploi de définitions et de systèmes de classification communs et l'utilisation de métadonnées uniformes et efficaces.

Les *comités consultatifs* – le Conseil national de la statistique, le Comité consultatif fédéral-provincial de la politique statistique et ses comités subsidiaires – émettent des avis sur de grandes questions d'orientation, les priorités, les besoins en données, les besoins des utilisateurs et les produits.

Treize *comités de conseils professionnels* couvrant les principaux domaines de spécialisation du Bureau étudient en permanence les produits et les méthodes statistiques du Bureau et contribuent à établir les priorités dans ces domaines.

Il appartient aux secteurs des programmes d'assurer la mise en oeuvre, la gestion et l'examen des programmes statistiques dont ils ont la responsabilité. Aux fins du développement et du réaménagement, ils mettent en oeuvre et gèrent les programmes en s'appuyant sur un système de gestion matricielle qui fait appel à des équipes de projet. Les secteurs des programmes fournissent les compétences professionnelles (et certaines compétences techniques), et les secteurs de l'infrastructure, des compétences et des services techniques et opérationnels. Divers éléments de l'infrastructure des services de gestion leur offrent un appui direct ou indirect (p. ex. ressources humaines, services financiers, formation, vérification, évaluation des programmes).

Le gestionnaire d'un programme statistique est responsable de la qualité des données et de l'information produites, compte tenu de la portée du mandat du programme. Les secteurs de l'infrastructure technique et des opérations statistiques ont la responsabilité d'offrir des méthodes techniques et opérationnelles, des systèmes et des opérations scientifiquement judicieux et efficaces.

Cette organisation fournit un système d'autocontrôle qui contribue au maintien de pratiques statistiques judicieuses et uniformes.

3.3 Cadre d'assurance de la qualité

Contrairement à la plupart des utilisateurs de produits manufacturés, les utilisateurs de données statistiques ne sont généralement pas en mesure de vérifier d'eux-mêmes la qualité des produits qu'ils obtiennent uniquement en les examinant, en les mettant à l'essai et en les utilisant. Alors que la pertinence et l'actualité des données statistiques peuvent être évidentes pour l'utilisateur, une simple inspection des données ne suffit pas pour donner une idée d'autres dimensions de leur qualité, tout particulièrement leur exactitude. Le producteur de données statistiques doit donc mettre en place des mécanismes de gestion et de mesure de la qualité qui permettent de fixer les principales exigences en matière de qualité, de faire en sorte que les produits s'y conforment et de fournir ensuite à l'utilisateur assez de renseignements pour voir si les données sont bien adaptées ce à quoi elles sont dédiées.

La gestion de la qualité exige une connaissance des risques et des possibilités susceptibles d'influer sur la qualité des données, ainsi que le recours à des méthodes administratives et scientifiques permettant d'atténuer ces risques et de tirer parti de ces possibilités. Au nombre

des risques éventuels, mentionnons notamment le fait de perdre de vue les besoins des clients, la diminution de la coopération des répondants, les contraintes budgétaires et les pénuries de compétences. Parmi les possibilités, on retrouve les progrès rapides de la technologie, l'avancement des méthodes statistiques et la demande croissante d'analyses des politiques. Une gestion attentive peut transformer des risques en possibilités; malheureusement, le contraire est également vrai.

Le cadre d'assurance de la qualité (Statistique Canada, 1997a) décrit les mécanismes mis en place par le Bureau pour gérer la qualité au sein de sa structure organisationnelle, compte tenu des risques et des possibilités connus. Ces mécanismes visent essentiellement a) à aider le Bureau à prendre des décisions administratives judicieuses et éclairées au sujet de l'affectation des ressources et à donner des avis techniques judicieux et éclairés qui s'appuient sur sa connaissance et son savoir-faire; b) à faire en sorte que la qualité soit prise en compte comme il se doit dans l'exécution des opérations émanant de ces décisions. Les éléments du cadre d'assurance de la qualité sont décrits ci-dessous sous trois grandes rubriques : *assurance permanente de la pertinence; conception et exécution; et environnement.*

3.3.1 Pertinence

L'élément « *pertinence* » comprend les mécanismes qui permettent au Bureau de déterminer quelles données il doit produire et quel niveau de ressources il doit affecter à chaque programme. Cela consiste essentiellement à traduire les besoins des utilisateurs en autorisations de programmes et en décisions budgétaires à l'intérieur du Bureau. Comme l'appellation de cet élément le laisse entendre, ces mécanismes sont intimement liés aux plus importants facteurs permettant d'assurer la dimension de la qualité liée à la pertinence pour chacun des programmes statistiques et l'ensemble des programmes. Les mécanismes qui servent à assurer la pertinence de l'information permettent aussi d'exercer un contrôle de base sur les autres aspects de la qualité et d'évaluer ainsi l'importance que leur accordent les utilisateurs.

Tout en reconnaissant qu'environ 95 % des ressources budgétaires du Bureau sont affectées à des programmes permanents qui ne sont pas discrétionnaires, et en tenant compte de l'interdépendance des différents programmes, le Bureau a mis en place des mécanismes qui servent à contrôler la pertinence des programmes existants, qui mettent en évidence les lacunes des données que le programme existant ne permet pas de combler et qui se traduisent par des décisions touchant des modifications du programme ou de nouvelles mesures de financement visant à accroître la pertinence du programme général du Bureau. Ces mécanismes peuvent être répartis entre quatre grands groupes : les commentaires des clients et des intervenants; l'évaluation des programmes; l'analyse des données; et la planification. Les trois premiers mécanismes comprennent les activités qui visent à recueillir des données et des renseignements sur les besoins en statistiques des clients, sur l'efficacité des programmes existants dans l'optique de ces besoins ainsi que sur les lacunes et les faiblesses des programmes existants. Le quatrième renvoie au processus visant à déterminer quels changements doivent être apportés à la lumière de ces renseignements.

Cependant, les besoins des clients ne tiennent pas uniquement à une question de pertinence des programmes, pris individuellement ou collectivement. Il faut voir aussi aux autres éléments de la qualité. Pour assurer la pertinence des programmes et prendre des décisions au moment de la planification, il faut évaluer tous les éléments de la qualité recherchée et en tenir compte. Cette évaluation suppose en partie une analyse des données produites et un examen des mécanismes permettant de cerner et de corriger les faiblesses ou les défauts d'uniformité, et en partie une évaluation des résultats obtenus en fonction des coûts.

3.3.2 Conception et exécution

L'élément « *conception et exécution* » comprend la conception, l'exécution et l'évaluation des mécanismes de production de données dans le cadre des programmes approuvés, compte tenu des contraintes budgétaires. En plus de déterminer la pertinence des données et d'être associées à l'établissement des priorités, les mesures de conception et d'exécution ont une incidence prépondérante sur l'exactitude, l'actualité, l'accessibilité et l'intelligibilité des données.

En général, des comités de gestion et de direction de projet voient à ce que les programmes statistiques soient réalisés en conformité avec leur mandat. Grâce à une combinaison d'approches de gestion fonctionnelle et matricielle, on s'assure que les ressources et le savoir-faire requis sont planifiés et affectés aux projets. Cette façon de faire permet par ailleurs d'assurer un suivi des progrès, des problèmes et des questions et d'en faire rapport de façon que le mandat et les objectifs soient correctement interprétés et que les jugements portés et mis en oeuvre soient appropriés. La forme exacte des pratiques retenues dépend de la taille, de la pertinence, de la complexité et de la nature du programme et des ressources qui y sont affectées.

Le recours à une approche interdisciplinaire de gestion par projet/équipe de projet dans la conception et la mise en oeuvre des programmes, étayée par l'organisation fonctionnelle, contribuent pour beaucoup à faire en sorte que la qualité bénéficie de l'attention voulue. Les méthodologistes des équipes de projet, notamment, ont la responsabilité explicite de mettre leurs connaissances à profit pour parvenir à un juste équilibre entre les divers aspects de la qualité des données d'un projet. Le fait qu'ils appartiennent à une structure fonctionnelle spécialisée leur permet de développer pleinement leur savoir-faire et d'avoir recours, au besoin, à des ressources spécialisées et à leurs supérieurs pour résoudre des différends qui ne pourraient pas l'être au sein de l'équipe de projet.

La collecte ou l'acquisition, l'exploitation et la compilation des données nécessitent l'application de méthodes et de modèles statistiques et analytiques judicieux ainsi que l'emploi de plans, d'instruments, de méthodes et de procédures opérationnelles efficaces, et de systèmes et d'algorithmes efficaces. La qualité obtenue (exactitude, actualité et cohérence) dépendra des méthodes explicites utilisées et des mécanismes d'assurance de la qualité prévus pour mettre en évidence et contrôler les erreurs éventuelles à chacun des stades du programme. Les gestionnaires des programmes disposent d'une grande latitude dans

l'application des pratiques et des méthodes. C'est à eux qu'il revient de déterminer et de justifier le niveau de qualité qui sera jugé acceptable et les critères correspondants, compte tenu des circonstances, des contraintes, des possibilités et des objectifs, dans le cadre du mandat approuvé par le Comité des politiques.

Quelles que soient les méthodes retenues, elles doivent entrer dans le domaine des pratiques statistiques généralement reconnues et défendables, compte tenu des circonstances. L'emploi de nouvelles technologies et le recours à l'innovation pour améliorer la qualité et l'efficacité sont toujours encouragés, mais doivent au préalable avoir fait l'objet d'essais propres à minimiser les risques. Les questionnaires, en particulier, doivent faire l'objet d'essais démontrant que les répondants pourront et voudront fournir des données de qualité acceptable. Il doit être possible de contrôler la qualité, de réagir efficacement aux problèmes imprévus et de vérifier ou de justifier la crédibilité des résultats et d'en comprendre les limites. Les pratiques et les méthodes retenues peuvent évoluer en fonction des décisions d'orientation ou par suite de l'évaluation permanente des programmes.

À l'étape de la conception ou du remaniement et dans le cadre des examens permanents, on procède à des évaluations techniques des méthodes proposées ou en place et on évalue l'efficacité opérationnelle et l'évolution des coûts. Ces évaluations permettent de déterminer si les pratiques ou les propositions techniques sont appropriées. Elles contribuent également à améliorer et à orienter la mise en oeuvre d'éléments précis de la méthodologie et des opérations dans les programmes et à l'échelle du Bureau.

En plus de faire connaître et de décrire les phénomènes statistiques et de mettre en évidence les lacunes des données, l'analyse permet également d'évaluer ou de mesurer l'exactitude et la cohérence des données. Dans ce contexte, les résultats des analyses peuvent avoir des conséquences diverses : par exemple, ajout ou modification de méthodes de vérification des données; modifications des questionnaires; ajout de méthodes de collecte de données; nouveaux programmes de formation du personnel; utilisation de nouveaux systèmes ou méthodes; remaniement.

La diffusion des données et des produits statistiques est l'aboutissement des efforts visant à assurer la qualité, puisque c'est à cette étape que les produits parviennent enfin aux utilisateurs. Même si tous les aspects de la qualité sont importants à cette étape, l'intelligibilité, la cohérence et l'accessibilité jouent un rôle clé dans la qualité ultime des données diffusées.

3.3.3 Environnement

L'élément « *environnement* » comprend les initiatives et les mécanismes généraux qui permettent à Statistique Canada de maintenir un climat qui encourage le personnel à se préoccuper de la qualité et à produire des statistiques de la meilleure qualité possible, compte tenu des contraintes opérationnelles et budgétaires. Ces mesures comprennent le recrutement d'employés de talent, la sensibilisation du personnel aux questions relatives à la qualité, la

création d'un réseau de communications internes ouvert et efficace, l'adoption de mesures explicites de création de partenariats et d'ententes avec les fournisseurs du Bureau (les répondants, notamment), la création de programmes d'analyse de données et de recherche méthodologique propices à l'amélioration, et l'établissement et la mise à jour de définitions, de classifications, de cadres et d'outils méthodologiques normalisés qui contribuent à l'intelligibilité et à la cohérence des données.

Bien que ces trois éléments soient décrits séparément, ils sont en interaction étroite. Par exemple, il y a un rapport important entre les évaluations qui sont faites dans le cadre de l'élément « conception et exécution » et les décisions qui sont prises dans le cadre de l'élément « pertinence » au titre des investissements dans les programmes.

Même si le cadre permet une gestion et une assurance de la qualité cohérentes, ses éléments n'ont pas été mis en place à un moment précis en vue de « gérer la qualité ». Le cadre a plutôt évolué au fil des ans en suivant l'évolution administrative et technique du Bureau et témoigne du rôle capital que le souci de la qualité joue dans la gestion d'un organisme de statistique. Les éléments en place sont issus d'un large éventail d'activités permanentes faisant intervenir tous les niveaux du Bureau, de même que les clients, les intervenants, les « fournisseurs » et les groupes consultatifs. Il appartient notamment aux comités de projet, aux comités consultatifs, aux comités de gestion ainsi qu'aux services de planification générale et de planification de projets du Bureau de cerner les questions de gestion de la qualité, de trouver et de faire connaître des pratiques exemplaires et de proposer des activités susceptibles de promouvoir la qualité.

3.4 Conclusion

Le régime 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 régime ne dépend pas d'un seul processus, mais bien de l'effet combiné de nombreuses mesures interdépendantes qui s'appuient sur l'intérêt professionnel et la motivation du personnel et qui se renforcent mutuellement grâce à l'attention qui est accordée aux besoins des clients, à l'accent qui est mis sur le professionnalisme et au souci constant de la qualité des données. Bien que toute description d'un régime complexe donne inévitablement l'impression d'en dissocier les éléments, le trait dominant du régime de Statistique Canada 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 moyens et l'uniformité des messages.

Références

Statistique Canada (1997a). Aperçu du cadre d'assurance de la qualité de Statistique Canada. Rapport non publié, Statistique Canada.

4. BIBLIOGRAPHIE

- Bankier, M., Filion, J.-M., Luc, M. et Nadeau, C. (1994). Imputing Numeric and Qualitative Variables Simultaneously. **Proceedings of the Section on Survey Research Methods**, American Statistical Association, 242-247.
- 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., Gratton, M., Hidioglou, M.A., Kumar, S. et Rao, J.N.K. (1984). Analyses de données qualitatives d'enquêtes complexes : quelques expériences canadiennes. **Techniques d'enquête**, 10, 155-170.
- Binder, D.A., et Hidioglou, M.A. (1988). Sampling in Time. **Handbook of Statistics**, P.K. Krishnaish et C.R. Rao (éd.), 187-211.
- Brackstone, G.J. (1987). Utilisation des dossiers administratifs à des fins statistiques. **Techniques d'enquête**, 13, 35-51.
- Brackstone, G.J. (1991). Shaping Statistical Services to Satisfy User Needs. **Statistical Journal of the United Nations**, ECE 8, 243-257.
- Brackstone, G.J. (1993). Data Relevance: Keeping Pace With User Needs. **Journal of Official Statistics**, 9, 49-56.
- Brick, J.M., Broene, P., James, P. et Severynse, J. (1997). **A User's Guide to WesVar PC**. Westat Inc., Maryland.
- 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. (1986). Outlier Robust Finite Population Estimation. **Journal of the American Statistical Association**, 81, 1063-1069.
- 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**. John Wiley and Sons, New York.
- Colledge, M.J. (1995). Frames and Business Registers: An Overview. In **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.
- Cotton, C. (1993). Functional Description of the Generalized Edit and Imputation System. Rapport technique de 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**. John Wiley and Sons, New York.
- Couper, M.P. et Groves, R.M. (1992). Le rôle de l'intervieweur dans la participation aux enquêtes. **Techniques d'enquête**, **18**, 279-294.
- 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. 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.

- 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.
- 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.
- 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.
- Fellegi, I.P. et Holt, D. (1976). Systematic Approach to Automatic Edit and Imputation. **Journal of the American Statistical Association**, 71, 17-35.
- Fellegi, I.P. (1980). Approximate Tests of Independence and Goodness of Fit Based on Stratified Multistage Samples. **Journal of the American Statistical Association**, 75, 261-268.
- Fowler, F.J. Jr. (1995). Improving Survey Questions: Design and Evaluation. **Applied Social Research Methods Series**, 38, Sage Publications, Thousand Oaks, Californie.
- Fuller, W.A. (1987). Estimators of the Factor Model for Survey Data. In **Applied Probability, Statistics and Sampling Theory**, I.B. MacNeill et G.J. Humphrey (éd.). Reidel, Boston, 265-284.
- Fuller, W. (1987). **Measurement Error Models**. Wiley, New York.
- Gagnon, F., Gough, H. et Yeo, D. (1994). Survey of Editing Practices in Statistics Canada. Rapport technique de Statistique Canada.

- Gagnon, F., Lee, H., Rancourt, E. et Särndal, C.E. (1996). Estimating the Variance of the Generalized Regression Estimator in the Presence of Imputation for the Generalized Estimation System. **Proceedings of the Survey Methods Section**, Société statistique du Canada, 151-156.
- 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 in **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? In **Survey Measurement and Process Quality**, Lyberg et coll. (éd.), Wiley & Sons, New York, 415-435.
- 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. (1989). **Survey Errors and Survey Costs**, New York: John Wiley
- 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., 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 Rao, J.N.K. (1987). Chi-squared Tests With Categorical Data from Complex Surveys, I and II. **Journal of Official Statistics**, **3**, 117-140.
- 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, 330-305.
- Kalton, G. et Kasprzyk, D. (1986). Le traitement des données d'enquête manquantes. **Techniques d'enquête**, 12, 1-17.
- Kish, L. (1965). **Survey Sampling**. Wiley, New York.
- 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.
- Kovar, J.G. et Whitridge, P. (1995). Imputation of Business Survey Data. In **Business Survey Methods**, B.G. Cox et al. (éd.), Wiley, New York, 403-423.
- 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, Editor), 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.
- Lee, H. (1995). Outliers In Business Surveys. in **Business Survey Methods**, B.G. Cox et coll. (éd). New York: John Wiley and Sons, 503-526.
- 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.

- 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.
- 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. in **Telephone Survey Methodology**, R.M. Groves et coll. (éd.), Wiley & Sons, 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., 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). Generalized Data Collection and Capture (DC2): Release 2.5.1, Sample Verification (SV). Rapport Technique de Statistique Canada.
- Mudryk, W. et Xiao, P. (1996). Quality Control Methodology for LFS Industry and Occupation Coding Operations. 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.
- Nelson, D.D. (1985). Informal Testing as a Means of Questionnaire Development. **Journal of Official Statistics**, 1, 179-188.
- 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.** John Wiley and Sons, 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. et Scott, A.J. (1987). On Simple Adjustments to Chi-square Tests with Sample Survey Data. **Annals of Statistics**, 15, 385-397.
- 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, in **L'échantillonnage : un guide non mathématique - deuxième édition**, 7-10, Publication n° 12-602F au catalogue, Statistique Canada.
- Schackis, D. (1993). **Manual for Disclosure Control.** Eurostat, Luxembourg.
- 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.

- Shah, B.V., Barnwell, B.G. et Bieler, G.S. (1997). SUDAAN User's Manual Release 7.5. Research Triangle Institute, Caroline du Nord.
- 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. (1996). Combining Information in Survey Sampling by Modified Regression. **Proceedings of the Section on Survey Research Methods**, American Statistical Association, 120-129.
- Singh, M.P., Drew, J.D., Gambino, J.G. et Mayda, F. (1990). **Méthodologie de l'Enquête sur la population active du Canada**. Publication n° 71-526 au catalogue, Statistique Canada.
- Singh, M.P., Gambino, J. et Mantel, H. (1992). Issues and Options in the Provision of Small Area Data. **Small Area Statistics and Survey Designs**, G. Kalton, J. Kordos et R. Platek (éd.). Bureau central de la statistique, Varsovie, 31-75.
- Sirken, M.G. et Casady, R.J. (1988). Sampling Variance and Nonresponse Rates in Dual Frame, Mixed Mode Surveys. Chapitre 11 in **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.
- Statistique Canada (1970). **Loi sur la Statistique**. Statistique Canada, Ottawa, Canada.
- Statistique Canada (1980). **Classification type des industries 1980**. Division des normes, Statistique Canada, Ministre des Approvisionnements et Services du Canada, Ottawa.
- Statistique Canada (1985a). Politique d'édition. **Manuel des politiques**, 3.1.
- Statistique Canada (1985b). Politique en matière de diffusion de logiciels. **Manuel des politiques**, 3.5.
- Statistique Canada (1986a). **Classification canadienne type des industries pour compagnies et entreprises, 1980**. Division des normes, Statistique Canada, Ministre des Approvisionnements et Services du Canada, Ottawa.
- Statistique Canada (1986b). 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 (1987a). Politique sur la diffusion des microdonnées. **Manuel des politiques**, 4.2.
- Statistique Canada (1987b). Politique sur les classifications types des industries. **Manuel des politiques**, 2.7.
- 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 (1992c). 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 (1992d). **Classification géographique type (CGT) 1991**. Division des normes, Statistique Canada, Ministre de l'Industrie, des Sciences et de la Technologie, Ottawa.
- 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 (1993c). Standards and guidelines for reporting of nonresponse rates. Rapport technique de Statistique Canada.
- Statistique Canada (1993d). **Classification type des biens 1992** (basée sur le Système harmonisé de désignation et de codification des marchandises). Division des normes, Statistique Canada, Ministre de l'Industrie, des Sciences et de la Technologie, Ottawa.
- Statistique Canada (1993e). **Classification type des professions 1991**. Division des normes, Statistique Canada, Ministre de l'Industrie, des Sciences et de la Technologie, Ottawa.

- Statistique Canada (1993f). Guidelines for Seasonal Adjustment of Current Observations and their Revisions. Document publié par le Comité des méthodes et des normes, 10 février.
- Statistique Canada (1994a). Politique de rabais. **Manuel des politiques**, 3.9.
- Statistique Canada (1994b). Politique sur les classifications types des professions. **Manuel des politiques**, 2.9.
- Statistique Canada (1994c). Politique sur l'élaboration, l'essai et l'évaluation des questionnaires. **Manuel des politiques**, 2.8. (Reproduite à l'annexe 1).
- Statistique Canada (1994d). Politique sur l'enregistrement des produits et des services d'information. **Manuel des politiques**, 3.10.
- Statistique Canada (1995a). Data Quality Criteria in the Dissemination of Statistical Information. Version provisoire de la politique rédigée par le Comité des méthodes et des normes, Statistique Canada.
- Statistique Canada (1995b). Politique concernant l'évaluation des produits d'information. **Manuel des politiques**, 2.5.
- Statistique Canada (1996a). Politique d'information des répondants aux enquêtes. **Manuel des politiques**, 1.1.
- Statistique Canada (1996b). Politique relative au couplage d'enregistrements. **Manuel des politiques**, 4.1.
- Statistique Canada (1997a). Aperçu du cadre d'assurance de la qualité de Statistique Canada. Rapport non publié, Statistique Canada.
- Statistique Canada (1997b). Guidelines on Writing Analytical Articles. Division des communications.
- Statistique Canada (1997c). Politique sur les unités géographiques types. **Manuel des politiques**, 2.4.
- Statistique Canada (1997d). Initiative d'harmonisation de la statistique sociale. Rapport non publié, version 2.1, Statistique Canada.
- Statistique Canada (1998a). **Système de classification des industries de l'Amérique du Nord, Canada 1997**. Publication n° 12-501 au catalogue, Statistique Canada.

- Statistique Canada (1998b). Rapport d'étape sur l'harmonisation des statistiques sociales. Article présenté à l'occasion de la Réunion de travail sur les systèmes de métainformation, Conférence des statisticiens européens, Commission de statistique et Commission économique pour l'Europe, février 1998.
- Statistique Canada (1998c). Politique concernant les normes. **Manuel de politiques**, 2.10.
- Statistique Canada (non daté). **Manuel des politiques**. Statistique Canada, Ottawa, Canada.
- Statistique Canada. (non daté). Le guide de rédaction officielle des Services de rédaction. Services de rédaction, Division des communications, site Intranet de Statistique Canada. Publication mise à jour régulièrement.
- 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.
- 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.
- U.S. Executive Office of the President, Office of Management and Budget, U.S. Department of Commerce, Statistique Canada et Office statistique des Communautés européennes (1997). Concordance internationale entre les classifications industrielles des Nations Unies (CITI Rev 3) et du Canada (CTI 1980), de l'Union Européenne (NACE Rev 1) et des États-Unis (1987 SIC).
- Willenborg, L. et de Waal, T. (1996). **Statistical Disclosure Control in Practice**. Notes de cours de statistique, 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.

5. ANNEXES

5.1 Annexe 1: Politique sur l'élaboration, l'essai et l'évaluation des questionnaires

Approuvée le 5 janvier 1994

Contexte

Les questionnaires jouent un rôle essentiel dans la collecte des données. La qualité des données et l'image de Statistique Canada auprès du public en dépendent dans une large mesure.

Un questionnaire bien conçu devrait permettre de recueillir des données de façon efficace et avec un minimum d'erreurs. Il devrait faciliter aussi le codage et la saisie des données. Il devrait minimiser les vérifications et les imputations qui sont nécessaires et entraîner une baisse générale du coût et du temps nécessaires à la collecte et au traitement des données.

Statistique Canada a pour objectifs que les questionnaires soient soigneusement mis à l'essai, simples à remplir et à dépouiller, d'utilisation facile pour le répondant comme pour l'intervieweur. Cette politique impose donc la mise à l'essai des questionnaires nouveaux et révisés. Elle prévoit aussi une évaluation périodique des questionnaires utilisés dans les enquêtes permanentes. La politique couvre tous les aspects de l'étude des questionnaires qui peuvent influencer la qualité des données, le comportement des répondants et le rendement des intervieweurs.

Définition

Aux fins de la présente politique, on définit un questionnaire comme un ensemble de questions destinées à recueillir des renseignements auprès d'un répondant. Un questionnaire peut être rempli par un intervieweur ou un répondant.

Domaine d'application

Cette politique couvre tous les questionnaires que Statistique Canada utilise pour les répondants extérieurs. Ceci comprend les questionnaires qui servent aux programmes de collecte de statistiques et de renseignements ainsi qu'à d'autres programmes comme la recherche de marché et l'évaluation des programmes. Le domaine d'application recouvre aussi la collecte de données par papier et crayon et d'autres moyens, comme les interviews assistées par ordinateur. Cette politique porte également sur les révisions à un questionnaire résultant du changement de la méthode de collecte des données (des interviews papier et crayon aux interviews assistées par ordinateur).

Politique

Statistique Canada a pour politique que:

1. Tous les nouveaux questionnaires et les révisions aux questionnaires existants doivent faire l'objet d'une mise à l'essai dans les deux langues officielles avant leur mise en oeuvre.
2. Les secteurs de programmes doivent consulter le Centre d'information sur la conception des questionnaires à propos des plans d'élaboration et d'essai de questionnaires nouveaux ou révisés. Les lignes directrices annexées donnent une vue d'ensemble des méthodes d'essai qui peuvent être utilisées.
3. Le Centre d'information sur la conception des questionnaires examine tous les questionnaires nouveaux et révisés avant la mise à l'essai et une nouvelle fois après que des révisions auront été apportées à la suite de la mise à l'essai.
4. Tous les questionnaires de Statistique Canada font l'objet d'une évaluation périodique.

Responsabilités

1. Secteurs de programmes

- Les secteurs de programmes veillent à ce que toutes les conditions de mise en oeuvre de cette politique soient réunies. Ils constituent des réserves budgétaires adéquates pour élaborer et mettre à l'essai les questionnaires nouveaux ou remaniés.
- Ils consultent le Centre d'information sur la conception des questionnaires à propos des projets d'élaboration et de mise à l'essai de questionnaires nouveaux ou révisés.
- Au moment approprié lors de la conception des questionnaires, ils consultent au besoin des services tels que la Division des communications (Section des relations publiques), la Division de la diffusion (Étude et production des formules) et la Division des normes.
- Ils soumettent chaque questionnaire nouveau ou révisé à la Division des services d'accès et de contrôle aux données.
- Ils soumettent la version finale de chaque questionnaire nouveau ou révisé à la Division des normes pour inclusion dans "l'index des enquêtes et questionnaires de Statistique Canada".

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

- La Division s'assure que les questionnaires sont conformes à la Politique d'information des répondants aux enquêtes.
- Elle s'assure que le Centre d'information sur la conception des questionnaires a été consulté et qu'il a examiné les questionnaires.
- Elle s'assure que le statisticien en chef donne son autorisation à la collecte des renseignements.

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

Le Centre offre des services d'examen et de consultation pour la conception des questionnaires.

- Il offre de l'aide et des conseils pour l'élaboration et la mise à l'essai de questionnaires nouveaux et révisés. Ceci comprend a) la mise à l'essai de questionnaires grâce à des méthodes comme les groupes de discussion, les techniques cognitives et les prétests, b) la coordination des activités d'élaboration et de mise à l'essai faisant l'objet de marchés de services.
- Il étudie les questionnaires sur les plans du libellé et de l'ordre des questions, la longueur, le format, la facilité d'utilisation pour le répondant et pour l'intervieweur.
- Il fournit la documentation, les lignes directrices et la formation nécessaires pour améliorer la conception des questionnaires.

4. Conception et production des imprimés (Division de la diffusion)

- Aide à la conception, à la présentation, à la composition et à l'impression des questionnaires et des formules d'enquête.
- Veille à ce que le questionnaire réponde aux exigences du Programme fédéral de coordination de l'image de marque.

5. Comité des méthodes et des normes

- Le Comité des méthodes et des normes surveille la mise en application de cette politique.

- Il approuve toute ligne directrice visant l'étude des questionnaires émise en vertu de la présente politique.

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

Approuvée le 7 avril 1992

Introduction

En sa qualité d'organisme professionnel responsable de la production de données officielles, Statistique Canada se doit de faire connaître aux utilisateurs les méthodes et les concepts utilisés dans la collecte et le traitement de ses données, de même que la qualité des données produites et les autres caractéristiques des données pouvant limiter leur utilité et leur interprétation.

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 l'erreur se rapportant aux données restreint, ou restreint davantage, l'utilisation de celles-ci.

L'évaluation de la qualité des données constitue toutefois une tâche complexe compte tenu de la nature pluridimensionnelle du concept de la qualité, de l'existence de nombreuses sources éventuelles d'erreur et, souvent, de l'absence de mesure exhaustive de la qualité des données. Dans l'état actuel de nos connaissances, 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 cependant insister sur la description et la quantification des principaux éléments de la qualité.

Politique

1. Statistique Canada fournira aux utilisateurs des indicateurs de la qualité des données diffusées et des descriptions des concepts, des définitions et des méthodes utilisés.
2. Il faut joindre aux produits statistiques de la documentation relative à la qualité et à la méthodologie ou prévoir d'y inclure un renvoi à celle-ci.
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 effectuées contre remboursement des frais 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 diffusées par Statistique Canada, quelle que soit la méthode de collecte, de calcul ou de regroupement utilisée, abstraction faite du mode de diffusion ou de la source de financement.

Responsabilités

Les 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;

Le Comité des méthodes et des normes assume la responsabilité:

- de rédiger des rapports périodiques sur l'application de la présente politique;
- d'évaluer périodiquement la mise en oeuvre 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 oeuvre 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
- d'étudier 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(x) président(s) du Comité des méthodes et des normes.

6. INDEX DES TERMES IMPORTANTS

- accessibilité (5)
- activité statistique subséquente (8)
- actualité (5)
- analyse des données (67)
- base de sondage (17)
- calibrage (45)
- certification (55)
- classification (13)
- codage (33)
- cohérence (5)
- collecte de données (33)
- comité de gestion (81)
- comités consultatifs (82)
- comités de conseils professionnels (82)
- comités de spécialistes du sujet (81)
- composante irrégulière (50)
- conception (6, 9)
- conception et exécution (84)
- concepts (13)
- confidentialité (59)
- contrôle de la divulgation (59)
- couverture (17)
- cycle (50)
- définitions (13)
- désaisonnalisation (50)
- diffusion (64)
- documentation (71)
- domaines (22, 46)
- données auxiliaires (21)
- données tabulaires (60)
- dossiers administratifs (8, 74)
- dossiers statistiques (74)
- échantillon (21)
- échantillonnage (21)
- échantillonnage à deux degrés (22)
- échantillonnage à deux phases (22)
- échantillonnage probabiliste (21)
- éléments de la qualité (4)
- enquête (8)
- enquête-échantillon (8)
- enquêtes (6)
- environnement (85)
- erreurs de couverture (57)
- erreurs de mesure (57)
- erreurs de traitement (57)
- erreurs dues à la non-réponse (57)
- erreurs d'échantillonnage (57)
- erreurs d'estimation (57)
- erreurs fatales (37)
- estimateur de Horvitz-Thompson (45)
- estimateurs pour petits domaines (46)
- estimation (45)
- estimation composite (47)
- estimation par domaine (46)
- estimations de la tendance-cycle (50)
- estimations sans biais par rapport au plan (45)
- études sur les sources d'erreurs (55)
- évaluation de la qualité (55)
- exactitude (5)
- facteurs de calibrage (45)
- gestion de la qualité totale (4)
- grappes (22)
- imputation (41)
- imputation déterministe (41)
- imputation stochastique (41)
- intelligibilité (5)
- lignes directrices (8, 9)
- méthodes de modification des données (62)
- méthodes de réduction des données (62)
- microdonnées (60)
- non-réponse partielle (29)
- objectifs (10)
- pertinence (4, 83)
- plan d'échantillonnage (21)
- poids d'échantillonnage (45)
- population-cible (17)
- portée et objet (8, 9)
- précision (23, 45)
- principes (8, 9)
- probabilité d'inclusion (21)
- qualité (4, 6)
- quasi-identificateurs (60)
- questionnaire (25)
- recensement (8)

règle (n,k) (61)
règle C-times (61)
regroupement (61)
regroupement des valeurs extrêmes
supérieures (61)
renouvellement (24)
réponse partielle (29)
stratification (21)
suppression de cellules (61)
tableaux de données quantitatives (60)
tableaux de fréquence (60)
taux de réponse (29)
taux de succès (38)
taux d'échantillonnage (23)
tendance (50)
unités primaires d'échantillonnage (22)
utilisateurs (10)
utilisations (10)
validation (55)
variables clés (60, 62)
variations saisonnières (50)
vérification (37)
vérifications exploratoires (37)