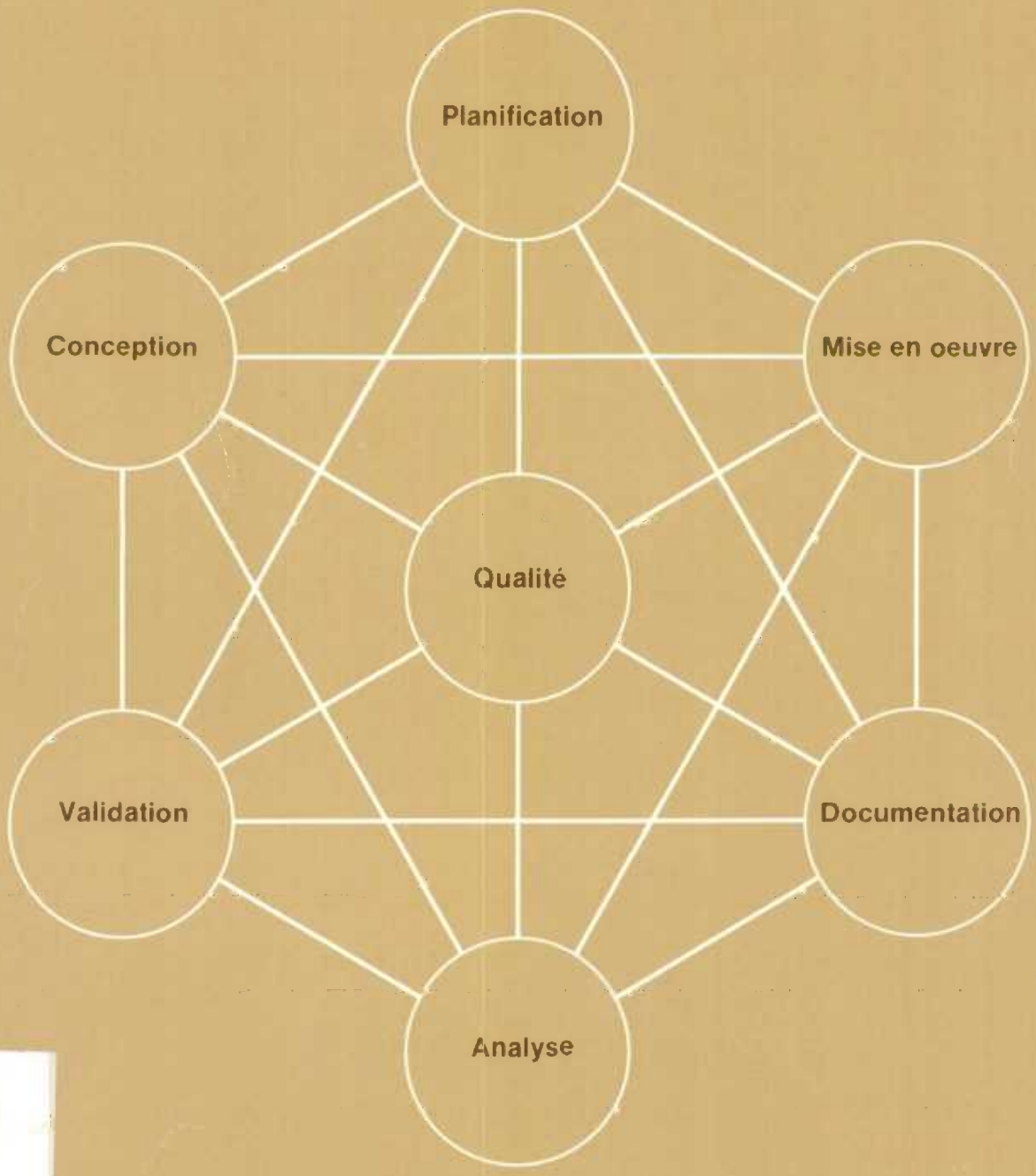
12-539-XPF
1985
c. 3

Statistique Canada Statistics Canada

LIGNES DIRECTRICES CONCERNANT LA QUALITÉ

PREMIÈRE ÉDITION · SEPTEMBRE 1985



Canada

STATISTICS CANADA STATISTIQUE CANADA
MAY MAI 10 2001
LIBRARY BIBLIOTHÈQUE

Statistique Canada

LIGNES DIRECTRICES CONCERNANT LA QUALITÉ

PREMIÈRE ÉDITION · SEPTEMBRE 1985

Publication autorisée par
le ministre des Approvisionnements et
Services Canada

© Ministre des Approvisionnements
et Services Canada 1985

Ottawa

STATISTIQUE CANADA

LIGNES DIRECTRICES POUR LA QUALITÉ DES PRODUITS STATISTIQUES:

Un manuel proposant des conseils pour la production, le maintien et la promotion de la qualité des produits statistiques.

Document préparé au nom du

comité des méthodes et normes

par

Harry Freedman, Division des industrie

Jim Booth, Division des méthodes d'enquête - institutions et agriculture

Jean-François Gosselin, Division des méthodes de recensement et d'enquêtes-ménages

Shaila Nijhowne, Division du produit national brut

Innis Sande, Division des méthodes d'enquêtes-entreprises

Septembre 1985
Première édition

THESE GUIDELINES ARE ALSO AVAILABLE IN ENGLISH

PRÉFACE

Ce manuel constitue un recueil succinct des méthodes, procédures et pratiques qui régissent la poursuite des objectifs qualitatifs dans l'exécution de l'activité statistique. Il résume les moyens qui ont été élaborés à Statistique Canada sous forme d'un ensemble de lignes directrices qui constitute un instrument de contrôle utile tant pour les planificateurs que pour les chefs des enquêtes.

Il faut bien préciser qu'il s'agit de lignes directrices et non de normes. Ces lignes directrices fournissent des indications supplémentaires, mais elles ne sauraient en aucune façon se substituer à des connaissances spécialisées et à un jugement éclairé. Il est certain que toutes les recommandations formulées ici ne s'appliquent pas uniformément à toutes l'activité statistique. Et même le cas échéant, la décision de suivre certaines voies de préférence à d'autres doit reposer sur un jugement éclairé quant à ce qui produira en définitive la meilleure qualité possible et offrira le maximum d'avantages à un coût acceptable.

L'objet de telles mises en garde est de situer le manuel dans son contexte, et non de justifier l'inaction.

Chaque responsable de l'activité statistique a l'obligation de faire en sorte que la recherche de la qualité préconisée par le Bureau se reflète adéquatement dans les méthodes et les procédures statistiques.

J'estime que la préparation de ce manuel marque une étape importante dans la recherche de la qualité à Statistique Canada, et je tiens à féliciter les auteurs et toutes les personnes qui y ont collaboré. Il est bien entendu qu'un tel document est un outil dynamique. L'usage, et l'évolution des techniques et des méthodes, donneront lieu à des mises à jour. Des commentaires ou des recommandations pourront être formulées à n'importe quel moment et adressées à Geoffrey Hole, directeur de la Division des méthodes d'enquêtes-institutions et agriculture.

Ivan P. Fellegi
Statisticien en chef

TABLE DES MATIÈRES

		PAGE
PRÉFACE		(i)
TABLE DES MATIÈRES		(iii)
0.	INTRODUCTION	1
	0.1 Introduction	1
	0.2 Mode de présentation et contenu des lignes directrices	1
	Bibliographie	2
1.	PLANIFICATION	5
	Introduction	5
	Lignes directrices concernant la qualité	6
	1.1 Consultation	6
	1.2 Planification	7
	Bibliographie	9
2.	CONCEPTION ET ÉLABORATION	11
	Introduction	11
	Lignes directrices concernant la qualité	11
	2.1 Base de sondage	11
	2.2 Échantillonnage	12
	2.3 Concepts	13
	2.4 Questionnaire	14
	2.5 Collecte	15
	2.6 Vérification manuelle/codage et saisie des données	15
	2.7 Vérification et correction	16
	2.8 Imputation/estimation	17
	2.9 Procédures	18
	2.10 Élaboration des systèmes	18
	Bibliographie	19
3.	MISE EN OEUVRE	21
	Introduction	21
	Lignes directrices concernant la qualité	21
	3.1 Évaluation des erreurs	21
	3.2 Couverture	21
	3.3 Échantillonnage	22
	3.4 Collecte des données	22
	Questionnaires remplis par les répondants	22
	Questionnaires remplis par les interviewers	23
	Mesures de suivi	23

TABLE DES MATIÈRES (suite)

		PAGE
3.5	Préparation des documents et codage	24
3.6	Saisie des données	25
3.7	Vérification et correction	26
3.8	Imputation et estimation	27
3.9	Traitement automatisé des données	28
	Bibliographie	29
4.	VALIDATION	31
	Introduction	31
	Lignes directrices concernant la qualité	31
4.1	Notions et définitions de base	31
4.2	Couverture	32
4.3	Réponse	34
4.4	Non-réponse	35
4.5	Saisie, codage et préparation des documents	36
4.6	Estimation	36
4.7	Plan d'échantilonnage	37
4.8	Modèles	38
	Bibliographie	38
5.	ANALYSE	39
	Introduction	39
	Lignes directrices concernant la qualité	39
5.1	Préparation des données	39
5.2	Distribution, corrélation, modélisation, réduction	40
5.3	Analyse longitudinale	42
5.4	Analyse des séries chronologiques	43
	Bibliographie	44
6.	DOCUMENTATION	47
	Introduction	47
	Lignes directrices concernant la qualité	47
	Exhaustivité	47
6.1	Objectifs et notions de base	47
6.2	Méthodologie	48
6.3	Tests	48
6.4	Systèmes	49
6.5	Contrôle de la qualité	49
6.6	Opérations	50
6.7	Mise en oeuvre	50

TABLE DES MATIÈRES (suite)

		PAGE
	6.8 Validation	51
	6.9 Analyse	51
	6.10 Ressources	51
	6.11 Documentation	51
	Facilité de Compréhension	52
	6.12 Facilité de compréhension et acceptation professionnelle	52
	Bibliographie	52
7.	DIFFUSION	53
	Introduction	53
	Lignes directrices concernant la qualité	53
	7.1 Planification et conception	53
	.1 Délais de production	53
	.2 Programmes régionaux	54
	.3 Données exploitables par machine	54
	.4 Publications courantes	54
	.5 Demandes spéciales	55
	7.2 Mise en oeuvre	55
	7.3 Documentation	55
	Bibliographie	57
8.	ACTIVITÉS STATISTIQUES FONDÉES SUR L'UTILISATION DE DONNÉES ADMINISTRATIVES	59
	Introduction	59
	Lignes directrices concernant la qualité	60
	8.1 Production des données	60
	8.2 Objectifs liés à l'activité statistique et à la qualité	60
	8.3 Couverture	60
	8.4 Conception et mise en oeuvre	60
	8.5 Vérification et imputation	60
	8.6 Méthodes d'estimation	61
	8.7 Documentation	61
	8.8 Évaluation de la qualité	61
	8.9 Communication suivie avec le fournisseur	62
	Bibliographie	62
9.	ACTIVITÉS STATISTIQUES DÉRIVÉES	65
	Introduction	65
	Lignes directrices concernant la qualité	65
	9.1 Du point de vue des activités dérivées	65
	9.2 Du point de vue des activités de base	66
	Bibliographie	67

0. *INTRODUCTION*

0.1 INTRODUCTION

Les présentes lignes directrices ont été établies pour aider les personnes responsables de la planification, de la gestion, de la conception et de la mise en oeuvre d'activités statistiques à Statistique Canada. Elles ont pour objet d'expliquer ce qu'on entend par "pratiques statistiques efficaces généralement reconnues" et de contribuer au maintien et à l'amélioration de la qualité des produits statistiques.

Il importe de souligner qu'on ne s'attend pas à ce que toutes les lignes directrices indiquées soient strictement observées. Un grand nombre d'entre elles sont déjà d'usage courant, et d'autres sont mentionnées par souci d'exhaustivité; certaines s'appliquent uniquement à certains genres d'activité, et d'autres sont coûteuses et ne peuvent pas être adoptées faute de ressources suffisantes. Une application absolue de toutes les lignes directrices pourrait en fait avoir un effet contraire à l'effet recherché, et même représenter un coût prohibitif, en ce sens que la fonction d'assurance de la qualité submergerait les messages importants sur lesquels il faut de toute évidence faire porter son attention. Ces lignes directrices n'en demeurent pas moins un moyen de contrôle utile pour déceler des faiblesses ou des lacunes évidentes. On n'a pas cherché à déterminer si certaines pratique sont généralement plus importantes que d'autres. Cette démarche est laissée à la discrétion de la personne chargée de chaque activité statistique.

On s'attend à ce que les responsables s'efforceront de suivre les recommandations formulées ici, dans la mesure où elles sont pertinentes et où les ressources le permettent. Comme il s'agit d'un premier effort, il est probable que les utilisateurs de ce manuel y trouveront des lacunes ou des ambiguités. Tous commentaires et toutes recommandations proposant des modifications peuvent être adressés à Geoffrey Hole, directeur de la Division des méthodes d'enquêtes-institutions et agriculture. Des éditions révisées de ce manuel seront produites au besoin.

0.2 MODE DE PRÉSENTATION ET CONTENU DES LIGNES DIRECTRICES

Le mode de présentation des lignes directrices a été déterminé en fonction du déroulement d'une activité statistique type. Elles suivent le cheminement du processus qui est illustré à la figure 1 - Modèle général du processus d'activité (enquête) statistique. Ce modèle vise à expliquer le processus complexe de planification qui précède la décision d'initier (ou de réviser) une activité statistique. Il illustre ensuite les différentes étapes liées à l'activité proprement dite, notamment la conception et l'élaboration, la mise en oeuvre, la validation, la diffusion, l'analyse et la documentation, et donne un aperçu de la rétroaction à obtenir entre ces étapes. Le modèle démontre enfin que le processus ne se termine pas avec la diffusion des données finales, mais qu'il s'étend plutôt jusqu'à l'utilisation qui est faite de ces données par la clientèle et jusqu'à l'analyse de leurs réactions en tant que "consommateurs" du produit. Les étapes susmentionnées sont traitées dans des chapitres distincts du présent document. La figure 2, qui est une simplification de la figure 1, indique le sujet de chaque chapitre. Les deux chapitres qui s'ajoutent à la fin traitent de questions particulières aux activités fondées sur l'utilisation de données administratives et aux activités dérivées. Les chapitres 2, 3 et 4 s'appliquent principalement aux enquêtes-rencensements et aux enquêtes-échantillon, tandis que les autres chapitres ont une portée plus générale. Les chapitres 8 et 9 devraient être lus conjointement avec ces autres chapitres.

BIBLIOGRAPHIE

BOOTH, J.K., "Product Reliability through Quality Assurance", document interne de Statistique Canada, avril 1979.

FELLEGI, I.P., "Data, Statistics, Information - Some Issues of the Canadian Social Statisticts Scene", *Méthodes d'enquête*,V No. 2 (1979), pp. 130-161.

FELLEGI, I.P., "Notes on Quality Assurance of Statistics Canada Outputs", document interne de Statistique Canada, 16 janvier 1981.

MOSER, SIR C. ET. COLL., "Statistics Canada Methodology Review", chapitre V, "Quality Assurance in Surveys", mars 1980.

MOSES, L.E. "Energy Information Validation", 15 novembre 1978.

PLATEK, R., "Methodological Validation Project: A Proposal", 17 septembre 1980 (document provisoire).

STATISTIQUE CANADA, "Ligne de conduite concernant la production et la diffusion des mesures de la qualité des données et les rapports de méthodologie pour les enquêtes de Statistique Canada et leurs résultats", 9 février 1978.

SWAIN, L. ET COLL., "Survey Design Standards and Guidelines", document interne de Statistique Canada, septembre 1980, (version provisoire incomplète).

U.N., ECONOMIC COMMISSION FOR EUROPE, Conférence des statisticiens européens "L'Actualité, le coût et la qualité des statistiques", 8 juin 1982 (CES/464).

U.S., BUREAU OF LABOUR STATISTICS, *Handbook of Methods*, Bulletin 1910. 1976.

U.S., ENERGY INFORMATION ADMINISTRATION, *An Assessment of the Quality of Principal Data Series of the Energy Information Administration*, avril 1983.

U.S., EXECUTIVE OFFICE OF THE PRESIDENT, OMB, Statistical Policy Division, *Federal Statistics - Coordination, Standards*, Guidelines 1976.

U.S., NATIONAL CENTER FOR HEALTH STATISTICS, *Manual on Standards and Procedures for Reviewing Statistical Reports*, mars 1976.

WELLS, J.S., "Quality Assurance at Statistics Canada", document interne de Statistique Canada, janvier 1981.

Figure 1: Modèle général du processus d'activité (enquête) statistique

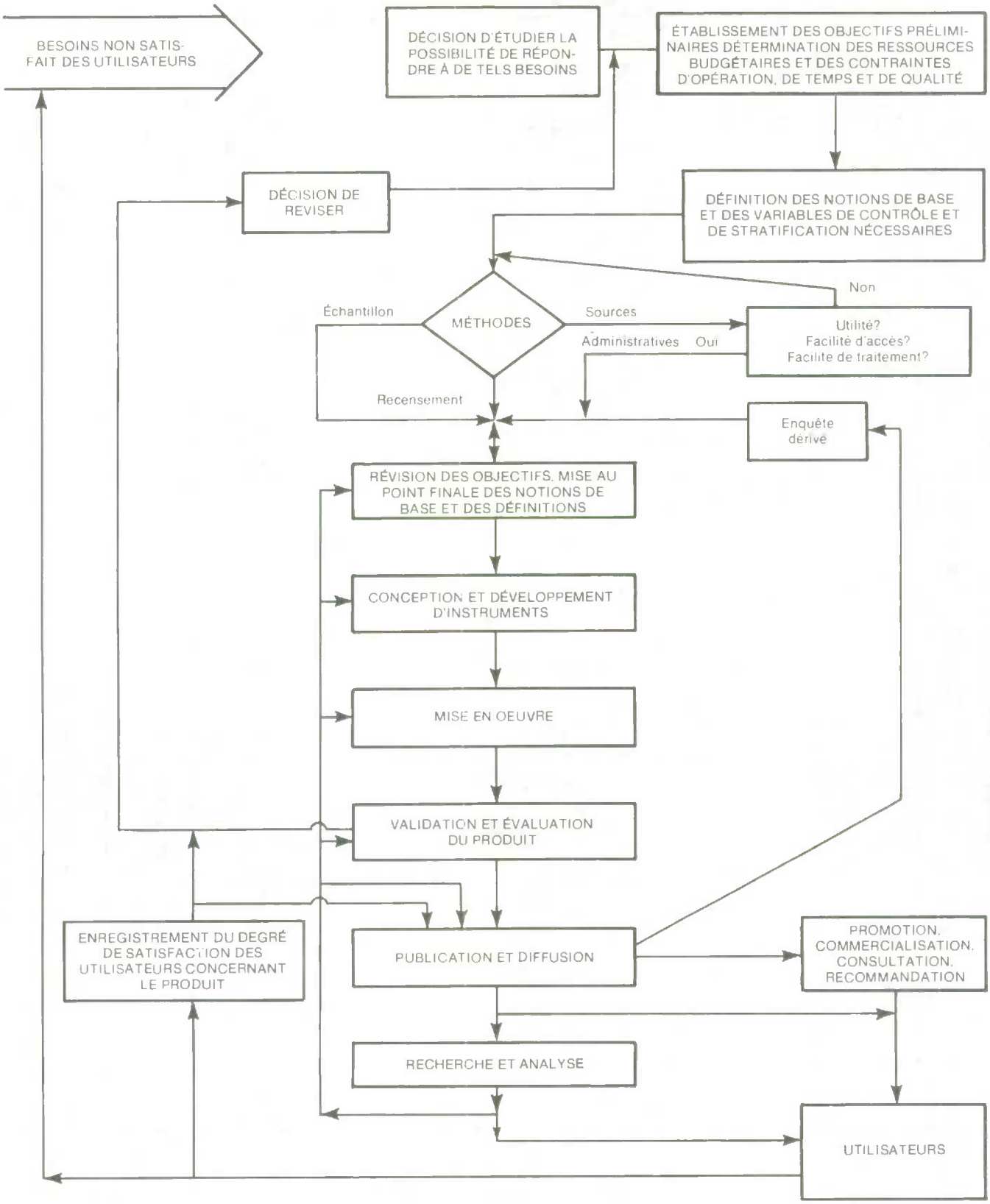
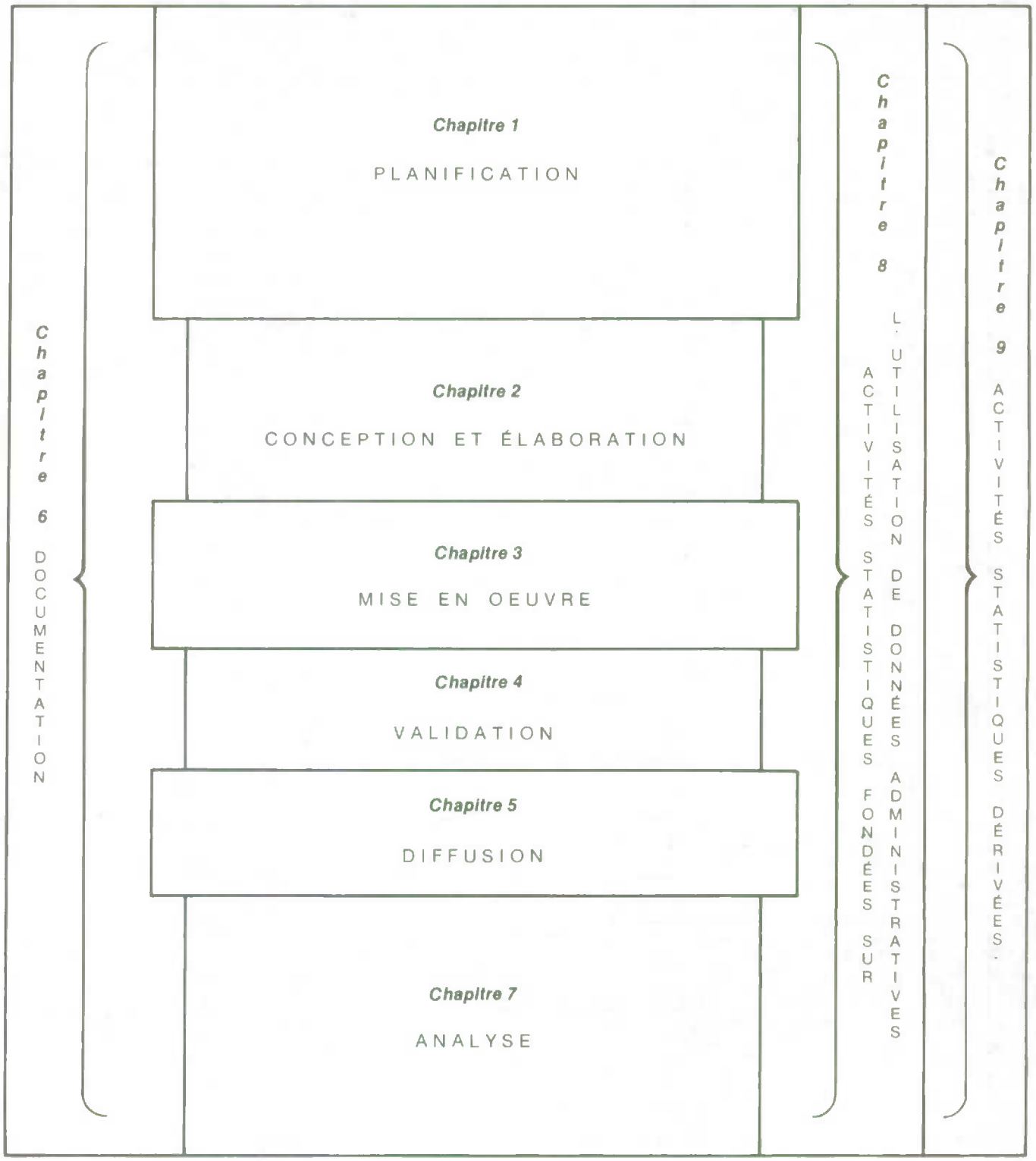


Figure 2: Simplification du modèle général et illustration du mode de présentation des lignes directrices



INTRODUCTION

Nous supposons que la décision d'entreprendre ou de réviser l'activité statistique visée a déjà été prise et qu'elle a été fondée sur une étude préliminaire, quoique sommaire, visant à offrir une vue d'ensemble du problème à résoudre ou de la clientèle à servir et de ses besoins d'analyse. De plus,la nécessité et la faisabilité générale du projet ou de l'activité en question ont généralement été reconnues, et on a fait une certaine estimation des coûts et des besoins en ressources probables.

L'étape suivante consiste donc à nommer un chef et à entreprendre une étude de planification plus détaillée de l'activité ou du projet.

On nomme normalement un chef qui est responsable de tous les aspects du travail. Il doit rendre compte à la haute direction (ou à un comité de direction formé de membres de la haute direction de Statistique Canada et, si possible, de représentants des utilisateurs). Il se peut qu'on crée également un groupe de travail ou d'une équipe de planification pour aider le chef à s'acquitter de la tâche de planification. Ce groupe de travail ou cette équipe de planification doit être formé de personnes possédant l'expérience et les connaissances requises dans tous les domaines de spécialisation visés, par exemple la méthodologie, le contenu et l'analyse, le contrôle statistique de la qualité, les opérations sur le terrain, la conception des systèmes, l'établissement des normes et la diffusion.

Le processus de planification est par nature un processus répétitif comprenant une série d'étapes qui reviennent une ou plusieurs fois. Les étapes suivantes sont normalement comprises dans ce processus:

- une bonne compréhension et une définition claire des objectifs visés, exprimés en termes du problème à résoudre, des hypothèses a vérifier ou des données a recueillir (c'est là un idéal; plus on s'en approche, plus il est facile de déterminer dans quelle mesure la conception et la mise en oeuvre correspondent aux plans);

- la détermination des besoins en données, l'établissement des notions de base et des définitions requises;

- l'analyse des statistiques déjà produites dans le domaine, plus précisément les sources, la fréquence, la qualité, etc. de ces données, de même que le relevé de leurs lacunes compte tenu des besoins établis plus haut;

- la conception de solutions suivant plusieurs approches méthodologiques et une étude de faisabilité de chaque solution en tenant compte des facteurs comme les méthodes de collecte de données, la fréquence de diffusion, le niveau de précision géographique, les coûts, les délais requis et le degré de précision souhaité;

- l'examen des diverses solutions possibles et le choix d'un plan d'action et, après consultation avec les utilisateurs, intégration de leurs représentants aux travaux de l'équipe de planification, dans la mesure du possible;

- la réalisation d'un compromis entre le budget, le fardeau de la réponse et le respect de la vie privée;

- l'établissement d'un plan et d'un échéancier détaillés.

Les activités statistiques permanentes doivent faire l'objet d'un examen à intervalles réguliers. Les objectifs de l'activité doivent être examinés pour assurer que les produits statistiques répondent toujours aux besoins des utilisateurs, lesquels ont pu évoluer ou changer. On doit également revoir les notions de base et les définitions utilisées pour s'assurer de leur pertinence en tenant compte de l'évolution du monde dans lequel nous vivons, des besoins des utilisateurs et de la capacité des répondants à fournir l'information recherchée (Voir Bonnen (1975), pp. 754-5).

Si une révision de l'activité s'impose à la suite d'un tel examen, on devra entreprendre un processus de planification semblable à celui qui vient d'être décrit.

La qualité de la phase de planification est déterminante pour la qualité de l'ensemble du projet. (Même s'il est impossible de faire une évaluation quantitative de la qualité du travail de planification, il est évident que, si on ne parvenait pas à s'entendre au sujet des objectifs, des définitions et des notions de base, il ne pourrait y avoir de compréhension commune de la portée du projet ou de l'activité. Par conséquent, des problèmes pourraient surgir, par exemple à la conception des questionnaires, à la production des publications ou à l'harmonisation des données produites avec celles d'autres sources. Si les ressources budgétaires n'étaient pas clairement définies, on ne pourrait prendre aucune décision rationnelle au moment de la conception de l'enquête. Si les responsabilités des membres de l'équipe de planification n'étaient pas précisées, certains travaux seraient faits en double ou certaines tâches de planification essentielles seraient oubliées, et ainsi de suite.)

LIGNES DIRECTRICES CONCERNANT LA QUALITÉ

OBJET

- Garantir la qualité du processus de planification.

LIGNES DIRECTRICES

1.1 CONSULTATION

C'est au cours de la phase de planification d'une activité statistique éventuelle que l'on définit de manière précise les besoins de la clientèle et que l'on détermine les moyens que Statistique Canada peut prendre pour satisfaire ces besoins ainsi que la mesure dans laquelle on prévoit que ces besoins seront satisfaits.

Il est par conséquent essentiel de s'assurer que les mécanismes d'organi-sation nécessaires sont en place et qu'une consultation a lieu, entre Statistique Canada et ses clients éventuels, au cours de chacune des étapes du processus de planification. Les objectifs visés sont les suivants :

- établir la nécessité du projet (ou de la révision de l'activité);
- s'entendre sur les objectifs;
- définir en détail les besoins des utilisateurs;
- établir les notions de base et les définitions; et
- évaluer dans quelle mesure l'activité proposée peut satisfaire à ces besoins compte tenu des ressources budgétaires et des délais impartis.

1.2 PLANIFICATION

1. Le processus de planification est normalement répété constamment, en reprenant les étapes suivantes:

   - la rédaction et l'acceptation d'un énoncé des objectifs, c'est-à-dire la description de l'objet de l'activité et la définition des besoins en données, y compris le niveau de qualité prévu, après un examen des données disponibles; et

   - la formulation de diverses solutions possibles, notamment des solutions de compromis (par exemple entre les objectifs de contenu et les objectifs de précision géographique) et la réalisation d'une étude de faisabilité des aspects tels que le champ d'observation, le niveau de précision géographique, la méthode de collecte, la fréquence de diffusion, la qualité, les coûts et les délais requis.

2. Le processus de discussion et de consultation vise à déterminer :

   - les contraintes sur le plan des ressources budgétaires et autres ressources dont on tient compte dans

   - le choix d'un plan, ou la modification des objectifs et la conception d'un nouveau plan d'action.

3. Le processus de planification prend fin avec la production, par le chef, d'un énoncé des objectifs et plan de mise en oeuvre, élaboré et documenté avec soin.

   Ce document fait normalement état des points suivants :

   OBJET

   L'objet de l'activité, c'est-à-dire

   - le ou les sujets d'intérêt;

   - les principaux problèmes de nature politique, sociale, économique ou autre sur lesquels les travaux d'analyse de l'activité portent;

   - les hypothèses, les modèles ou les notions qui sont à la base de l'activité statistique, s'il y a lieu.

   CONTENU:

   La nature de l'information requise, c'est-à-dire

   - les notions, les définitions et le degré de précision particuliers au phénomène que l'on veut mesurer;

   - la définition du cadre conceptuel des unités géographique et leur ventilation;

   - déterminer si les données requises visent à mesurer des niveaux ou des changements (tendances) ou les deux.

CHAMP D'OBSERVATION:

La détermination de l'univers ou de la population cible, qu'il s'agisse d'un recensement ou d'un sondage, et la façon dont les petites unités ou les cas problèmes seront pris en compte.

MÉTHODE DE COLLECTE OU DE COMPILATION:

Données d'enquête, données administratives ou données provenant d'autres sources; population complète ou échantillon et méthode utilisée pour la collecte ou la compilation des données.

COHÉRENCE:

Comment on procédera à une comparaison et à l'harmonisation des données avec des données connexes grâce à l'utilisation de notions ou de définitions uniformisées.

FRÉQUENCE:

La fréquence d'observation et de déclaration - unique ou périodique. Pour déterminer le moment propice à l'activité, on doit tenir compte des changements cycliques, saisonniers ou autres du phénomène que l'on désire mesurer.

CALENDRIER:

Déterminer le moment où les résultats devraient être rendus publics ainsi que le délai prévu entre la date du début de l'activité et la date où les utilisateurs pourront avoir accès aux données. À cette fin, on devra tenir compte du temps nécessaire aux essais préliminaires et aux différentes étapes de la conception, du contrôle, du codage, du traitement et des totalisations des données.

DÉLAI:

Combien de temps après la fin de la période d'observation les données seront publiées.

MODE DE DIFFUSION:

La forme sous laquelle les données seront présentées.

QUALITÉ:

Autant que possible, le niveau de la qualité des données visé exprimé en fonction d'aspects mesurables de la qualité (par exemple, taux de réponse, erreur d'échantillonnage, taux d'observation). Dans le cas des données administratives et des activités dérivées, la qualité de l'information produite peut être liée directement à celle de l'information introduite.

ORGANISATION:

L'organisation prévue sur le plan du personnel, des fonctions, des domaines de responsabilité, etc. Il faut assurer, dans toute la mesure du possible, la continuité entre le travail de planification et le travail de mise en oeuvre de l'activité visée.

PLAN D'ACTION:

Ventilation de l'activité proposée selon l'ordre des tâches à exécuter. Ce plan pourra être révisé au cours de la mise en oeuvre de l'activité, mais il est nécessaire d'analyser le programme en détail pour pouvoir planifier le travail et estimer les besoins en ressources.

ESTIMATION DES COÛTS OU ÉTABLISSEMENT DU BUDGET:

Les coûts prévus, qui déterminent les besoins en ressources financières et autres, basés sur les coûts de chacune des étapes de la collecte et du traitement des données, sans oublier les essais préliminaires et le contrôle de la qualité.

CONTRAINTES:

La liste de toutes les contraintes, par exemple celles qui ont trait aux méthodes de mise en oeuvre choisies ou à l'accessibilité des données.

Dans le cas des projets statistiques permanents faisant l'objet d'une révision, on doit entreprendre un processus de planification du même genre pour pouvoir apporter les modifications nécessaires à l'énoncé des objectifs, aux notions de base et définitions, aux exigences touchant la production et la qualité des données et au plan de mise en oeuvre détaillé, dans le sens des recommandations ci-dessus. Dans le cas d'une révision, un autre point doit être pris en compte:

CONTINUITÉ:

L'établissement de plans et de méthodes permettant de combler les écarts ou de concilier les données avant et après une révision en vue de minimiser les effets négatifs de l'interruption dans la série. Il s'agit là d'un élément essentiel du processus de planification entrepris dans le cadre d'une révision.

Grâce à un énoncé des objectifs et un plan de mise en oeuvre établis en fonction des recommandations ci-dessus, les utilisateurs pourront évaluer dans quelle mesure l'activité statistique répond à leurs besoins. Il en sera de même pour les utilisateurs dont les objectifs peuvent être différents. C'est également le moyen de faire connaître à tous les utilisateurs à quoi peuvent raisonnablement servir les produits de l'activité statistique visée et de leur permettre de juger dans quelle mesure ils doivent faire preuve de prudence si l'utilisation qu'ils désirent en faire déborde les limites prévues.

BIBLIOGRAPHIE

MAYDA, F. & TIMMONS, P., "Survey Maintenance - Philosophy and Planning", *Survey Methodology*, VII No. 1 (June 1981), pp 1-20.

BONNEN, JAMES T., "Improving Information on Agriculture and Rural Life", *Amer. J. of Agric. Econ.*, LVII (1975), pp. 753-63.

FELLEGI, I.P. & SUNTER, A.B., "Balance Between Different Sources of Survey Errors - Some Canadian Experiences", *Sankhya* XXXVI Series C, pp. 119-142.

GONZALEZ, MARIA E., ET. AL., "Standards for the Discussion and Presentation of Errors in Survey and Census Data", J.A.S.A. LXX (September 1975), 5-23.

LESSLER, J.T., FOLSOM, R.E. AND KALSBEEK, W.D., *A Taxonomy of Error Sources and Error Measures for Surveys*, Project RTI/1791/00-03F. (Research Triangle Institute, Research Triangle Park, North Carolina, 1983).

MAYER, C.S., "Quality Control in Research", *Journal of Advertising Research*, XXI No. 3 (June 1971).

ZARKOVICH, S.S., *Quality of Statistical Data*, (Rome, F.A.O., 1966).

INTRODUCTION

Maintenant qu'on a convenu des objectifs de l'activité et d'un plan de mise en oeuvre, la prochaine étape est celle de la conception et de l'élaboration des méthodes. C'est le moment d'examiner et de préciser en détail les méthodes qui seront utilisées pour produire les données requises. Il importe de tenir compte de tous les aspects de l'activité afin d'assurer l'uniformité des méthodes et des procédures décrites. Au cours de l'étape de la conception, il est également essentiel de déterminer les procédures qui permettront de contrôler la qualité des méthodes statistiques1 et d'évaluer la qualité des produits statistiques afin d'assurer que ces procédures seront intégrées à l'étape de la mise en oeuvre.

De façon générale, toutes les données recueillies à intervalles périodiques serviront à l'examen des changements dans le temps du phénomène mesuré, que cela soit ou non un objectif explicite. Cet aspect ne doit donc pas être négligé au cours de la phase de la conception de l'activité, particulièrement pour ce qui est du choix de la base de sondage et de la méthode d'échantillonnage.

Quelques-unes des lignes directrices qui suivent ne s'appliqueront peut-être pas à certaines activités. Dans le cas des activités dérivées, seuls les aspects touchant les méthodes d'estimation et l'élaboration des systèmes peuvent se révéler pertinents. Quant aux activités fondées sur les données administratives, on peut ne pas tenir compte des méthodes de collecte, mais il pourra être nécessaire de modifier quelque peu le système administratif afin d'atteindre les objectifs fixés en matière de production de statistiques.

LIGNES DIRECTRICES CONCERNANT LA QUALITE

L'objectif global de l'étape de la conception d'une activité statistique est d'établir un ensemble de méthodes et de procédures qui permettent de réaliser un équilibre adéquat entre les objectifs qualitatifs et les contraintes en matière de ressources. Cela signifie qu'il faut situer la conception dans une perspective globale où l'effet net total de l'ensemble des opérations au niveau de la qualité et au niveau des coûts est pris en considération. Cette idée de conception globale doit être retenue dans l'utilisation des lignes directrices suivantes qui décrivent la façon d'assurer la qualité pour chacun des aspects de la conception de l'activité.

2.1 BASE DE SONDAGE

OBJET

- Choisir une base de sondage ou concevoir un ensemble de procédures permettant de construire ou de mettre à jour une base de sondage de façon à couvrir l'ensemble de la population visée.

LIGNES DIRECTRICES

1. Il faut évaluer la pertinence et la qualité des diverses bases de sondage qui peuvent être utilisées, en tenant compte de facteurs comme la population cible, le degré de sous-dénombrement ou de sur-dénombrement (double compte ou répétitions y compris), la qualité de la classification des éléments de la base et les limites de coûts et autres contraintes opérationnelles. Des études d'évaluation spéciales, des tests pilotes ou autres mesures du genre pourront se révéler nécessaires.

2. Il sera parfois nécessaire de recourir à des techniques particulières en vue d'améliorer, de maintenir ou de contrôler le niveau de la qualité. En voici quelques exemples:

   - comparaison avec d'autres sources;

   - utilisation de bases multiples;

   - élimination des doubles comptes;

   - procédures de mise à jour pour tenir compte des éléments nouveaux (naissances), des éléments supprimés (décès) et des éléments reclassifiés;

   - inclusion de questions spéciales dans le questionnaire en vue d'obtenir une information directe ou prévisionnelle sur les erreurs de couverture possibles, par exemple des questions de classification (permettant d'exercer un contrôle sur les unités hors du champ de l'enquête) ou des questions dont les réponses peuvent aider à l'appariement à d'autres sources;

   - l'utilisation des résultats d'évaluation provenant d'autres enquêtes utilisant la même base de sondage.

3. Dans le cas de bases aréolaires, les cartes doivent être soigneusement vérifiées pour s'assurer que les secteurs géographiques utilisés comme unités d'échantillonnage sont bien délimités et ne se chevauchent pas.

2.2 ÉCHANTILLONNAGE

OBJET

- Concevoir des méthodes d'échantillonnage qui tiennent compte des niveaux de fiabilité cibles et des contraintes de coût.

LIGNES DIRECTRICES

1. Des études spéciales doivent être entreprises pour évaluer les diverses possibilités qui s'offrent pour le choix d'un plan d'échantillonnage, des variables de stratification, de la taille de l'échantillon et de la méthode de répartition de cet échantillon. De telles mesures exigent que l'on tienne compte du niveau d'erreur d'échantillonnage fixé et que l'on se fonde sur toute l'information pertinente, données d'enquêtes et de recensements antérieurs, données administratives et résultats d'enquêtes pilotes.

2. Dans le cas des enquêtes permanentes, il faut concevoir un plan d'échantillonnage suffisamment souple pour que certains de ses éléments puissent être modifiés ultérieurement, au besoin, par exemple si l'on veut accroître ou diminuer la taille de l'échantillon ou encore mettre à jour les probabilités de sélection.

3. Un plan de renouvellement de l'échantillon doit être utilisé chaque fois que l'on veut obtenir des estimations précises d'un changement: le fardeau sera réparti, durant la période de temps entre plusieurs répondants.

4. On doit élaborer des méthodes permettant de résoudre le problème des unités qui changent de strate ou de classification.

5. Au moment de déterminer la taille de l'échantillon, il faut tenir compte dans une certaine mesure des éléments nouveaux ou supprimés de la base de sondage, de la non-réponse et des autres sources de variance non échantillonnale aux étapes de collecte de traitement.

6. Dans le cas des enquêtes permanentes, il faut prévoir des méthodes de contrôle de l'efficacité du plan d'échantillonnage dans le temps.

7. Avant d'arrêter son choix sur un plan d'échantillonnage, on doit tenir compte des conditions dans lesquelles celui-ci sera mis en oeuvre. Cet aspect est particulièrement important dans le cas des enquêtes pour lesquelles la sélection s'effectue sur le terrain. En acceptant une perte d'efficacité, on peut souvent éviter des erreurs dans la sélec tion de l'échantillon qui peuvent entraîner des biais.

2.3 CONCEPTS (voir également 2.4 - Questionnaire)

OBJET

- Concevoir et élaborer des concepts (par exemple, populations statistiques, variables, indices, taux, périodes de référence, unités spatiales) qui satisfont aux objectifs établis de l'activité statistique, et pour lesquels il est possible d'obtenir des réponses précises et cohérentes.

LIGNES DIRECTRICES

Pour pouvoir tirer des conclusions valables d'un ensemble de données, il est essentiel que les utilisateurs aient à la fois une connaissance des concepts et des définitions supportant les données et une connaissance de la qualité de ces données. Autrement dit, ils doivent savoir exactement ce que les données visent à mesurer, et dans quelle mesurent elles y parviennent. Aussi:

1. Les désignations des concepts, et les définitions qui s'y rapportent, doivent refléter de façon précise l'information statistique recueillie.

2. Dans les cas où le concept est clair (par exemple, année civile), mais où la "réalité" est tellement variable qu'il est impossible sur le plan opérationnel d'insister pour qu'on l'emploi, il faut:

   - prévoir recueillir suffisamment de données pour que la réponse puisse être adaptée aux concepts; ou

   - élaborer des procédures qui permettent de mesurer l'effet sur la qualité attribuable à l'écart par rapport au concept et publier les résultats en les associant aux estimations; ou

   - modifier la désignation du concept pour se conformer à l'usage, éliminant ainsi l'erreur d'interprétation éventuelle; ou, si aucun des moyens susmentionnés n'est possible,

   - ne pas employer le concept.

3. Avant de choisir une désignation pour un concept, il faut déterminer:

   - si la désignation communique bien le sens voulu à l'utilisateur;

- si le concept, ou une étroite approximation, est déjà employé par une autre activité statistique. Le cas échéant, on tâche d'utiliser le même concept, la même désignation et la même définition afin d'accroître la cohérence des deux séries;

- si la désignation est déjà employée pour un concept différent (c'est-à-dire, les définitions sont différentes). Le cas échéant, il faut modifier la désignation pour éviter toute erreur d'interprétation entre les deux séries.

2.4 QUESTIONNAIRE

OBJET

- Concevoir un questionnaire (ou de façon plus générale un instrument de collecte) dont on pourra tirer l'information requise selon les notions et définitions déjà établies;

- S'assurer que le questionnaire satisfait aux exigences de toutes les étapes du plan de sondage (collecte, codage, saisie des données, etc.);

- S'assurer que les répondants peuvent répondre aux questions posées de façon précise, tout en réduisant au minimum le fardeau de la réponse.

LIGNES DIRECTRICES

1. Le questionnaire est le moyen de communication entre le répondant et la personne chargée de la compilation des statistiques. Il est essentiel que le répondant comprenne ce que le compilateur attend de lui, et que celui-ci sache ce que l'autre est en mesure de fournir (y compris les systèmes de tenue de dossiers dans le cas des enquêtes-entreprises). (Voir également la Section 2.3 Concepts). Une communication déficiente au niveau du questionnaire se traduit par une erreur de réponse qui aurait pu être évités.

2. On doit se conformer aux normes et règles de conception des questionnaires (Platek et coll. (1983) et Statistique Canada (1979)) notamment pour ce qui a trait au contenu, à la formulation, au mode de présentation et aux formalités administratives (par exemple, la politique concernant les langues officielles).

3. La possibilité d'avoir plusieurs versions d'un questionnaire, chacune visant un segment différent de la population, ou même un questionnaire fait sur mesure (par exemple, le questionnaire de l'enquête sur la population active) peut être considérée du point de vue des relations avec les répondants, des coûts et des avantages. Il faut toutefois user de prudence pour que les différentes versions n'entraînent pas de distorsions dans les réponses en raison des variations au niveau de la présentation et de la formulation des divers questionnaires.

4. Le projet de questionnaire doit être examiné à fond par les participants au projet (responsable, spécialiste de la matière, méthodologiste, analyste de systèmes, personnel des opérations régionales, etc.). De plus, le processus d'examen doit englober les personnes associées aux opérations (par exemple, personnel de la saisie des données) et le plus grand nombre possible de travailleurs spécialisés.

5. On recommande d'effectuer des essais sur le terrain pour vérifier une ou plusieurs versions du questionnaire, particulièrement si on applique de nouvelles notions de base ou de nouvelles questions.

2.5 COLLECTE

OBJET

- Concevoir une méthode de collecte permettant de recueillir des données d'une qualité convenable (celle-ci est habituellement mesurée par le taux de réponse).

LIGNES DIRECTRICES

1. La méthode de collecte des données doit comprendre des procédures de vérification et de suivi pour régler le problème de la non-réponse totale ou partielle de façon à atteindre un taux de réponse conforme aux objectifs de qualité fixés. L'enregistrement de la raison de la non-réponse (refus de répondre, aucun contact) doit aussi être prévu en vue de faciliter l'analyse subséquente.

2. Il ne faut pas porter une attention seulement au taux de réponse global, mais aussi aux variations possibles du niveau de réponse dans les divers segments de la population observée. Les variations importantes doivent être réduites au minimum dans la mesure du possible. Il faudra envisager d'accorder une certaine priorité à des secteurs où la non-réponse est élevée, ou encore à certaines unités (par exemple les grandes entreprises).

3. Des tests pilotes peuvent être effectués en vue d'élaborer une méthode efficace de collecte des données et de préciser certains facteurs comme la pertinence des techniques utilisées, le nombre de rappels nécessaires, les coûts, etc. Ces tests servent également à contrôler la qualité des procédures, de la formation, du questionnaire et des autres instruments de collecte utilisés.

4. Dans le cas des activités statistiques fondées sur des données de fichiers administratifs, il importe de comprendre parfaitement les notions et les procédures qui sont à la base de leur création et de leur maintien afin de pouvoir utiliser et interpréter correctement de telles données.

2.6 VÉRIFICATION MANUELLE, CODAGE ET SAISIE DES DONNÉES

OBJET

- Concevoir des Méthodes de Vérification Manuelle, de Codage et de Saisie des Données permettant d'assurer une correspondance suffisamment fidèle entre les données saisies et les données recueillies, de manière à atteindre les objectifs prévus.

LIGNES DIRECTRICES

1. On devrait évaluer diverses approches possibles de vérification manuelle, de codage et de saisie des données. Dans la mesure du possible, on doit se fonder à cette fin sur les données réelles de questionnaires ayant servi à des enquêtes pilotes, quoique des données fictives peuvent dans certains cas se révéler suffisantes.

2. Se servir autant que possible de listes de codes et de classes uniformisées.

3. Les cas problèmes doivent pouvoir être soumis à des spécialistes (par exemple les codeurs spécialisés).

4. Dans la mesure du possible, un système de correction interactif (par exemple les contrôles d'amplitude et de validité) doit être intégré aux opérations de saisie des données. On doit éviter de procéder à des vérifications manuelles complexes qui exigent un travail d'interprétation ou qui ralentissent la production.

2.7 VÉRIFICATION ET CORRECTION

OBJET

- Repérer tous les éléments d'information dont les valeurs ne se situent pas dans les limites ou les rapports établis;

- Dans tous les cas où c'est possible et faisable, corriger à la main les données rejetées au moyen des liens logiques qui ressortent du questionnaire ou, exceptionnellement, en faisant un suivi auprès des répondants.

LIGNES DIRECTRICES

1. Il faut prendre beaucoup de soin, dans la conception et l'élaboration des lignes directrices 2-7 qui suivent, pour s'assurer qu'il existe un équilibre adéquat entre la découverte de l'erreur et la possibilité d'une survérification (ce qui peut entraîner une dépense excessive de ressources pour modifier des valeurs qui sont extrêmes, mais valides).

2. Les contrôles de vérification et de validation supposent habituellement :

   - l'application de règles de vérification en vue de déceler les données manquantes, non valides ou non compatibles;

   - des comparaisons au micro-niveau avec d'autres sources de données aussi précises;

   - le repérage de données aberrantes.

   Certains de ces contrôles supposent l'établissement de rapports logiques qui découlent directement des notions et des définitions de base tandis que d'autres sont plus empiriques ou découlent de l'application de tests ou de procédés statistiques (par exemple les techniques d'analyse des données aberrantes). On doit se servir autant que possible de données d'enquêtes ou de recensements antérieurs ou de données d'autres sources pour élaborer ou évaluer d'autres méthodes possibles de vérification et de validation.

3. Il faut veiller à assurer l'uniformité entre les différents contrôles de vérification. Certains programmes généralisés, comme le CANEDIT, sont dotés d'un analyseur de règles de vérification pouvant être d'une très grande utilité.

4. Si les ressources budgétaires le permettent, tous les cas de rejet à la vérification doivent être analysés en tenant compte de toute information qui s'y rapporte. Les corrections manuelles sont apportées seulement lorsqu'il est possible de trouver la source du problème et de déterminer la valeur "exacte" (par exemple dans le cas d'une erreur de saisie des données ou lorsqu'un commentaire formulé sur le questionnaire permet de déterminer la bonne réponse). Tous les autres cas doivent être soumis à l'attention des spécialistes du contenu et de l'analyse ou faire l'objet d'une imputation automatique.

5. Lorsqu'il n'est pas possible de passer en revue tous les cas de rejet à la vérification, on conseille de réduire le nombre de ces cas en établissant un ordre de priorité selon la nature ou la gravité de l'erreur ou l'importance de la variable. Quelle que soit l'ampleur du contrôle, on recommande de faire porter l'analyse sur au moins un échantillon des rejets à la vérification en vue de pouvoir déterminer la cause des problèmes, de vérifier fondement des procédures et d'apporter des modifications aux opérations précédentes.

6. Les unités auxquelles on a apporté des corrections doivent faire l'objet d'un nouveau contrôle de vérification pour déterminer si d'autres erreurs n'ont pas été introduites directement ou indirectement. Idéalement toutefois, un mécanisme spécial du système de correction devrait permettre a certains cas d'exception de ne pas subir les contrôles.

7. L'opération de vérification doit être soumise à un contrôle permanent qui permet de déterminer la fréquence des cas de rejet à la vérification ainsi que le nombre et le genre de corrections à apporter. Un tel contrôle permet de faire ressortir les problèmes particuliers qui nécessitent une analyse plus poussée si on veut que les mesures correctives voulues puissent être prises. Ce programme de contrôle constitue un sous-produit du système informatique mis au point pour l'opération de vérification et de validation.

2.8 IMPUTATION ET ESTIMATION

OBJET

- Élaborer une méthode d'imputation permettant de réduire au minimum l'effet des données manquantes ou rejetées;

- Concevoir et élaborer des méthodes d'estimation approximativement sans biais et qui permettront de réduire au minimum l'erreur type ou le coefficient de variation.

LIGNES DIRECTRICES

1. On doit autant que possible évaluer diverses méthodes d'imputation, de pondération et d'estimation possibles en se fondant sur les données de recensements ou d'enquêtes antérieurs, d'enquêtes pilotes ou de sources administratives.

2. La méthode d'imputation doit être conçue de façon à faciliter son contrôle et le repérage des problèmes au besoin. On recommande l'utilisation de symboles de signalisation ou d'indicateurs. L'utilisation de tableaux montrant la fréquence de cas d'imputation, lesquels, permet également d'effectuer une analyse suivie.

3. Des procédés doivent être mis au point pour tenir compte des cas extrêmes, des erreurs de classification, ainsi que des éléments nouveaux ou supprimés de la base de sondage.

4. Une méthode d'estimation de la variance doit être élaborée. La technique d'estimation doit être conforme au plan d'échantillonnage. Elle peut également servir à estimer d'autres sources de variation, comme les erreurs dues à l'interviewer, les erreurs de codage, les erreurs d'imputation, etc.

2.9 PROCÉDURES

OBJET

- Élaborer un ensemble de procédures pour chaque opération d'enquête qui permet d'assurer le respect uniforme de toutes les exigences du plan de sondage. Les procédures opérationelles liées à l'exécution des programmes informatiques sont aussi visées.

LIGNES DIRECTRICES

1. Des manuels de procédures portant sur chacune des opérations doivent être préparés et passés entièrement en revue par tous les participants au projet en vue d'assurer que les notions, les définitions et le plan de sondage soient mis en place correctement, qu'il y ait cohérence entre toutes les opérations et que les procédures soient claires, concises et faciles à interpréter et à suivre.

2. Il faut assurer la formation adéquate de toutes les personnes responsables de l'application des procédures, avec, à l'appui, une documentation bien conçue, pour assurer la bonne marche de chacune de ces activités. La formation doit être prévue, conçue et mise en oeuvre avant que l'opération visée ne soit exécutée. Dans la mesure du possible, ces séances de formation doivent faire l'objet d'une surveillance ou d'un contrôle par l'équipe chargée de la conception.

3. Les manuels de procédures et les documents de formation doivent faire l'objet de tests réalisés dans des conditions qui reproduisent le plus fidèlement possible les conditions réelles de l'enquête.

2.10 ÉLABORATION DES SYSTÈMES

OBJET

- Établir un ensemble de spécifications qui sont exactes, complètes, qui tiennent compte de tous les aspects du plan de sondage, qui sont souples et qui peuvent être maintenues à peu de frais;

- Élaborer des systèmes informatiques permettant de satisfaire ces spécifications;

- Élaborer les procédures opérationnelles nécessaires à l'exécution des programmes.

LIGNES DIRECTRICES

1. Tous les participants au projet doivent passer soigneusement en revue l'ensemble des spécifications afin d'assurer l'application en tout point conforme des notions de base et du plan de sondage.

2. Chaque fois qu'il est possible, on doit utiliser des programmes uniformisés (généralisés) pour l'élaboration des systèmes informatiques.

3. Les systèmes doivent être conçus de façon à faciliter les changements et la mise à jour.

4. Les systèmes doivent être conçus de facon à réduire au minimum les coûts d'entretient.

5. Des tests de logique et de volume doivent être effectués à l'aide de données créées ou recommandées par l'auteur des spécifications ou le personnel responsable de la production.

6. Les systèmes informatiques doivent permettre l'établissement de rapports qui facilitent le contrôle des opérations et le repérage des problèmes.

7. Les procédures opérationnelles nécessaires à l'exécution des programmes doivent être élaborées conformément aux lignes directrices énoncées en 2.9.

BIBLIOGRAPHIE

.1 GÉNÉRALITÉS

MOSER, C.E., ET KALTON, G., *Survey Methods in Social Investigations*, (New York, Basic Books, 1972).

STATISTIQUE CANADA, *Répertoire des concepts sociaux*, (Ottawa, Division des normes, 1980) no. 12-560 au catalogue.

WARWICK, D.P. ET LININGER, C.A., *The sample Survey: Theory and Practice*, (New York, McGraw Hill, 1975).

.2 CONCEPTION DES QUESTIONNAIRES

CARSON, E.M., "Questionnaire Design: Some Principles and Related Topics", document interne de Statistique Canada, 1973.

PLATEK, R., PIERRE-PIERRE, F., ET STEVENS, P., *Élaboration et conception des questionnaires d'enquête*, (Ottawa, Statistique Canada, 1983).

PLATEK, R., ET ROYCE, D., "The role of the Questionnaire in Survey Design", *Méthodes d'enquête*, VIII (1982) pp. 1-16.

STATISTIQUE CANADA, *CONCEPTION DES QUESTIONNAIRES: Manuel d'atelier*, (Ottawa, troisième édition, 1981).

SWAIN, L., *Survey Design Standards and Guidelines*, chapitre X "Questionnaire Design", document provisoire de Statistique Canada non publié, 1980.

.3 ÉCHANTILLONNAGE

COCHRAN, W.G., *Sampling Techniques*, (New York, Wiley, 1977).

HORVITZ, D.G., "Some Design Issues in Sample Surveys", Dans *Survey Sampling and Measurement*, éd. N. Krishnan Namboodiri, (New York, Academic Press, 1978), pp. 1-11.

MURTHY, M.N., *Sampling Theory and Methods*, (Calcutta, Statistical Publishing Society, 1967).

.4 COLLECTE

DILLMAN, D.A., *Mail and Telephone Surveys: The Total Design Method* (New York, Wiley, 1978).

GROVES, R.M., ET KAHN, R.L., *Surveys by Telephone*, (New York, Academic Press, 19-.

.5 IMPUTATION

HILL, C.J., ET PUDERER, H.A., "Data Adjustment Procedures in the 1981 Canadian Census of Population and Housing". Dans *Current Topics in Survey Sampling*, Kreuski, D., Platek, R., et Rao, J.N.K. (éditeurs) (New York Academic Press, 1981).

SANDE, I.G., "Imputation in surveys: Coping with reality", *Techniques d'enquête*, VII (1) pp. 21-43 et The American Statistician XXXVI (1982), pp. 145-52.

INTRODUCTION

Le champ couvert par l'enquête et la qualité des données recueillies et traitées déterminent les utilisations possibles et les avantages qui pourront être tirés des données de cette enquête. Ces deux aspects importants de toute enquête dépendent principalement de la mesure dans laquelle les méthodes d'enquête et les procédures élaborées à l'étape de la conception sont complètes, à jour (c'est-à-dire conformes aux objectifs fixés) et appliquées de manière adéquate.

LIGNES DIRECTRICES CONCERNANT LA QUALITÉ

Les lignes directrices ci-dessous décrivent les activités qui visent à assurer que les opérations se déroulent conformément aux méthodes et procédures établies.

OBJET

- Assurer des niveaux de réponse adéquats des répondants.

- Prévenir, réduire au minimum ou limiter les erreurs possibles dans le traitement des données recueillies.

LIGNES DIRECTRICES

3.1 ÉVALUATION DES ERREURS

Des erreurs d'observation peuvent se produire et se produisent effectivement à toute étape de l'exécution d'un recensement ou d'une enquête, et il s'agit généralement d'erreurs de couverture, de réponse, de mesure ou de traitement. Chacune de ces sources d'erreur peut avoir un effet différent sur la qualité et l'utilité des résultats. C'est pourquoi il est important de relever les principales sources d'erreur d'une enquête, et d'y apporter une attention particulière.

Pour satisfaire aux objectifs de qualité des données d'enquête (c'est-à-dire réduire au minimum ou prévenir les erreurs au niveau des opéations qui constituent ou pourraient constituer des sources d'erreur importantes), il faut élaborer et incorporer dans le plan de sondage général des méthodes permettant de déceler, de mésurer et d'évaluer ces erreurs. La réaction aux résultats de ces évaluations doit faire partie intégrante de la mise en oeuvre.

3.2 COUVERTURE

Les mesures suivantes peuvent être prises pour aider à évaluer et corriger le problème du surdénombrement ou du sous-dénombrement dans les enquêtes.

1. On peut chercher à mettre à contribution les connaissances et l'expérience du personnel des bureaux régionaux. À cette fin, celui-ci doit être bien informé au sujet de l'objet de l'enquête, du champ d'observation visé, de la base de sondage, des unités faisant partie de la base de sondage et du mode de délimitation des unités directement visées par l'enquête. Les unités inexistantes, hors du champ de l'enquête ou manquantes pourront ainsi être repérées plus facilement.

2. On peut procéder à une vérification des secteurs choisis sur des cartes locales à jour ou grâce à une inspection sur le terrain. Les listes d'unités d'échantillonnage établies peuvent être comparées à des listes locales, par exemple les listes d'adresses des bureaux de poste.

3. La construction et/ou la mise à jour des bases de sondage peut se fonder sur des sources d'information locale. Cela suppose une vérification périodique sur le terrain des secteurs choisis à l'intérieur de la base de sondage.

4. Le personnel responsable du contenu et de l'analyse et le personnel chargé de la conception de l'enquête doivent superviser les séances de formation, particulièrement celles à l'intention des responsables des opérations.

5. Les enquêtes-entreprises ou enquêtes économiques sont caractérisées par un degré élevé de volatilité au niveau du plan de sondage. Celui-ci doit constamment être mis à jour. Il faut y incorporer le plus rapidement possile les créations, les disparitions et les changementsde structure ou de classification. Il faut effectuer des vérifications fréquentes de l'information contenue dans les fichiers suivant une formule d'échantillonnage, et il faut aller chercher certaines données de classification par le moyen de formulaires d'enquête aux fins de la mise à jour. La transmission des informations de l'enquête à la base de sondage doit se faire rapidement et efficacement. Toutes les enquêtes portant sur le même secteur de l'univers économique doivent être effectuées au moyen de la même base de sondage, de façon à éviter les incohérences.

6. Des méthodes d'évaluation de la couverture sont présentées dans la Section 4.2.

3.3 ÉCHANTILLONNAGE

Durant la mise en place de l'échantillonnage, le rendement de l'échantillon doit être surveillé car, à chaque niveau du plan d'échantillonnage, il doit correspondre aux comptes obtenus durant l'élaboration du plan. De plus, si possible, la précision des estimations devrait être surveillée afin d'atteindre les objectifs prévus. A cette étape de l'enquête il peut être possible d'accroître la taille de l'échantillon.

3.4 COLLECTE DES DONNÉES

Le taux de réponse et la qualité des données recueillies sont deux facteurs essentiels au succès de l'enquête.

Dépendamment des méthodes de collecte, les activités permettant d'assurer que ces deux exigences soient satisfaites sont décrites dans les paragraphes qui suivent.

Questionnaires remplis par les répondants

1. Dans la mesure du possible, une bonne publicité dans les différents médias ou une lettre adressée personnellement aux répondants peut

2. Lorsque les répondants remplissent eux-mêmes les questionnaires, il faut les renseigner sur l'objets de l'enquête et sur l'utilisation des données demandées. Pour toutes les enquêtes, on doit fournir au répondants des instructions claires sur la façon de remplir le questionnaire et, si possible, leur indiquer où s'adresser pour obtenir de l'aide.

Questionnaires remplis par les interviewers

3. Des lignes de conduite détaillées doivent être établies pour toutes les tâches des interviewers, soulignant l'objet et l'importance de l'enquête et précisant les vérifications à effectuer au moment de la collecte des données. Ces lignes de conduite doivent être mises à jour pour tenir compte de toute modification apportée aux opérations ou aux méthodes.

4. Les interviewers doivent bénéficier d'une formation poussée sur tous les aspects de la collecte des données. L'accent doit être mis sur l'importance d'obtenir un taux de réponse élevé et d'effectuer les premières vérifications pour s'assurer que les interviewers comprennent parfaitement les exigences concernant la qualité des données et le niveau de réponse. Ces séances de formation doivent faire l'objet d'un programme de supervision ou d'observation, et les observateur doivent documenter leurs constatations pour qu'elles soient réinjectées dans le processus de formation.

5. La formation adéquate et la compétence du personnel chargé de la collecte des données sont particulièrement importantes dans les enquêtes complexes. On recommande de simuler des interviews ou d'utiliser des exercices écrits.

6. Un contrôle de qualité sur l'exhaustivité et l'exactitude du travail de chaque interviewer doit faire partie intégrante du processus de mise en oeuvre. La fréquence ou l'ampleur de tels contrôles dépend de la complexité de l'opération ainsi que de la compétence de chacun des membres de l'équipe. Cette vérification sert également à s'assurer que les lignes de conduite ont été suivies. Les résultats de ces vérifications doivent être consignés aux fins d'utilisation future.

Mesures de suivi

7. On doit établir des lignes de conduite sur la façon de résoudre le problème de la non-réponse. Celles-ci doivent porter sur l'ensemble des mesures, depuis les premières tentatives de contact, suivies d'un appel téléphonique ou de plusieurs appels jusqu'aux visites personnelles. Il faut également prévoir des méthodes particulières dans le cas des secteurs géographiques où le taux de non-réponse est très élevé. De tels contrôles sont parmi les principaux facteurs qui contribuent aux améliorations futures, qu'il s'agisse notamment de modifier les méthodes ou de déterminer l'utilité des lettres de suivi.

8. Des mesures de suivi des cas de non-réponse partielle doivent aussi être prévues. Il importe de faire ressortir les éléments de données les plus importants afin que les mesures de suivi portent en priorité sur les cas les plus importants.

9. Une bonne gestion des documents est nécessaire pour que le personnel chargé de la collecte des données soit au courant des mesures de suivi déjà prises au moment d'entrer en contact avec le répondant visé.

10. Des lignes directrices peuvent être fournies au personnel chargé de la collecte des données relativement au taux de réponse visé ou prévu selon le secteur et le type de répondant. L'ampleur du suivi dépend de l'écart entre le taux de réponse réel et le taux prévu. Ce genre de comparaison doit cependant être documenté et diffusé.

11. Les surveillants doivent contrôler périodiquement le travail de leurs subalternes. La fréquence et l'ampleur de telles évaluations dépendent de la complexité de l'opération, de l'importance des éléments de données visés et de la compétence du

personnel. Ce genre de contrôle permet de déterminer dans quelle mesure les lignes de conduite établies ont été suivies et d'apporter les changements qui s'imposent.

12. Il faut veiller à transmettre aux gestionnaires d'enquête les résultats des contrôles des taux de réponse totale ou partielle.

3.5 PRÉPARATION DES DOCUMENTS ET CODAGE

Ces opérations ont pour objet d'examiner, de transcrire ou de convertir les données recueillies sous une forme qui en facilite la saisie. Comme ces opérations sont le plus souvent manuelles, des erreurs peuvent se glisser. Par conséquent, voici quelques points importants à retenir :

1. Il faut utiliser chaque fois que possible des listes de codes uniformisées.

2. On doit assurer une formation poussée portant sur tous les aspects de l'opération, particulièrement sur l'importance de la qualité des données et les répercussions possibles des erreurs. Les premières séances de formation peuvent faire l'objet d'une supervision afin de s'assurer que tout le personnel a bien compris les objectifs de l'opération et les méthodes devant être utilisées pour les atteindre.

3. Les cas problèmes doivent pouvoir être soumis à l'attention de spécialistes.

4. Étant donné que le pourcentage global d'erreur de codage peut avoir une incidence sur la qualité des données, on doit préciser dans la documentation de formation les normes ou exigences établies à cette fin.

5. Des procédures complètes et faciles à comprendre doivent être établies pour toutes les opérations d'écriture et mises à jour chaque fois qu'un changement est apporté aux opérations ou aux méthodes.

6. Les surveillants doivent contrôler périodiquement le travail de leurs subalternes afin de s'assurer que les lignes de conduite sont bien suivies. Ils doivent également évaluer dans quelle mesure les codeurs respectent les normes établies relativement à la qualité des données. Au début, une vérification intégrale du travail de chaque codeur est habituellement effectuée. Les résultats de cette opération (c'est-à-dire les taux d'erreur enregistrés et leurs répercussions possibles) déterminent le moment où le travail doit être soumis au contrôle statistique de la qualité ou aux vérifications de qualité d'usage ou encore s'il faut prévoir des évaluations ultérieures. La méthodologie employée actuellement pour le contrôle de la qualité des statistiques offre l'assurance que les objectifs concernant la qualité des données découlant des opérations d'enquête seront atteints. Elle permet de déceler les erreurs qui sont introduites, de contrôler et de limiter les erreurs ainsi produites avec un minimum de ressources, de fournir de la rétroinformation à l'intention du codage et des opérations, et de documenter la qualité de l'opération pour ce qui concerne les données produites. La rétroinformation permet d'améliorer le travail de chaque codeur, et le contrôle de la qualité engendre généralement une attitude positive vis-à-vis de la qualité des données.

7. Des procédures de vérification doivent être établies. Il est préférable que ces vérifications soient indépendantes. On doit prévoir à la fois une vérification à 100% et une vérification à l'aide d'un échantillon. Il ne faut jamais perdre de vue les objectifs de qualité des données et d'utilisation optimale des ressources.

8. De la documentation doit être fournie sur le niveau de qualité atteint dans le cas des opérations dont on sait qu'elles peuvent contribuer de façon importante au niveau général d'erreur.

9. Les résultats de l'analyse de la qualité des données doivent être communiqués aux personnes intéressés à tous les niveaux, depuis les codeurs jusqu'aux gestionnaires d'enquête.

10. De bonnes méthodes de contrôle des documents doivent être mises au point pour assurer la bonne marche des opérations principales et des opérations de vérification.

3.6 SAISIE DES DONNÉES

Cette opération est aussi en grande partie manuelle et des erreurs peuvent donc être facilement introduites. Ce travail ne nécessite habituellement pas une connaissance approfondie des différents domaines de spécialisation, car il suffit dans la plupart des cas d'introduire en machine ce qui est écrit sur des documents. Voici les règles à suivre pour assurer la qualité des données dans ce genre d'opération :

1. Dans la mesure du possible, des vérifications interactives doivent être effectuées, particulièrement dans le cas des erreurs pouvant être corrigées par l'opérateur chargé de la saisie des données (c'est-à-dire des vérifications préliminaires des erreurs d'enregistrement). Il faut éviter autant que possible de procéder à une imputation manuelle des données lorsque celles-ci ne figurent pas sur le document ou le questionnaire.

2. En raison des répercussions possibles des erreurs d'introduction en machine, on doit établir des normes ou des exigences touchant la qualité des données. Celles-ci doivent être conformes à celles déjà établies pour les opérations précédentes. On doit évaluer dans quelle mesure les opérateurs satisfont à ces exigences.

3. On doit pouvoir recourir, au besoin, à l'aide de spécialistes.

4. On doit fournir des directives sur la saisie des données ainsi que des lignes de conduite détaillées au sujet de tous les cas de rejet à la vérification.

5. Une formation doit être offerte à tous les opérateurs chargés de la saisie des données afin de s'assurer qu'ils comprennent bien comment procéder pour tous les cas de rejet à la vérification. Cette formation doit également mettre l'accent sur l'importance de la qualité des données et les répercussions possibles des erreurs de saisie.

6. Des méthodes de vérification doivent être établies. Au début, une vérification intégrale de la qualité du travail de l'opérateur s'impose et, selon les résultats, on saura quand il doit être soumis à un contrôle de la qualité statistique, aux vérifications de qualité d'usage ou quand devra avoir lieu la prochaine évaluation de l'opérateur.

7. On doit faire rapport sur le niveau de qualité atteint à cette opération.

8. Les résultats de l'analyse doivent être communiqués à tous les niveaux, depuis les opérateurs jusqu'aux chargés de projet.

9. De bonnes méthodes de contrôle des documents doivent être établies en vue de s'assurer qu'ils peuvent être soumis aux opérations de saisie et de vérification.

10. Il importe que le matériel utilisé pour la saisie des données soit doté d'un logiciel ou de programmes opérationnels qui permettent une vérification par échantillon. De cette façon, la qualité de n'importe quelle combinaison des éléments de données peut être garantie avec un minimum de ressources.

3.7 VÉRIFICATION ET CORRECTION

Le programme détaillé de vérification permet de repérer tous les éléments d'information dont les valeurs ne se situent pas dans les limites ou rapports établis. La correction peut s'effectuer par un procédé manuel d'examen ou à l'aide d'une méthode automatisée d'imputation.

1. Des lignes de conduite détaillées doivent être fournies pour toutes les opérations d'écriture et elles doivent être mises à jour pour tenir compte des modifications apportées aux opérations ou aux méthodes.

2. Une formation approfondie doit être offerte sur tous les aspects de l'opération et doit mettre l'accent sur les principaux cas d'erreur et les mesures qui peuvent être prises pour remédier à ces problèmes, y compris les mesures de suivi.

3. Des tables faisant état de la fréquence des cas de rejet à la vérification et des corrections apportées selon le genre de données doivent être établies et contrôlées sur une base continue. Tous les cas particuliers doivent faire l'objet d'une analyse plus poussée.

4. On doit soumettre les données corrigés au même programme de vérification qui a causé leur rejet.

5. Dans la mesure du possible, on ne doit pas à ce stade imputer des données, sinon établir un rapport sur ces cas (contenant la valeur avant et après l'imputation) en vue de leur analyse ultérieure.

6. On doit inscrire sur les questionnaires des répondants les données manquantes obtenues à la suite d'un suivi téléphonique et préciser de quelle façon et à quel moment ces renseignements ont été recueillis.

7. Comme la charge de travail peut devenir trop lourde pour le personnel en place, tous les cas de rejet à la vérification doivent être classés selon leur importance et la priorité à accorder à leur correction.

8. Les surveillants doivent contrôler périodiquement le travail de leurs subalternes afin de s'assurer qu'ils se conforment aux lignes de conduite établies. Ils doivent également évaluer la qualité de leur travail. Les résultats (c'est-à-dire les taux d'erreur et les répercussions possibles de ces erreurs) déterminent si une certaine forme de contrôle de la qualité est nécessaire.

9. De bonnes méthodes de contrôle des documents doivent être établies en vue d'assurer leur conformité à tous les aspects de l'opération de correction ou aux contrôles d'évaluation appliqués par les surveillants.

10. L'opération de saisie des données corrigées doit se faire conformément aux mesures d'assurance de la qualité énumérées au point 3.5.

11. Les lignes directrices concernant l'assurance de la qualité énoncées au point 3.8 s'appliquent également à toutes les activités opérationnelles qui s'appuient sur l'utilisation de programmes informatiques.

3.8 IMPUTATION ET ESTIMATION

L'imputation est la dernière opération mise en oeuvre pour remplacer des données manquantes ou inacceptables, et l'estimation comprend l'ensemble des méthodes utilisées pour produire les totalisations, les publications et les tableaux définitifs de l'enquête.

Imputation

1. Le programme d'imputation ou la documentation établie à cette fin, si cette opération est manuelle, doivent inclure des indicateurs du nombre et de la fréquence des imputations. Voici quelques exemples de ce genre de mesures:

   (a) identification des cases de non-réponse ou de réponse incohérente;

   (b) comptage des cas de rejet à la vérification;

   (c) repérage des données corrigées ou imputées et identification de la méthode d'imputation qui a été appliquée, s'il y a lieu;

   (d) registre des données avant ou après imputation.

2. Des lignes de conduite détaillées doivent être établies pour toutes les opérations manuelles ou automatisées particulières à ces deux étapes importantes du processus d'enquête. Les méthodes d'imputation doivent prévoir la façon de traiter les cas particuliers, comme les données aberrantes.

3. Une formation approfondie doit être offerte sur toutes les exigences liées à ces opérations.

4. Le processus d'imputation doit faire l'objet d'un contrôle continu qui suppose habituellement :

   - l'analyse de la fréquence des cas d'imputation et de leurs répercussions sur les estimations finales;

   - la production et la comparaison des distributions des données avant et après imputation.

   Les cas de fréquences ou d'écarts élevés doivent faire l'objet d'une analyse plus poussée.

5. Un contrôle périodique de ces opérations doit être effectué afin de s'assurer que les lignes de conduite ont été suivies et que toutes les sources d'information pertinentes ont été utilisées.

Estimation

6. Avant de mettre en place les technique d'estimation, les résultats de l'enquête doivent être analysés:

- pour s'assurer que les conditions prévues pour que les estimations aient une erreur quadratique moyenne minimale existant toujours pour l'estimation du quotient, il doit toujours y avoir un compromis entre la variance d'échantillonnage et la taille du bias;

- pour permettre un traitement adéquat des données aberrante;

- pour permettre un traitement adéquat des cas où un grand nombre d'unités change de strate.

7. Pour produire des séries de données chronologiques, on a souvent recours à des méthodes fondées sur l'utilisation de "jalons". Selon ce genre de méthode, les valeurs obtenues à partir d'une source de données (habituellement un échantillon) sont ajustées en fonction de totaux obtenus à intervalles moins fréquents, mais jugés plus fiables (les jalons).

   La méthode ainsi choisie doit viser à fausser le moins possible l'orientation générale de la série en cause (c'est-à-dire l'évolution des caractéristiques observées de la série originale) tout en ramenant le niveau des estimations à celui de la série de contrôle (c'est-à-dire la série des données jugées plus fiables). (Voir E. Dagum (1974)).

3.9 TRAITEMENT AUTOMATISÉ DES DONNÉES

Il s'agit des programmes informatiques utilisés pour le traitement des données saisies en vue de constituer ce que l'on appelle un fichier de "données épurées". Dans l'optique de l'assurance de la qualité:

1. On doit fournir une documentation détaillée incluant les lignes de conduite sur chaque programme informatique conçu, vérifié et accepté aux fins de l'enquête. En effet, le personnel chargé des opérations doit connaître:

   (a) Les objectifs de chaque programme.

   (b) Les directives de mise en forme et d'exécution de chaque programme, notamment la préparation des données d'entrée, un aspect très important puisque les méthodes peuvent varier selon les régions et le type de données.

   (c) Les fichiers de données utilisés dans le cadre de chaque programme et la façon d'appliquer des contrôles pour assurer l'utilisation des bons fichiers.

   (d) Les directives concernant les résultats de chaque programme, notamment l'importance de chaque produit par rapport à l'ensemble de l'opération et le nom des personnes responsables des résultats du traitement.

2. On doit également établir des directives sur la façon de résoudre les problèmes opérationnels ou mécaniques qui peuvent se produire. Les résultats des tests des programmes et des systèmes peuvent se révéler très utiles à cette fin.

3. Tous les employés chargés des opérations doivent avoir reçu la formation informatique nécessaire à l'exécution de leurs tâches.

4. Une formation approfondie sur tous les aspects des opérations manuelles directement liées aux systèmes informatiques doit également être dispensée.

BIBLIOGRAPHIE

.1 GÉNÉRALITÉS

BAILAR, B.A., ET LANPHIER, C.M., *Development of Survey Methods to Assess Survey Practices* (A.S.A., 1978).

BOOTH, J.K., "Product Reliability through Quality Assurance", document interne de Statistique Canada, avril 1979.

BOOTH, J., ET MUDRYK, W., "Quality Control: Its role in Quality Assurance and its Relation to Survey Operations", (Statistique Canada, 1980), document préparé en vue d'un colloque de formation.

GOSSELIN, J.F., CHINNAPPA, B.N., GHANGURDE, P.D., ET TOURIGNY, J., *Répertoire des méthodes d'évaluation des erreurs dans les recensements et les enquêtes*, (Ottawa, Statistique Canada, 1978), no 13-564F au catalogue.

JURAN, J.M., *Quality Planning and Analysis*, (New York, McGraw-Hill, 1970).

.2 CODAGE/SAISIE DES DONNÉES

JURAN, J.M., *Quality Control Handbook*, troisième édition, (New York, McGraw-Hill, 1974).

DUNCAN, A.J., *Quality Control and Industrial Statistics*, (Illinois, Richard D. Irwin, 1974).

HALD, A., *Statistical Theory of Sampling Inspection By Attributes*, (New York, Academic Press, 1981).

DODGE, H., ET ROMIG, H., *Sampling Inspection Tables*, (New York, J. Wiley and Sons, 1959).

.3 VÉRIFICATION ET CORRECTION

NAUS, JOSEPH I., *Data Quality Control and Editing*, (New York, M. Dekker, 1975).

.4 ESTIMATION

DAGUM, E.B., Comparisons of Various Interpolation Procedures for Benchmarking Economic Time Series, document interne de Statistique Canada, mai 1977.

4. *VALIDATION*

INTRODUCTION

La validation est le terme que l'on utilise pour décrire le processus qui consiste à déterminer dans quelle mesure le produit final de l'enquête satisfait aux objectifs établis.

Certaines études de validation (par ex. les études portant sur la couverture et les macrovérifications) doivent être entreprises de manière systématique tandis que d'autres (par ex. les analyses de validation des notions de base ou des réponses par une réinterview plus détaillée) ne sont réalisées que si les ressources le permettent ou si elles sont jugées particulièrement importantes. Dans le cas d'une enquête ou d'une activité permanente de collecte de données, les études de validation doivent être entreprises périodiquement et non seulement lorsqu'un problème survient, si l'on veut s'assurer de la pertinence du projet.

LIGNES DIRECTRICES CONCERNANT LA QUALITÉ

Les lignes directrices ci-dessous décrivent les activités qui visent à valider le produit statistique et à évaluer dans quelle mesure les objectifs ont été atteints.

4.1 NOTIONS ET DÉFINITIONS DE BASE

OBJET

- Déterminer dans quelle mesure les données ou les estimations fournissent une mesure fidèle des notions et définitions énoncées au point de départ.

LIGNES DIRECTRICES

Microniveau

Des interviews détaillées sur le terrain peuvent être menées auprès d'un échantillon de répondants en vue de déterminer:

(i) si le répondant répond à la définition établie,

(ii) si le répondant a été classé dans la bonne catégorie,

(iii) si le répondant a compris de la manière prévue les notions et les définitions qui lui ont été présentées,

(iv) si le répondant a fourni une information exacte.

Des interviews peuvent aussi être menées auprès d'un échantillon d'unités hors du champ de l'enquête pour déterminer:

(v) si certains répondants éventuels n'auraient pas été oubliés.

À noter que les points (i), (ii) et (v) peuvent aussi être précisés dans le cadre d'une étude de validation de la couverture et que le point (iv) est un élément dont on tient compte dans une étude de validation des réponses. Il convient également de souligner que les études de validation se font après l'enquête proprement dite. Une étude de ce genre peut aussi être menée à l'étape de la planification pour tester la base de sondage, le questionnaire ou la capacité des répondants éventuels à fournir l'information requise.

Selon le nombre de cas qui satisfont aux conditions (i), (ii), (iii), (iv) ou (v), on peut obtenir une mesure de la qualité de concrétisation des notions de base.

Macroniveau

Dans certains cas, les statistiques sont fondées sur des systèmes de classification qui supposent l'homogénéité pour des unités qui sont en fait essentiellement hétérogènes. C'est ainsi que, par exemple, les entreprises sont classées dans une catégorie industrielle en fonction de leur activité principale, et les activités secondaires ne sont pas représentées. Lorsqu'on recueille des données pour un univers de ce genre, une partie de l'activité totale d'une industrie spécifique porte sur des aspects qui se situent en dehors de la catégorie industrielle pertinente. Pareillement, une partie de l'activité industrielle pertinente se situe dans des catégories qui sont en dehors de l'industrie spécifique.

À ce moment-là, deux mesures de la qualité de la classification (le système comme étant appliqué à l'univers complet et non à l'opération de codage) peuvent être calculées:

- le **coefficient de spécialisation**, ou la proportion de l'activité spécifique à l'industrie (par exemple, valeur des livraisons) par rapport à l'activité totale déclarée par cette industrie; et

- le **coefficient de couverture**, ou la proportion de l'activité totale spécifique à l'industrie représentée par les entreprises portant le code de cette industrie.

Ces mesures sont difficiles à estimer, de sorte qu'il faut prendre des moyens afin de recueillir assez de données pour y parvenir[1]. Autrement, il faut trouver un substitut quelconque.

4.2 COUVERTURE

OBJET

- Déterminer dans quelle mesure la population cible visée a été observée.

LIGNES DIRECTRICES

Microniveau

1. À partir de listes d'unités comparables (ou d'un échantillon d'unités de ces listes), on peut procéder à un appariement des unités de la base de sondage courante et des unités de ces autres listes possibles. Après analyse des résultats, le pourcentage de cas inexpliqués de non-appariement sur chaque liste fournit une mesure de l'erreur

---

[1]On peut trouver des exemples de ces mesures dans:
U.S. Dept. of Commerce, Bureau of the Census **1977 Census of Manufactures** (divers bulletins concernant la spécificité des industries); Ministère fédéral de la consommation et des corporations, **Concentration dans les industries manufacturières au Canada**, Ottawa 1971; U.S Dept. of Commerce, Bureau of the Census, **1977 Census of Manufactures: Concentration Ratios in Manufacturing** (Wahsington D.C., U.S.G.P.O. mai 1981).

couverture (sous-dénombrement ou surdénombrement). Il est possible d'estimer le nombre des unités qui ne figurent sur aucune liste grâce à l'application de techniques d'estimation par saisie-resaisie (voir Bishop, Fienberg et Holland, **Discrete Multivariate Analysis,** chapitre 6). Ces techniques peuvent aider à déterminer les causes d'un sous-dénombrement.

2. On peut prélever un échantillon de données d'une source externe très fiable correspondant à un sous-ensemble de la population cible visée et on procède à un appariement des unités de l'échantillon et de la base de l'enquête. Les unités de l'échantillon qui ne se retrouvent pas dans la base de sondage fournissent une estimation du sous-dénombrement (p. ex. après contre-vérification des dossiers).

3. Les questionnaires d'enquête peuvent être analysés pour déceler les cas d'éléments supprimés, de répétitions, de changements de classification ainsi que les unités hors du champ d'enquête. Les taux de fréquence suffisent habituellement à déterminer la qualité de la base de sondage. Une analyse plus poussée peut révéler des problèmes particuliers qui touchent la construction même de la base.

4. On peut aussi analyser les données tirées d'échantillons répétés conçus spécialement pour produire le même nombre d'unités. Des écarts importants peuvent être l'indice de problèmes au niveau de la construction de la base.

5. Dans le cas des enquêtes-entreprises, il peut se produire des incohérences au niveau de la couverture en raison des créations, des disparitions ou des changements de classification qui sont survenus depuis le moment où l'échantillon a été choisi. (Tous ces changements ne sont peut-être pas encore reflétés dans la base de sondage mise à jour.) L'ampleur de l'ereur de couverture peut être déterminée par les moyens suivants:

   (i) l'utilisation d'une base aréolaire - un échantillon de secteurs est visité, et toutes les entreprises situées dans ces secteurs sont consignées et comparées avec toutes les entreprises énumérées pour ces secteurs dans la base initiale;

   (ii) la comparaison entre la base mise à jour au moment de la période de référence de l'enquête et la base au moment de l'échantillonnage (cette méthode est moins satisfaisante que (i), mais elle coûte beaucoup moins cher);

   (iii) la révision du profil - la détermination de la structure et la classification des éléments d'un échantillon d'entreprises à partir de la base initiale (cette méthode ne permet pas de déceler les créations);

   (iv) la comparaison avec les listes commerciales (comme dans 4.2.1).

Macroniveau

6. Les chiffres estimés peuvent être comparés avec des chiffres estimés à partir d'une autre source. Les estimations provenant d'une source externe fiable, dans certains cas une version antérieure de la même enquête, peuvent être corrigées pour tenir compte des éléments nouveaux, des éléments supprimés, de la migration et des changements observés dans les populations cibles. Ces estimations corrigées peuvent ensuite être comparées aux estimations faites à partir des données de l'enquête courante. L'écart en pourcentage de chaque estimation peut servir de mesure de la qualité ou de la fiabilité. Toutefois, il ne faut jamais oublier que les

données corrigées de la source externe ne sont pas toujours le reflet exact de la réalité.

7. On peut aussi effectuer des contrôles d'uniformité (par ex. le ratio hommes/femmes du recensement) pour évaluer l'erreur de couverture.

Les analyses décrites aux points 3 et 4 ci-dessus peuvent aussi révéler des problèmes de mise à jour de la base de sondage, de sélection de l'échantillon ou des changements réels dans la population.

4.3 RÉPONSE

OBJET

- Évaluer le degré d'exactitude des réponses.

LIGNES DIRECTRICES

Microniveau

1. On peut produire des distributions de fréquence globales et stratifiées des variables observées, des diagrammes de dispersion ainsi que des totalisations de combinaisons de variables. Une telle analyse se fait à un niveau plus détaillé que l'analyse courante des résultats globaux de l'enquête. Les variables visées par ce genre d'analyse sont les variables de contrôle et les codes indicateurs. On doit porter une attention particulière aux valeurs de codes qu'on s'attendait à obtenir, mais qui n'ont pas été rapportées ou qui ont été rapportées peu de fois ainsi qu'à celles qui présentent des irrégularités évidentes, par exemple dans le cas des codes non valides. Ce genre d'analyse permet de déceler des problèmes qui touchent non seulement la réponse, mais aussi la saisie et l'enregistrement des données.

2. On devrait faire l'analyse des enregistrements échouant aux vérifications, ou qui contiennent des données manquantes. Cette analyse devrait traiter en partie de l'existence de liens possibles entre le rejet d'une vérification donnée (ou l'absence d'un ou plusieurs éléments d'information) et la valeur d'autres éléments d'information ou de variables de classification.

3. On peut comparer les résultats de l'enquête avec des données provenant d'une autre source. Le fichier des données de l'enquête est apparié avec un autre fichier qui comprend des données ou des unités semblables. Même si le nombre des cas de non-appariement peut être élevé, ce genre de comparaison peut dévoiler des biais ou des écarts systématiques qui exigent plus ample analyse.

4. Une réinterview peut être menée auprès d'un échantillon de répondants, qu'on fasse accorder ou non les réponses. Pour cette réinterview, l'interviewer peut utiliser un questionnaire plus détaillé, ou encore le questionnaire initial et chercher à comparer les réponses nouvelles et initiales. Avec une réinterview plus détaillée, on peut obtenir une estimation de la fiabilité de la réponse. Dans une réinterview simple, on doit tenter de recréer le plus possible les conditions dans lesquelles l'enquête originale a été menée, ce qui permet d'obtenir une estimation de la variance de réponse et de certaines de ses composantes.

5. On peut faire une analyse des résultats des études d'évaluation ayant pour objet de mesurer l'erreur de réponse. Ces études, qui visent aussi à déterminer si les

réponses fournies par les répondants sont exactes, doivent faire partie du plan de sondage original et être effectuées le plus rapidement possible après l'enquête proprement dite.

Macroniveau

6. Les résultats devraient être comparés à des macrodonnées provenant d'une autre source ou à des données chronologiques (dans le cas des enquêtes permanentes). On peut aussi recourir à des sources externes qui fournissent des données équivalentes ou de bons indices d'approximation pour les éléments d'information recueillis dans le cadre de l'enquête, pourvu évidemment que le champ d'observation ne soit pas trop différent. En ce qui concerne les données chronologiques, on peut utiliser les estimations qui portent sur des éléments d'information qui ne changent pas beaucoup avec le temps, et il est également possible de mettre à jour les estimations. Même lorsque le champ d'observation n'est pas le même, il est possible de comparer les rapports entre certaines estimations des deux fichiers. Il s'agit de l'un des modes de validation les plus importants.

7. On devrait procéder à des vérifications globales. Les estimations elles-mêmes doivent satisfaire certaines règle internes ou conditions "raisonnables". On doit également enregistrer un très faible niveau de fréquence des caractéristiques jugées rares.

8. Les estimations de la précision devraient être analysées à tous les niveaux. Une précision beaucoup plus grande ou beaucoup plus faible que prévu peut révéler un problème de réponse.

9. Dans le cas des enquêtes permanentes, un contrôle devrait être exercé sur les estimations de la précision. Un changement soudain dans une série d'estimations de la précision peut traduire un problème de réponse exigeant une analyse plus poussée.

4.4 NON-RÉPONSE

OBJET

- Évaluer le degré de non-réponse et ses répercussions sur la précision des estimations.

LIGNES DIRECTRICES

Microniveau

1. Les caractéristiques des non-répondants peuvent être déterminées à partir des données de la base de sondage ou de celles recueillies sur le terrain.

2. Les caractéristiques des non-répondants peuvent être déterminées à l'aide d'un appariement avec des données de source extérieure ou à l'aide d'un suivi.

3. Les tendances de la non-réponse partielle (absence systématique de réponse à un ensemble de questions) et les caractéristiques correspondantes doivent être déterminées.

4.5 SAISIE, CODAGE ET PRÉPARATION DES DOCUMENTS

OBJET

- Évaluer l'incidence des méthodes de saisie et de codage des données et de préparation des documents sur la qualité des données.

LIGNES DIRECTRICES

Microniveau

1. Les résultats des contrôles de la qualité devraient être analysés. C'est là un moyen d'obtenir des renseignements sur les erreurs les plus fréquentes de codage ou d'enregistrement et de déterminer si certains opérateurs ont besoin d'une formation supplémentaire. On doit constituer des dossiers sur les changements observés dans la qualité des données et s'en servir pour améliorer les opérations de saisie.

2. Les données sur la progression du travail devraient aussi être analysées. On peut ainsi se rendre compte si des goulots d'étranglement se forment et exigent que des mesures soient prises. Lorsqu'il se présente de tels goulots d'étranglement, on a souvent tendance à recourir à des expédients qui nuisent à la qualité des données.

3. Des études de recodage peuvent être entreprises. Des codeurs spécialisés procèdent à un nouveau codage d'un échantillon des questionnaires originaux, et les nouveaux codes sont comparés aux codes initialement attribués au cours de l'enquête. On peut donc ensuite estimer l'erreur de codage en pourcentage ainsi que la part de l'erreur de codage dans l'erreur totale.

4.6 ESTIMATION

OBJET

- Évaluer la robustesse et l'efficacité des techniques d'estimation en tenant compte des caractéristiques des données réelles observées. On se fonde très souvent sur des conditions "idéales" pour élaborer les techniques d'estimation; en pratique toutefois, les données accusent des comportements bien différents de ceux que l'on avait prévus et les méthodes d'estimation ne donnent donc pas toujours les résultats escomptés.

LIGNES DIRECTRICES

Microniveau

1. Des études de validation des estimations synthétiques peuvent être entreprises. La technique de l'estimation synthétique peut être utilisée pour produire des estimations à un niveau habituellement plus détaillé que ne le prévoit le plan d'échantillonnage. Une étude de validation peut prendre la forme d'une enquête de portée restreinte dont l'objet est d'estimer directement certains détails requis (par exemple dans le cas des petites régions) et dont les résultats permettent d'évaluer la technique ou le modèle utilisé pour produire les estimations synthétiques.

Macroniveau

2. On peut procéder à une nouvelle estimation en appliquant des techniques différentes, notamment en ce qui touche le traitement des données aberrantes, l'imputation et l'estimation de la variance. Si les nouvelles estimations obtenues enregistrent un écart trop important par rapport aux estimations originales, il sera sans doute nécessaire de pousser plus loin l'analyse. Idéalement, une étude de ce genre doit être entreprise au cours de l'étape de planification au moment où l'on met au point les techniques d'estimation. En l'absence de données chronologiques ou de données de tests pertinentes, on pourra utiliser des données synthétiques plausibles construites à partir d'un modèle convenable, mais les résultats devront alors être corroborés une fois que les données réelles seront disponibles.

3. On doit entreprendre une analyse du processus d'imputation. L'ampleur de cette analyse dépend de la méthode d'imputation utilisée. Quand l'imputation est fondée sur l'utilisation de tables de décision, on doit procéder à une répartition de la fréquence d'utilisation de chaque ensemble de possibilités. Quand l'imputation est fondée sur l'utilisation de "donneurs" (méthode du "hot deck" et du plus proche voisin), on doit recueillir des statistiques sur la fréquence d'utilisation de chaque donneur, sur le nombre de tentatives de completer l'imputation soumises aux vérifications avant que les vérifications ne se révèlent satisfaisantes, sur le nombre de cas de rejet à la vérification (et que l'on peut vraisemblablement attribuer à une irrégularité de traitement ou à une imputation manuelle) et sur les distributions initiales et finales des variables.

4. On devrait analyser les effets de l'imputation. Cela peut se faire à l'aide de modèles de simulation ou par un appariement avec un fichier de données externe dont les variables sont sensiblement les mêmes ou qui offrent de bonnes variables de substitution.

4.7 PLAN D'ÉCHANTILLONNAGE

OBJET

- Suivre l'efficacité du plan d'échantillonnage. À mesure que la population ou que la base de sondage évolue, la qualité du plan d'échantillonnage se détériore. Pour une enquête permanente, la baisse de la qualité du plan d'échantillonnage révèle la nécessité de procéder à une révision.

LIGNES DIRECTRICES

Macroniveau

1. L'effet du plan d'échantillonnage sur diverses estimations devrait être comparé aux effets obtenus au début de l'enquête.

2. Des plans d'échantillonnage différents peuvent être simulés (par ex. des méthodes différentes de construction des unités primaires d'échantillonnage, des méthodes d'échantillonnage différentes et d'autres méthodes de stratification) et on peut comparer les effets de ces plans et ceux du plan actuel.

Un grand nombre des lignes directrices énoncées dans la présente partie sont présentées plus en détail dans Gosselin et coll., 1978.

4.8 MODÈLES

OBJET

- Assurer que le modèle utilisé est approprié ou encore valide.

LIGNES DIRECTRICES

Comme le sens du mot "modèle" diffère selon les disciplines, il est difficile de fournir des lignes directrices concrètes concernant la validation des modèles. Nous posons ici comme hypothèse qu'un modèle est "quelque chose dont le comportement se rapproche de la réalité", mais qui présente une structure beaucoup plus simple (notamment parce que certaines sources de variation sont éliminées).

Selon le contexte, les mesures suivantes peuvent être prises:

1. Vérification des hypothèses sur lesquelles s'appuie le modèle.

2. Contrôle de la capacité de prédiction du modèle grâce à l'utilisation de données chronologiques en vue de la "prédiction" de valeurs connues.

3. Comparaison des résultats produits par le modèle avec ceux d'autres sources d'information connexes.

4. Recours à des études de sélection et de validation croisée (voir plus loin la section 5.2.13).

5. Vérification de la sensibilité du modèle aux paramètres utilisés.

6. Validation des données d'entrée du modèle.

BIBLIOGRAPHIE

BISHOP, Y.M.M., FIENBERG, S.E. AND HOLLAND, P.W., *Discrete Multivariate Analysis: Theory and Practice*, (Cambridge, Mass., M.I.T. Press, 1974).

DEPT. OF CONSUMER AND CORPORATE AFFAIRS, *Concentration in the Manufacturing Industries of Canada*, (Ottawa, Information Canada, 1971).

GOSSELIN, J.F., CHINNAPPA, B.N., GHANGURDE, P.D., ET TOURIGNY, J., *Répertoire des méthodes d'évaluation des erreurs dans les recensements et les enquêtes*, (Ottawa, Statistique Canada, 1978), no 13-564F au catalogue.

LESSLER, J.T., FOLSOM, R.E. AND KALSBEEK, W.D., *A Taxonomy of Error Sources and Error Measures for Surveys*, Project RTI/1791/00-03F. (Research Triangle Institue, Research Triangle Park, North Carolina, 1983).

U.S. DEPT. OF COMMERCE, Bureau of the Census *1977 Census of Manufactures* (Washington, D.C., U.S. Government Printing Office.

U.S. DEPT. OF COMMERCE, Bureau of the Census, *1977 Census of Manufactures*: Concentration Ratios in Manufacturing, (Washington D.C., U.S. Government Printing Office, May 1981).

5. *ANALYSE*

INTRODUCTION

L'analyse, qui consiste à résumer et à interpréter les données, comporte une étude plus poussée de l'information que celle qui est offerte par les estimations courantes d'enquête. Elle est importante pour la réinjection d'informations découlant d'enquêtes précédentes et d'études pilotes dans le processus de planification des nouvelles activités statistiques, la fourniture d'informations essentielles à la conception des enquêtes, la formulation d'objectifs réalistes concernant la qualité, la découverte des problèmes et les exigences du traitement des données. Il est souvent possible d'utiliser également à cette fin des fichiers indépendants de données connexes ou auxiliaires.

Le processus de validation exige également de l'analyse, notamment pour expliquer les différences entre les résultats de l'activité et les données connexes.

L'analyse peut également englober l'exploration d'aspects économiques et sociaux par l'examen de données provenant d'une ou de plusieurs sources.

LIGNES DIRECTRICES CONCERNANT LA QUALITÉ

OBJET

- Assurer ou améliorer la qualité de l'analyse.

LIGNES DIRECTRICES

Les présentes lignes directrices ne visent pas à couvrir l'ensemble des méthodes statistiques utilisées pour résumer une information, mais plutôt à mettre l'accent sur les principales techniques dont on doit évaluer la capacité à analyser tout ensemble de données. Ces techniques sont présentées dans les paragraphes qui suivent; on y explique aussi brièvement les raisons pour lesquelles ces techniques sont jugées utiles.

5.1 PRÉPARATION DES DONNÉES

1. **On doit faire une étude de la documentation existante sur les notions et définitions de base, les méthodes de collecte des données, le plan d'échantillonnage, la qualité des données, etc.** Pour pouvoir interpréter et analyser correctement les données qui lui sont soumises, l'analyste doit être parfaitement au courant des principes et des méthodes qui ont été appliqués en vue de leur production. Il arrive souvent que des ensembles de données distincts, qui portent sur la même caractéristique, ont trait en fait à des variables qui sont désignées de façon semblable mais reposent sur une définition différente. Si l'analyse ne tient pas compte de la méthode complexe de collecte des données qui a été utilisée, sa validité peut être sérieusement compromise. Les données ayant déjà été traitées peuvent ne pas convenir au genre d'analyse que l'on souhaite faire (par exemple, dans le cas d'une analyse d'une série chronologique, c'est l'objet de l'analyse qui détermine s'il est préférable d'utiliser des données brutes ou des données désaisonnalisées).

2. **On doit procéder à une étude de la documentation existante sur le fichier des données et vérifier la signification de tous les codes et symboles indicateurs.** De façon générale, on trouve dans un dossier une grande quantité de renseignements non statistiques. Voici quelques exemples de renseignements qu'on peut y trouver : le

code-réponse de l'unité visée, la date à laquelle on a reçu la réponse (s'il y a lieu), les éléments d'information manquants, les éléments d'information ayant été acceptés à la vérification et ceux ayant fait l'objet d'une imputation, le poids d'échantillonnage attribué, etc. Les codes particuliers de certains éléments de données peuvent comprendre des valeurs spéciales à attribuer à des données manquantes. S'il s'agit de valeurs numériques, il est possible qu'aucune distinction ne soit faite entre un "0" attribué en tant que valeur nulle fondée et un "0" attribué pour une donnée manquante. Rien ne garantit que tous les dossiers sont également fiables; il importe donc de déterminer à l'avance quels dossiers peuvent servir à une analyse donnée.

3. **On doit entrer en communication avec le personnel responsable des activités de planification, de conception et de collecte des données.** Malheureusement, la documentation est souvent incomplète ou périmée. Dans certains cas, les procédures ont été modifiées au pied levé ou encore mal appliquées à la suite d'erreurs humaines. Il importe que l'analyste se tienne le plus possible au courant de l'existence de ce genre d'information verbale et qu'il sache pourquoi certaines tâches ont été exécutées d'une certaine façon quand, en rétrospective, il aurait été plus logique d'agir autrement.

4. Si le fichier contient des données imputées, **on doit faire un examen des méthodes d'imputation utilisées** en vue de déterminer si ces données conviennent à l'analyse. Habituellement, l'imputation est faite en fonction des besoins immédiats de l'enquête. L'imputation des données manquantes ou inacceptables permet d'assurer la simplicité relative des méthodes d'estimation et de produire des estimations raisonnables. Certaines méthodes d'imputation peuvent toutefois introduire un biais dans l'analyse, particulièrement si le nombre de cas d'imputation est très élevé. Ainsi, lorsque l'imputation est fondée sur l'utilisation d'un modèle linéaire, toute l'analyse subséquente des données fera ressortir le même modèle linéaire, et la relation semblera probablement beaucoup plus étroite. (Voir Sande, 1982).

5. **On doit éliminer du fichier de travail analytique tous les dossiers qui ne conviennent pas à l'analyse.** Pour les besoins de l'analyse, les données du fichier principal sont trop nombreuses et classées de manière trop générale. Il importe également de n'oublier aucun des dossiers que l'on prévoit utiliser pour l'analyse. Il est donc habituellement plus simple d'établir un fichier de travail (peut-être même plusieurs) contenant toute l'information pertinente ainsi que tous les renseignements qui en découlent (par exemple les coefficients de pondération ajustés). Évidemment, ce genre de fichier doit être bien documenté.

5.2 DISTRIBUTION, CORRÉLATION, MODÉLISATION, RÉDUCTION

1. **Les statistiques descriptives telles que les quantiles et les histogrammes.** Même lorsqu'il a l'intention d'appliquer une technique complexe, l'analyste ne doit pas négliger ces moyens simples de décrire les données, descriptions qui n'obligent pas à formuler des hypothèses au sujet des modèles de base et dont l'interprétation est facile.

2. **Les analyses d'exploration des données** (voir Mosteller et Tukey, 1977). Lorsqu'un analyste ne sait pas trop dans quelle direction s'orienter ou quelles hypothèses semblent les plus réalistes, ce genre de technique lui permet d'avoir une vue d'ensemble de la situation.

3. **La représentation graphique des distributions de probabilité faite en vue d'évaluer si les distributions théoriques conviennent à l'ajustement des données.** Malgré la possibilité d'appliquer certains tests théoriques aux distributions (par exemple le test de Wilk-Shapiro, le test de Kolmogorov-Smirnov), ce genre de graphique (qui fait correspondre les valeurs prises par une série de données aux quantiles prévus selon l'hypothèse de distribution que l'on vérifie) a l'avantage de montrer la qualité de l'ajustement et, s'il y a lieu, de faire ressortir les problèmes. Un mauvais ajustement peut être dû à une seule valeur extrême, à la faiblesse de l'hypothèse ou à un effet de la méthode de collecte des données. (Voir Wilk et Gnanadesikan, 1968; Shapiro et Wilk, 1965).

4. **Les graphiques et les diagrammes.** Les points 1, 2 et 3 ci-dessus montrent qu'on n'insistera jamais assez sur l'importance des techniques de représentation graphique dans l'analyse des données et leur présentation, sans qu'il soit nécessaire de produire des graphiques d'une grande complexité. (Voir Tufte, 1982; Chambers et coll., 1983).

5. **L'utilisation de techniques robustes pour l'estimation des paramètres.** Les méthodes d'estimation courantes peuvent produire de piètres estimateurs lorsqu'on constate la présence de données aberrantes ou d'anomalies. Les données aberrantes ne peuvent pas toujours être découvertes facilement, soit parce que l'ensemble des données est trop vaste ou parce que l'observation porte sur un trop grand nombre de variables. Cependant, la plupart des systèmes d'estimation automatisés que l'on utilise aujourd'hui peuvent repérer et prendre en charge les données aberrantes sans qu'une opération manuelle soit nécessaire.

6. **Les diagnostics de régression.** Lorsqu'ils sont utilisés sans discernement, les programmes de régression peuvent fausser les estimations. Les diagnostics de régression font ressortir les observations qui ont l'influence la plus grande sur la régression. (Voir Belsey et coll., 1980; Hocking, 1983).

7. **La prise en compte de la complexité du plan d'échantillonnage.** Quand on utilise un plan d'échantillonnage complexe, on attribue un poids différent à chaque observation, et on rend les observations dépendantes les unes des autres. La plupart des programmes d'analyse statistique posent comme hypothèse l'indépendance des observations et l'attribution d'un poids égal à chacune; ces hypothèses sont donc contredites dans le cas d'un échantillon complexe. En particulier, il n'est pas possible d'appliquer les techniques habituelles d'analyse des tables de contingence. Dans le cas de l'analyse de régression, les techniques habituelles suffisent pour estimer les coefficients, mais non pour évaluer leur importance. Le problème peut être résolu par l'application de techniques de répétition (par exemple la méthode dite du jack-knife) ou l'utilisation d'autres méthodes spéciales. (Voir Scott et Holt, 1982; Rao et Scott; 1982; Shah, 1978; Fellegi, 1980).

8. **L'évaluation de la validité de l'ajustement.** Quand un modèle est ajusté, soit en vue d'"expliquer" la variation soit à des fins de prévision, les tests de signification courants ne suffisent pas pour confirmer ou rejeter le modèle et l'analyste doit avoir recours à d'autres techniques, tant graphiques que théoriques, pour juger de la qualité de l'ajustement (voir les points 2, 3, 6 et 9 à 13). L'analyste ne doit pas oublier de vérifier la vraisemblance de son modèle ajusté, par exemple une droite de régression ajustée doit-elle passer par l'origine? Si les résultats sont invraisemblables, il se peut que cela provienne de données imparfaites.

9. **L'analyse des résidus** par des moyens tels que les diagrammes de dispersion des résidus, les graphiques des résidus en fonction des variables dépendantes et

indépendantes, les diagrammes de dispersion des résidus sur papier à échelle gaussienne et les graphiques des résidus partiels. Ce genre d'analyse peut révéler des lacunes dans le modèle, contredire l'hypothèse selon laquelle les erreurs obéissent à une loi normale ou indiquer des problèmes dans les données (voir les points 2, 3, 4 et 8). (Voir Mosteller et Tukey,1977; Hocking, 1983).

10. **La détection des données aberrantes.** Les données aberrantes non repérées peuvent nuire à l'exactitude des estimations (voir les points 2, 3, 4, 5, 8 et 9). La détection des données aberrantes peut, en soi, entraîner une reformulation du modèle de façon que les données aberrantes puissent être prises en compte par certaines variables indépendantes non encore utilisées, ou elle peut révéler des problèmes qui nécessitent une étude plus approfondie.

11. **La transformation des variables tant dépendantes qu'indépendantes.** On a souvent recours à des transformations pour accroître la validité de l'ajustement d'un modèle, démontrer un lien plus étroit entre les variables et stabiliser la variance d'une variable indépendante. Si une transformation améliore de beaucoup la qualité de l'ajustement, on peut souvent se faire une meilleure idée de la structure de base sur laquelle s'appuie le modèle. (Voir Mosteller et Tukey, 1977; Hocking, 1983).

12. **Le recours à la régression non linéaire dans l'ajustement des modèles.** Un modèle non linéaire peut offrir une explication plus rationnelle au sujet des rapports qui existent entre les variables, et les progiciels statistiques modernes sont conçus de façon à faciliter leur traitement.

13. **Les études de sélection préliminaire et de validation croisée.** Les déductions tirées d'une analyse ne sont peut-être pas aussi générales qu'on le souhaiterait. Grâce à ce genre de technique, les données sont réparties en deux catégories et on se sert des données de la deuxième catégorie pour confirmer l'analyse qui a été faite des données de la première catégorie. (Voir Mosteller et Tukey, 1977.)

14. **La consultation d'un spécialisteen analyse statistique de données.** Ce genre de consultation peut être utile à toute personne spécialisée dans un domaine bien précis en lui permettant d'élargir ses connais-sances des diverses techniques d'analyse des données, d'être avertie des problèmes qui peuvent se poser dans le cadre de leur application et même de se voir recommander des solutions pour ces problèmes.

15. **Évaluation par les pairs.** L'échange d'idées et la discussion de son travail avec des collègues donneront souvent lieu à des suggestions constructives sur le plan des techniques et des genres d'enquêtes.

5.3 ANALYSE LONGITUDINALE

De façon générale, les commentaires formulés dans les parties 5.1 et 5.2 s'appliquent à l'analyse longitudinale qui comprend un facteur temps. On désigne parfois ce genre d'analyse une "analyse de séries chronologiques", quoiqu'on réserve habituellement cette expression pour décrire un ensemble plus structuré et complexe de méthodes ayant pour objet d'expliquer et de prévoir les fluctuations dans le temps à l'aide de techniques de distribution de fréquence et (ou) d'autorégression, lesquelles reposent sur un ensemble relativement figé d'hypothèses et exigent normalement un gros volume de données (c'est-à-dire une série longue et continue de données).

Voici quelques lignes directrices à suivre si l'on prévoit procéder à une analyse longitudinale des résultats d'une enquête permanente ou d'une activité de collecte de données répétée autre qu'un recensement.

1. Dans le cas d'une enquête permanente, un échantillon de données longitudinales doit être construit. Cela a pour effet de diminuer la variance des paramètres qui mesurent le changement. Même lorsque l'analyse longitudinale n'est pas considérée comme un des objectifs principaux, on doit la plupart du temps faire certaines estimations du changement; à cette fin, le plan d'échantillonnage prévoit habituellement le renouvellement d'une partie de l'échantillon.

Quand une population est observée à intervalles réguliers (par exemple dans le cas de l'EPA et de l'EERH), l'analyse longitudinale peut aider à évaluer la qualité des données, car elle permet de mieux comprendre le comportement de la population et d'orienter les décisions qui doivent être prises au sujet du questionnaire, de la vérification, de l'imputation et des totalisations. Aussi, l'analyse doit-elle tenir compte des points suivants:

2. La stabilité de la population dans le temps, par exemple en ce qui a trait à la répartition par âge, secteur d'activité, etc.

3. Les taux de changements de classification (migrations), d'unités nouvelles (naissances) et d'unités supprimées (décès) dans la population visée.

4. L'évolution du comportement de chaque unité.

Quand l'analyse longitudinale s'inscrit dans les objectifs, l'assurance de la qualité vise à améliorer l'analyse grâce à des mesures comme:

5. Les représentations graphiques des données dans le temps.

6. L'élimination des perturbations dans les séries par la méthode de lissage.

7. L'utilisation de méthodes robustes.

5.4 ANALYSE DES SÉRIES CHRONOLOGIQUES

1. Si l'on envisage de procéder à une analyse structurée des séries chronologiques, il convient de consulter un spécialiste et de s'assurer que les données conviennent à cette fin. (Voir D.G. Horovitz, 1978).

   Un grand nombre des enquêtes transversales qui sont menées périodiquement par Statistique Canada entraînent la production de données relatives à des périodes consécutives, qu'on peut généralement considérer comme une série chronologique d'observations dont on se sert pour mesurer l'évolution dans le temps de phénomènes sociaux ou économiques, qu'une telle utilisation ait ou non été prévue au moment de la conception de ces enquêtes. De fait, quand l'analyse en bonne et due forme d'une série chronologique est un des objectifs d'une enquête, il importe d'en tenir compte dès l'étape de la conception afin que le plan d'échantillonnage ou les méthodes d'estimation utilisés ne confondent pas les effets, par exemple les effets saisonniers, que l'analyse des séries chronologiques est censée évaluer.

2. Un grand nombre des séries chronologiques de Statistique Canada sont publiées à la fois sous leur forme originale et après désaisonnalisation ou décomposition (en tendances, cycles, fluctuations saisonnières et mouvements irréguliers). Étant donné que le processus suppose l'utilisation d'un modèle statistique d'estimation, une autre source d'erreur est introduite dans les données finales. La désaisonnalisation des données joue un rôle important dans l'analyse de la situation économique,

analyse dont l'efficacité dépend de la mesure dans laquelle les données révèlent les caractéristiques et les points de retournement du cycle économique ou encore le stade du cycle qui a été atteint. Les estimations du comportement saisonnier des indicateurs économiques et sociaux en tant que telle, si elle est fiable, a une importance déterminante dans l'élaboration des politiques. Il est donc important que l'utilisateur connaisse le degré de fiabilité des données désaisonnalisées. (voir E.B. Dagum, 1974).

3. Voici quelles sont les lignes directrices à suivre pour assurer la qualité de la désaisonnalisation ou de la décomposition des séries chronologiques[1]:

   - On doit d'abord procéder à un examen des séries chronologiques en vue de déterminer s'il y a une saisonnalité dans les données et, dans l'affirmative, si celle-ci est additive ou multiplicative. (Voir Lothian et Morry, 1978a);

   - S'il y a saisonnalité et que l'on décide d'isoler les variations saisonnières, un degré élevé de précision doit être associé aux techniques de repérage et d'estimation du modèle qui sera utilisé. On doit apporter un soin particulier au choix de la méthode de décomposition du modèle (additive, multiplicative ou additive logarithmique) ainsi qu'à la sélection des systèmes de pondération utilisés en vue d'estimer les composantes saisonnières stables ou mobiles ainsi que les tendances associées aux changements structurels, pour repérer et remplacer les données aberrantes, et ainsi de suite. (Voir E.B. Dagum (1983));

   - Des critères de qualité doivent être choisis. Il existe divers critères qui permettent d'évaluer la qualité des séries désaisonnalisées; on peut viser par exemple à produire des séries désaisonnalisées bien lissées, à éliminer la saisonnalité résiduelle dans les séries, à réduire au minimum l'ampleur des révisions, et ainsi de suite;

   - Des mesures de contrôle de la qualité devraient être produites et examinées et servir à évaluer la qualité des séries chronologiques après leur désaisonnalisation ou leur décomposition. (Voir Lothian et Morry (1978b));

   - On devrait procéder à un examen périodique de la méthode de désaisonnalisation choisie ou utilisée étant donné que les variations saisonnières liées aux phénomènes observés peuvent changer avec le temps.

BIBLIOGRAPHIE

BELSLEY, D.A., KUH, E., ET WELSCH, R.E., Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, (New York, Wiley, 1980).

BINDER, D.A., "On the Variances of Asymptotically Normal Estimators from Complex Surveys", *International Statistical Review*, LI (1983), pp. 279-292.

BINDER, D.A., GRATTON, M., HIDIROGLOU, M.A., KUMAR, S., AND RAO, J.N.K., "Analysis of Categorical Data from Surveys with Complex Design". Paper presented at the Seminar on recent Developments in the Analysis of Large Scale Data Sets, sponsored by the Statistical Orfice of European Communities, Luxemburg, (1983).

CHAMBERS, J.M., CLEVELAND, W.S., KLEINER, B., ET TUKEY, P.A., *Graphical Methods for Data Analysis*, (Boston, Duxbury Press, 1983).

DAGUM, E.B., *La méthode de désaisonnalisation X-II-ARMMI*, (Ottawa, Statistique Canada, 1980) no 12-564F au catalogue.

DAGUM, E.B., "A note on the seasonal adjustment of economic time series", *Revue statistique du Canada*, août 1974.

FELLEGI, I.P., "Approximate Tests of Independence and Goodness of Fit Based on Stratified Multistage Samples", *Journal of the American Statistical Association*, LXXV (1980), pp. 261-268.

GOSSELIN, J.F., CHINNAPPA, B.N., GHANGURDE, P.D., ET TOURIGNY, J., *Répertoire de méthodes d'évaluation des erreurs dans les recensements et les enquêtes*, (Ottawa, Statistique Canada, 1978) no 13-564F au catalogue.

HIDIROGLOU, M., AND RAO, J.N.K., "Chisquare Tests for the Analysis of Categorical Data from the Canadian Health survey". Paper presented at the 43rd ISI Conference, Buenos Aires (1981).

HOCKING, R.R., "Developments in Linear Regression Methodology: 1959-1982", *Technometrics* XXV (1983), pp. 263-270.

HORVITZ, D.G., "Some Design Issues in Sample Surveys". Dans *Survey Sampling and Measurement*, (Ed. N. Krishnan Namboodiri), (New York, Academic Press, 1978) pp. 3-11.

LOTHIAN, J., ET MORRY, M., "A test for the presence of identifiable seasonality when using the X-II Program". Document de recherche de la Division des séries chronologiques - Recherche et analyse, Statistique Canada, 1978.

LOTHIAN, J., AND MORRY, M., "A set of quality control statistics for the X-II-ARIMA seasonal adjustment program". Document de recherche de la Division des séries chronologiques - Recherche et analyse, Statistique Canada, 1978.

MOSTELLER, F., ET TUKEY, J.W., *Data Analysis and Regression*, (Addison-Wesley, 1977).

RAO, J.N.K., ET SCOTT, A.J., "The Analysis of Categorical data from Complex Sample Surveys: Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables", *Journal of the American Statistical Association*, vol. LXXVI (1982), pp. 221-230.

SANDE, I.G., "Imputation in Surveys: Coping with Reality". *The American Statistician*, vol. XXXVI (1982), pp. 145-152.

SCHMIDT, C.F., *Statistical Graphics. Design, Principles and Practices*, (Wiley-Interscience, 1983).

SCOTT, A.J., ET HOLT, D., "The Effect of Two-Stage Sampling on Ordinary Least Squares Methods", *Journal of the American Statistical Association*, LXXVII (1982), pp. 848-854.

SHAH, B.V., "Variance Estimates for Complex Statistics from Multistage Sample Surveys". Dans *Survey Sampling and Measurement* (Ed. N. Krishnan Namboodiri), (New York, Academic Press, 1978).

SHAPIRO, S.S., ET WILK, M.B., "An Analysis of Variance Test for Normality (Complete Samples)", *Biometrika*, vol. LII (1965), pp. 591-611.

TUFTE, G.R., *The Visual Display of Quantitative Information*, (Cheshire, Connecticut, Graphic Press, 1982).

WILK, M.B., ET GNANADESIKAN, R., "Probability Plotting Methods for the Analysis of Data", *Biometrika*, vol. LV (1968), pp. 1-18.

6. *DOCUMENTATION*

INTRODUCTION

La documentation est l'établissement d'un dossier sur un projet. À partir de la documentation, il devrait idéalement être possible de reconstituer l'évolution des travaux et les décisions liées au projet. Il importe non seulement de connaître les décisions qui ont été prises, mais aussi de savoir pourquoi elles ont été prises. De la même façon, les théories qui sont à la base des méthodes utilisées sont tout aussi importantes que les applications de ces méthodes. Quel que soit l'aspect du projet visé, la documentation peut être établie à l'intention d'un ou de plusieurs groupes différents, par exemple la direction, le personnel spécialisé, les responsables de la planification d'autres enquêtes, les utilisateurs de l'extérieur (à des niveaux de complexité différents), et ainsi de suite.

LIGNES DIRECTRICES CONCERNANT LA QUALITÉ

OBJET

- Établir une documentation complète et facile à comprendre sur le projet.

LIGNES DIRECTRICES

On trouvera ci-dessous une liste des principales questions dont on doit tenir compte dans le cadre de la documentation des différents aspects de l'activité statistique. (La documentation relative aux étapes de la planification et de la diffusion est également décrite au chapitre 1, et aux sections 7.3 et 7.4).

Pour des considérations d'ordre pratique (volume), il n'est pas toujours possible de réunir toute la documentation qui est proposée ici. Il appartient au chef de l'activité statistique visée de décider quels documents sont essentiels et lesquels peuvent être omis sans trop nuire aux exigences touchant l'information et la tenue de dossiers. Il importe également de souligner que toute documentation, qu'elle soit complète ou non, sera plus utile si elle est à jour, bien ordonnée et facile à consulter (voir section 6.11) et également bien rédigée et techniquement fiable (voir section 6.12).

EXHAUSTIVITÉ

6.1 OBJECTIFS ET NOTIONS DE BASE

(i) Les objectifs quant aux renseignements requis au sujet du champ d'observation, du niveau de précision, du délai de production des données, de la fréquence de diffusion, de la précision des estimations et de la qualité des données.

(ii) Les notions et définitions de base, les normes et les rapports que l'activité présente avec d'autres enquêtes ou projets et d'autres domaines d'étude connexes.

(iii) L'évolution des objectifs dans le temps, documentation qui est habituellement présentée sous forme de compte rendu ou de mise à jour de l'information. Les objectifs changent souvent au cours de la mise en oeuvre du projet, en raison de l'existence de contraintes budgétaires, à la suite des résultats d'études de faisabilité ou de tests pilotes, parce que de nouvelles techniques sont découvertes et pour bien d'autres raisons. Il est important de faire état de tels changements et d'en expliquer les raisons par écrit.

(iv) L'élaboration du questionnaire et tout particulièrement une analyse de la mesure dans laquelle cet instrument tient compte des définitions et notions de base adoptées. La documentation peut prendre la forme d'un test pilote du questionnaire et d'une analyse subséquente des résultats du test.

6.2 MÉTHODOLOGIE

(i) **Une étude des différents plans de sondage possibles.** Des recommandations peuvent être formulées dans le cadre de cette étude, mais on ne doit pas chercher à prévoir ce que sera la décision finale. Il est en effet très important de faire un examen et une analyse de toutes les possibilités qui s'offrent afin de pouvoir prendre une décision finale éclairée. On doit aussi justifier les raisons d'un tel choix.

(ii) **Des études sur le choix des méthodes** (stratification, estimation, etc.), y compris des tests pilotes.

(iii) **Des documents sur:**

Le plan d'échantillonnage L'estimation Le traitement des données aberrantes La vérification L'imputation	**pour**	Les responsables des méthodes (mathématiciens, statisticiens) Les responsables du contenu et de l'analyse Le personnel du service informatique La haute direction.

Ces documents, qui doivent être révisés au besoin, sont importants pour la diffusion d'information interne, particulièrement à l'intention des nouveaux participants au projet à tous les niveaux. Naturellement, l'ampleur et la complexité du projet détermineront le degré dans lequel les questions pourront être combinées pour un groupe cible donné. Selon le niveau et la spécialisation des personnes auxquelles ils s'adressent, les documents doivent mettre l'accent sur les différents aspects plus ou moins précis des méthodes visées et doivent doser en conséquence l'utilisation de termes techniques.

(iv) **Un document de synthèse à l'intention des spécialistes sur tous les aspects techniques** des documents mentionnés ci-dessus. Présenté comme un tout cohérent, il doit pouvoir servir de document de référence final.

(v) **Un aperçu général des méthodes utilisées,** qui doit être inclus dans la publication où l'on présente les résultats de l'enquête.

(vi) **Des articles destinés à être publiés dans les revues spécialisées.** La meilleure façon de s'assurer de la qualité de tels documents consiste à les soumettre à l'examen minutieux de ses homologues spécialistes en vue d'obtenir leur approbation. Cette façon de procéder a en outre l'avantage de faire profiter de l'expérience de spécialistes ceux qui ont à faire face à des problèmes semblables.

6.3 TESTS

(i) Une documentation sur les tests **sur le terrain** ou les **enquêtes pilotes** réalisés dans le cadre du projet. Chaque test est un mini-projet en soi; les objectifs sont toutefois limités. La planification et l'exécution de ces activités exigent donc autant d'attention que le projet proprement dit.

(ii) Des études sur les méthodes de mise à l'essai des **systèmes** et des rapports sur les résultats. Un système est conçu de façon à répondre à certaines exigences précises. Pour la mise à l'essai d'un système, on soumet des données d'entrée sous une forme que le système peut traiter et on obtient des données de sortie conformes à celles que le système est censé produire. Les données des tests doivent permettre de découvrir tous les cas imprévus. On doit faire un compte rendu détaillé des résultats des tests en vue de déterminer dans quelle mesure les exigences ont été respectées. (Quand un test révèle que le système ne répond pas aux normes établies, celui-ci est habituellement modifié jusqu'à ce que les résultats des tests soient concluants.)

6.4 SYSTÈMES

(i) Une documentation sur les fichiers de données - explications au sujet de la présentation, des codes, des fréquences de base et des méthodes de vérification. Ces documents sont essentiels si l'on veut que les fichiers soient faciles à comprendre et à utiliser.

(ii) Une documentation sur les systèmes - leur construction, les algorithmes et leurs applications. Une telle information est essentielle si on veut modifier toute partie d'un système.

(iii) Les révisions. Toute révision, qu'elle qu'en soit l'ampleur, doit être documentée et cette documentation doit être intégrée à celle déjà établie sur le système. Il n'est pas suffisant, même si cela vaut mieux que rien, de rédiger une série de notes de services pour apporter des corrections et de les annexer à la documentation principale sur le système.

(iv) Les rapports de contrôle - c'est-à-dire la tenue de dossiers pour des aspects aussi variés que le temps consacré aux différentes opérations, les parties du système qui posent le plus de problèmes, etc.

(v) L'ordonnancement. Dans un système complexe, la synchronisation des différentes composantes est essentielle si l'on veut que les résultats soient publiés en temps opportun. Il importe de déterminer à l'avance les dates des différentes phases d'exécution afin que l'on puisse se rendre rapidement compte de tout retard de traitement.

6.5 CONTRÔLE DE LA QUALITÉ

(i) L'analyse des possibilités qui s'offrent en vue du choix d'une méthode de contrôle de la qualité. Quelle que soit l'opération visée, le choix d'une méthode en particulier n'est pas simple et il importe d'analyser tous les facteurs dont on a dû tenir compte avant de prendre une décision finale à ce sujet.

(ii) La marche à suivre. On doit établir un ensemble de lignes de conduite, ou rédiger un manuel, à l'intention des surveillants et des responsables de la vérification.

(iii) Les rapports. On doit établir des rapports périodiques sur les résultats des méthodes de contrôle de la qualité et le rendement des différents opérateurs en vue de pouvoir évaluer la qualité et déterminer si certains opérateurs ont besoin d'une formation plus poussée.

6.6 OPÉRATIONS

(i) Des guides de formation doivent être produits en vue d'assurer l'exhaustivité et l'uniformité du travail de formation et de servir de documents de référence aux personnes responsables de la planification des modifications à apporter.

(ii) Des guides à l'intention des opérateurs et des interviewers. Ces documents doivent faire état de l'expérience acquise par les opérateurs ou les interviewers et être mis à jour pour tenir compte des changements. Il s'agit d'un outil de référence très important pour les personnes chargées de planifier les opérations et de relever les cas problèmes.

(iii) Des rapports sur le rendement des opérateurs ou des interviewers. Ces rapports intéressent les responsables d'enquête actuels et futurs en vue de la modification des méthodes de travail ou des guides de formation. Ils peuvent aussi aider à déterminer quels opérateurs ou interviewers ont besoin d'une nouvelle formation ou devraient être retirés du projet.

(iv) Les rapports et comptes rendus sur les opérations. Ces documents sont nécessaires en vue de la modification des méthodes de travail, des guides ou des questionnaires.

6.7 MISE EN OEUVRE

(i) **Les procédures.** Les procédures relatives à la mise en oeuvre se situent à un niveau plus global que celles qui ont trait aux opérations. Néanmoins, les diverses opérations doivent être ordonnées, les intrants et les extrants doivent être clairement définis ainsi que la marche à suivre dans chaque cas particulier. De tels détails sont essentiels si l'on veut assurer la bonne marche d'une activité d'envergure, surtout quand on sait que le temps est une ressource rare et que l'absence de documentation sur les procédures de mise en oeuvre peut entraîner des "erreurs" comme l'omission d'une étape importante, l'utilisation d'un mauvais fichier, et ainsi de suite.

(ii) **L'ordonnancement.** Le temps dont on dispose pour exécuter chacune des étapes de la mise en oeuvre est souvent limité, d'où l'importance de l'établissement d'un échéancier précis. Il en va de même pour la répartition des différentes tâches entre le personnel.

(iii) **Des rapports sur les cas problèmes.** Malgré tout le soin apporté à la planification, il est certain que des problèmes se produiront. L'analyse de ce genre de rapports permet aux gestionnaires de déceler les causes de problèmes et de déterminer si des modifica tions doivent être apportées aux opérations, aux systèmes ou aux méthodes, et selon quel ordre de priorité. Des ajustements temporaires peuvent parfois être apportés pour pallier au plus urgent jusqu'à ce qu'une révision en profondeur puisse être mise en oeuvre, mais le chef de l'activité statistique visé doit faire en sorte que ces mesures constituent une exception et non la règle.

(iv) **Des rapports sur l'avancement du travail.** Ces documents montrent si l'ordonnancement est adéquat et si les ressources disponibles sont suffisantes ou s'il se forme des goulots d'étranglement. Pour tout genre d'activité, ces rapports doivent être produits sur une base assez fréquente, au moins hebdomadaire et quelquefois quotidienne dans les périodes de pointe.

6.8 VALIDATION

On doit produire **des rapports sur toutes les études de validation.** Toute étude ou analyse qui est entreprise à des fins de validation n'a aucune utilité s'il n'y a pas de rapport écrit sur les résultats. Un compte rendu verbal ne suffit pas. Même une étude qui doit être abandonnée en raison de complications particulières ou de l'impossibilité d'y donner suite doit faire l'objet d'une certaine description, ne serait-ce que pour indiquer qu'une méthode quelconque a été essayée et expliquer les raisons de l'abandon. De cette façon, les chercheurs futurs essaieront quelque chose d'autre ou s'y prendront différemment. On peut toujours tirer des renseignements utiles d'une étude, même inachevée.

6.9 ANALYSE

Des rapports sur les résultats. Une étude analytique doit comprendre un exposé exhaustif portant sur les sources de données, les hypothèses formulées, les techniques utilisées et les résultats obtenus. Toutefois, les documents destinés à des auditoires particuliers doivent faire ressortir l'information qui intéresse ces auditoires à un niveau de détail approprié.

Une étude doit présenter les graphiques et les diagrammes jugés les plus intéressants et les plus pertinents. Le lecteur doit être convaincu que toutes les mesures appropriées ont été prises pour assurer la qualité de l'analyse. Les résultats des différents traitements faits dans le cadre d'une analyse sont habituellement trop nombreux; l'analyste doit faire un choix judicieux parmi ces renseignements tout en s'assurant de ne pas fausser les conclusions. Par exemple, il n'est pas nécessaire que le rapport d'analyse fournisse des détails au sujet des études de validation des modèles, à moins que ceux-ci ne présentent un intérêt technique particulier, mais il peut donner un aperçu de la façon dont un modèle a été construit. Les détails techniques peuvent souvent être présentés en annexe. L'étude ne doit pas sembler accentuer les procédures, sauf si elle s'adresse à un auditoire qui s'y intéresse.

L'analyste ne doit jamais supprimer des résultats qui ne corroborent pas ses théories ou qui vont à l'encontre du climat politique tel qu'il est perçu. Il doit être scrupuleusement objectif et présenter ses conclusions d'une manière scientifique et impartiale, et éviter d'exprimer son opinion ou de faire des recommandations.

6.10 RESSOURCES

On doit établir des rapports sur la consommation réelle des ressources, mesurée en fonction du temps. Dans le cas d'une activité complexe, il est rare que le budget prévu soit respecté. Il est donc important de rendre compte de toutes les dépenses en ressources financières et humaines. Ce genre d'information est aussi essentiel en vue de l'amélioration de la planification budgétaire, c'est-à-dire pour déterminer le personnel, les fonds et les délais requis.

6.11 DOCUMENTATION

(i) On doit établir une liste d'ouvrages de référence, c'est-à-dire des études et des documents théoriques et généraux présentant un lien avec le projet, mais non produits dans le cadre de ce projet. Une telle liste est établie pour toute la durée du projet et est utile à ceux qui désirent élargir leur connaissance du domaine général visé ou qui veulent obtenir plus de détails sur des aspects tels que les sources des méthodes.

(ii) On doit mettre en ordre et classer toute la documentation produite dans le cadre du projet. Pour être utile, la documentation doit pouvoir être consultée facilement. Il arrive trop souvent que l'unique exemplaire d'un document soit conservé par son auteur ou encore soit classé dans un ficher général mal ordonné, lorsque la personne en question ne fait plus partie de l'équipe de projet ou du Bureau. Il peut très bien arriver que l'on oublie l'existence de certains documents et que les mêmes travaux soient repris.

(iii) Établir un système efficace et systématique de mise à jour de la documentation, notamment dans le cas des guides ou des systèmes automatisés. Quand un changement est apporté, il n'est pas suffisant de rédiger une note de service pour demander que la documentation courante soit modifiée en conséquence. Un tel système de mise à jour doit aussi être maintenu pendant toute la durée du projet, même une fois le travail d'élaboration terminé.

FACILITÉ DE COMPRÉHENSION

6.12 FACILITÉ DE COMPRÉHENSION ET ACCEPTATION PROFESSIONNELLE

1. Les auteurs de documents doivent essayer de tenir compte de la catégorie de lecteurs à laquelle s'adresse chaque document. L'ébauche du document doit être soumise à un examen attentif fait par les membres de la direction, des représentants des lecteurs visés et des homologues en vue d'assurer tant la qualité du contenu que sa facilité de compréhension.

2. Tous les documents doivent être vérifiés avec soin. Dans la mesure du possible, on conseille d'utiliser des machines de traitement de textes (plutôt que des machines à écrire) en vue de faciliter la correction et d'éviter que les différentes versions soient entachées de nouvelles erreurs. (On ne doit pas oublier de dater les différentes versions.)Les machines de traitement de textes qui peuvent reproduire les symboles mathématiques sont particulièrement indiquées.

3. Les bibliographies doivent aussi être vérifiées, car il peut arriver que le numéro de page cité soit erroné, que les titres de revues spécialisées soient intervertis ou que l'orthographe des noms d'auteurs ne soit pas correct. Compte tenu de la fréquence des ajouts, des suppressions ou de la réorganisation des renvois faits dans le texte, on doit aussi s'assurer que chacune des sources mentionnées dans le texte figure dans la bibliographie et vice versa.

BIBLIOGRAPHIE

CONFÉRENCE DES STATISTICIENS EUROPÉENS, "Projet de directives pour l'établissement d'indications concernant la portée et la qualité des statistiques, à l'intention des utilisateurs", Genève, du 21 au 24 novembre 1983.

UNITED NATIONS, *Recommendations for the Preparation of Sample Survey Reports* (Provisional Edition), Statistical Series C No. 1 (New York, 1964).

INTRODUCTION

Les chapitres précédents portaient sur les mesures qui doivent être prises pour fabriquer un produit de qualité. Tous ces efforts peuvent être gravement compromis si l'on n'apporte pas le même soin à mettre l'information à la disposition des utilisateurs en faisant en sorte qu'elle soit facile à obtenir, à comprendre et à utiliser.

LIGNES DIRECTRICES CONCERNANT LA QUALITÉ

OBJET

- Fournir aux utilisateurs suffisamment de renseignements sur les statistiques publiées de façon qu'ils puissent interpréter et utiliser correctement l'information;

- s'assurer que les produits respectent le caractère confidentiel des données fournies à la source par les répondants;

- présenter les produits statistiques de manière à faciliter leur utilisation;

- assurer une commercialisation efficace des produits (c'est-à-dire fournir les produits dont les utilisateurs ont besoin);

- assurer une mise en marché efficace et économique des produits (c'est-à-dire rationaliser le processus de diffusion et choisir un mode de présentation correspondant à la valeur du produit, laquelle est fondée sur la demande et la volonté de payer des utilisateurs).

LIGNES DIRECTRICES

7.1 PLANIFICATION ET CONCEPTION

La planification et la conception des programmes de diffusion doivent tenir compte des facteurs suivants:

Qui sont les utilisateurs des données?
À quelles fins les utilisent-ils?
De quelle manière les utilisent-ils?
De quelle taille est la clientèle (réelle ou éventuelle)?

Les réponses à ces questions favorisent une planification et une conception efficace des systèmes de diffusion et, par voie de conséquence, facilitent l'utilisation de l'information fournie.

Par exemple:

1. Délais de production

   Certains utilisateurs sont très conscients du problème des délais de production. Dans certains cas, il est possible de produire des indicateurs sommaires bruts des principales variables afin de les diffuser dans des communiqués comme le **Quotidien, l'Infomat ou les bulletins de service** et ce bien avant que les données définitives ne soient connues. Une telle information est cependant présentée à titre provisoire et

sous réserve de modifications. On peut au besoin se servir de mesures comme les coefficients ou les ratios puisque ceux-ci peuvent atténuer dans une certaine mesure le caractère incertain de l'information. Dans ce genre de communiqués, on doit cependant mentionner les dates de publication de renseignements plus détaillés afin de réduire le nombre des demandes présentées à cette fin. Comme il n'est pas nécessaire de fournir une documentation détaillée quand on publie des données préliminaires, un renvoi suffit pour indiquer que l'on peut obtenir plus de renseignements sur demande.

2. Programmes régionaux

Dans les domaines où les utilisateurs ont besoin d'une gamme variée de données ventilées à un niveau régional, on peut envisager d'établir des programmes de diffusion axés sur une région ou un secteur géographique précis (par exemple, pour certaines enquêtes, des publications visent une province, une RMR et même une ville en particulier). Dans le cas des enquêtes dont la fréquence et les délais de production sont les mêmes, on peut aussi produire simultanément un certain nombre de publications axées sur des secteurs géographiques donnés.

3. Données exploitables par machine

Un grand nombre de chercheurs et d'analystes qui utilisent nos produits apprécieraient que les données soient présentées sous une forme ordinologue afin de pouvoir les transférer directement à leurs systèmes informatiques. Nombre d'enquêtes utilisent CANSIM, certaines produisent des bandes de microdonnées, des bandes sommaires, des bandes de totalisations, etc. La demande pour des données sur disquettes de micro-ordinateur ou pour des systèmes de communication tels que TÉLIDON est appelés à croître. Tous ces produits exigent la transmission de documentation. À l'heure actuelle, il s'agit surtout de renseignements techniques sur les moyens de transfert et d'accès aux données. Avec en plus une documentation sur la qualité, les utilisateurs seront mieux informés sur les données qu'ils utilisent et sur la façon de les obtenir.

4. Publications courantes

Les publications courantes représentent encore le mode le plus important de diffusion des données d'enquête. Toutefois, compte tenu de leur coût élevé et de la demande souvent restreinte, un examen périodique doit être entrepris en vue de décider s'il est justifié d'en poursuivre la production. Quand les bibliothèques sont les principaux utilisateurs d'une publication, on doit étudier la possibilité de publier les résultats sur microfiches directement à partir des états mécanographiques. Quand la demande ne justifie plus la production d'une publication en tout ou en partie, voici les mesures qui peuvent être prises:

1. Annoncer dans la publication visée, aussi longtemps que possible à l'avance, la décision de mettre fin à la publication et la date à laquelle cette mesure entrera en vigueur.

2. On doit aussi citer les autres sources d'information existantes sur le sujet en vue de causer le moins de dérangements possible aux utilisateurs.

3. L'information peut continuer d'être produite sous une forme moins coûteuse (par exemple imprimés d'ordinateur ou de XEROX 9700, bulletins de service, etc.) pendant une certaine période de transition, ce qui est un moyen d'assurer la continuité si l'on devait découvrir que la demande a été sous-estimée.

4. Diffuser l'information en "rétrospective" dans une publication de moindre fréquence.

5. Demandes spéciales

Les utilisateurs deviennent de plus en plus exigeants et constatent que nos produits préprogrammés ne répondent plus à leurs besoins. Ils devront être prêts à payer un prix raisonnable pour des produits dont la présentation est plus conforme à leurs besoins. Il faudra à cette fin mettre au point des systèmes d'extraction efficaces et souples qui permettront de produire ce genre de données sans trop de frais. Des méthodes de contrôle devront aussi être soigneusement élaborées pour éliminer toute possibilité de divulgation de renseignements confidentiels.

7.2 MISE EN OEUVRE

Après analyse, une stratégie de commercialisation doit être mise en oeuvre pour chaque enquête, de façon à consacrer une partie des ressources et des efforts aux différents modes de diffusion. Un rapport sur le sujet doit être établi et distribué sur demande.

7.3 DOCUMENTATION

Quel que soit le mode de diffusion choisi, une documentation aussi détaillée que possible doit être fournie aux fins de l'assurance de la qualité, c'est-à-dire:

- identifier l'activité statistique;

- énoncer les objectifs de l'activité statistique visée;

- joindre une copie du ou des questionnaires ayant servi à recueillir les données;

- déterminer le genre d'activité statistique (recensement, enquête par sondage, enquête fondée sur des données administratives, enquêtes hybrides et enquêtes dérivées);

- fournir les caractéristiques générales et particulières de ce genre d'activité statistique;

- déterminer les erreurs d'échantillonnage et d'observation qui découlent des différences entre la théorie et la pratique;

- définir le champ d'observation et les variables. Quand toutes les définitions ne peuvent être présentées pour des raisons pratiques, seules les variables les plus importantes (ou celles susceptibles d'être mal interprétées) peuvent être définies, mais comme elles doivent être toutes définies, on doit dans ce cas indiquer par un renvoi que des renseignements supplémentaires peuvent être obtenus;

- présenter une vue d'ensemble de la méthode de collecte utilisée et faire ressortir les différences qui existent du point de vue analytique entre cette méthode et d'autres méthodes utilisées à l'intérieur du Bureau ou ailleurs pour produire des données "semblables" ou connexes;

- analyser la qualité des données. Cette analyse doit comprendre une description des problèmes liés aux données et, s'il y a lieu, des mesures correctives prises pour y remédier. Elle doit aussi fournir des mesures qualitatives ou quantitatives de la qualité obtenue aux différentes étapes du processus de l'activité statistique.

Une documentation type doit comprendre les renseignements suivants:

- une évaluation de la qualité globale;
- la nature et l'incidence de la non-réponse; les effets des biais;
- les taux d'erreur (résultats des méthodes de contrôle de la qualité, etc.);
- des mesures de précision pour toutes les principales variables;
- l'incidence des méthodes d'échantillonnage et de pondération sur la désagrégation par domaine ou par secteur, s'il y a lieu;
- l'incidence sur la qualité des autres erreurs ou biais non attribuables à l'échantillonnage;
- des mises en garde concernant l'utilisation des données aux fins d'analyse (par exemple, quand on sait que des données seront utilisées sous forme de séries chronologiques, il est important de faire ressortir l'incidence des changements dans les notions de base utilisées, des modifications du champ d'observation, des interruptions dans la série et de tous les autres facteurs qui peuvent nuire à la comparabilité des données dans le temps);
- la documentation connexe, son contenu et la façon dont on peut l'obtenir (y compris le coût).

Dans le cas de séries dérivées ou de renseignements compilés à partir de sources auxiliaires, la documentation fournie avec le produit diffusé doit:

- expliquer s'il y a lieu le modèle ou la structure qui a été utilisé;
- expliquer les notions et définitions de base appliquées;

Dans le cas de statistiques ou d'analyses fondées sur l'utilisation de données administratives, la documentation fournie avec le produit diffusé doit:

- décrire la ou les sources des données;
- expliquer à quelles fins les données ont été recueillies à l'origine;
- analyser les avantages et les désavantages des données eu égard aux usages auxquels elles sont destinées (par exemple en tenant compte des biais liés aux théories utilisées et au champ d'observation);
- décrire la façon dont les données sont traitées après leur collecte et les mesures correctives qui sont prises au besoin pour remédier aux problèmes constatés dans la série originale;
- analyser la fiabilité des estimations et formuler des mises en garde au besoin.

BIBLIOGRAPHIE

CENTRAL STATISTICAL OFFICE (UK), *National Accounts Statistics Sources and Methods* (HMSO, London, 1968) pp. 39-42 et autres sections consacrées à la fiabilité des statistiques.

GONZALEZ, M.E., OGUS, J.L., SHAPIRO, G., ET TEPPING, B.J., "Standards for Discussion and Presentation of Errors in Survey and Census Data", *J.A.S.A.*, LXX (1975) No. 351, partie II.

STATISTIQUE CANADA, "Ligne de conduite concernant la production et la diffusion des mesures de la qualité des données et les rapports de méthodologie pour les enquêtes de Statistique Canada et leurs resultats", (Ottawa, 9 février 1978).

U.N., ECONOMIC COMMISSION FOR EUROPE, Conférence des statisticiens d'Europe, "Timeliness, Cost and Quality Attributes of Statistics", 8 juin 1982 (CES/464), "Some Reflections on Different Approaches, to the Problem of Presenting the Quality of Statistics", (document préparé par le Bureau central de la statistique de Suède), 26-28 mai 1982 (CES, document de travail No. 3).

8. *ACTIVITÉS STATISTIQUES FONDÉES SUR L'UTILISATION DE DONNÉES ADMINISTRATIVES*

INTRODUCTION

Les données administratives sont des données recueillies à une fin administrative quelconque, par exemple les déclarations d'impôt, les dossiers sur l'assurance chômage, les documents sur les importations et les exportations, et ainsi de suite. Il arrive souvent que Statistique Canada exploite ces données à des fins statistiques, soit parce qu'il s'agit de sources de renseignements uniques ou parce que ces sources fournissent une information d'une qualité relativement élevée. On évite alors d'accroître le fardeau des répondants et les dépenses de collecte sont réduites de manière considérable.

En général, les documents administratifs ne contiennent qu'un ensemble limité de variables qui sont nécessaires aux fins de l'administration, et de ce fait ils ne peuvent servir qu'à des applications particulières et non à des analyses générales. De plus, les concepts et les définitions employés dans le système administratif ne coincident pas toujours avec ceux qu'utilisent ou qu'estiment importants les analystes statistiques. Les données administratives sont sujettes aux changements d'ordre législatif ou réglementaire qui affectent le programme. Les programmes d'assurance de la qualité peuvent mettre l'accent sur des variables qui sont essentielles au programme administratif, et non sur celles qui intéressent l'analyste statistique. Les dossieurs administratifs qui son conservés à l'échelon municipal ou provincial peuvent être incompatibles entre eux. Des dossiers de données administratives complètes, comme ceux par exemple de l'impôt sur le revenu, ne portent pas sur la famille, et pourtant dans certains cas, des données fondées sur les familles seraient plus significatives que des données fondées sur les individus.

Les données administratives peuvent être fournies sous forme de copies des questionnaires administratifs originaux, de fichiers de micro-données ou de données globales. Aussi, nombre des procédures liées aux activités "traditionnelles" de collecte des données sont éliminées, réduites ou d'une certaine façon modifiées. Il est rare que l'on tienne compte de l'utilisation à des fins statistique des données dans le cadre du processus de planification de la collecte. Les objectifs statistiques ont habituellement moins d'importance que les objectifs administratifs. Le problème de la qualité du produit statistique qui découle de sources administratives, lequel n'est pas différent en principe de celui que posent les produits d'autres activités statistiques, doit être considéré sous un angle différent.

"L'organisme qui planifie, conçoit et met en oeuvre le processus de collecte des données administratives est responsable de l'assurance de la qualité ainsi que de l'évaluation et de la documentation de cette qualité, mais il se fonde à cette fin sur les objectifs administratifs fixés au départ". Statistique Canada "est dans la situation d'un utilisateur externe des données et doit par conséquent en déterminer la pertinence à des fins statistiques" (TRADUCTION, voir L. Swain, 1981). Une telle tâche se révèle beaucoup plus facile lorsque les données ont été relativement peu traitées (par exemple dans le cas des déclarations d'impôt) que si les données sont traitées sous forme de fichiers de macro-données.

Par conséquent, on ne procède pas de la même manière pour assurer la qualité d'activités statistiques fondées sur l'utilisation de données administratives et pour assurer celle d'activités dites "normales" puisque dans le premier cas, on doit obtenir, examiner et valider la documentation du fournisseur, c'est-à-dire évaluer la qualité des données en entrée. La qualité des processus de traitement et de documentation subséquents demeure la responsabilité du chef de l'activité statistique visée, qui doit appliquer les lignes directrices énoncées dans les chapitres 1 à 7.

LIGNES DIRECTRICES CONCERNANT LA QUALITÉ

OBJET

- Assurer et évaluer la qualité des données administratives utilisées en entrée.

LIGNES DIRECTRICES

8.1 PRODUCTION DES DONNÉES

L'utilisateur doit obtenir du fournisseur toute la documentation qui existe sur les données : notions et définitions de base, champ d'observation, méthodes de collecte, de codage, de saisie des données, de vérification, de correction des erreurs, d'imputation, de traitement, de mise à jour et de contrôle de la qualité, et rapports sur la qualité globale. On doit aussi chercher à obtenir une copie de toute étude d'évaluation réalisée par le fournisseur. Dans la mesure du possible, l'utilisateur doit trouver une personne ressource au bureau du fournisseur qui peut lui donner des renseignements précis sur le traitement, lui expliquer la documentation et le tenir informé de tout changement touchant les politiques, les règlements, les notions de base, les méthodes ou les procédures.

Il importe de souligner que cette documentation est non seulement utile pour l'évaluation de la qualité, mais aussi pour la structuration des vérifications et imputations subséquentes au micro-niveau et la résolution des problèmes au macro-niveau.

8.2 OBJECTIFS LIÉS À L'ACTIVITÉ STATISTIQUE ET À LA QUALITÉ

Les objectifs de l'activité statistique fondée sur l'utilisation de données administratives doivent être définis conformément aux lignes directrices énoncées au chapitre 1.

8.3 COUVERTURE

On doit évaluer dans quelle mesure le champ d'observation des données administratives correspond aux objectifs de l'activité statistique visée et si des ajustements doivent être apportés pour améliorer la relation. Au nombre de ces ajustements, on compte la repondération des données du fichier, la recherche de données supplémentaires provenant d'une autre source ou d'autres formes d'adaptation (voir 8.6).

Des méthodes permettant d'évaluer la qualité de la couverture sont proposées au chapitre 5.

8.4 CONCEPTION ET MISE EN OEUVRE

Les procédures liées à l'assurance de la qualité décrites aux chapitres 2 et 3 doivent être suivies.

8.5 VÉRIFICATION ET IMPUTATION

On doit déterminer les taux de non-réponse à des questions particulières (habituellement des questions laissées en blanc) et les taux de rejet à la vérification associés aux documents ou aux fichiers reçus. Les données doivent alors être soumises à une nouvelle vérification, car il est presque certain que les vérifications jugées essentielles par l'utilisateur diffèrent de celles imposées par le fournisseur. S'il y a lieu,

les codes indicateurs appropriés doivent être ajoutés au fichier. On peut alors au besoin procéder aux imputations qui conviennent.

Les systèmes de vérification et d'imputation doivent être suffisamment souples pour permettre la prise en compte des changements qui pourront être apportés par la suite aux lois, règlements et notions de base utilisées et aussi pour que des mises à jour puissent être apportées une fois que l'on connaîtra et que l'on comprendra mieux le processus de production des données.

On doit prêter une attention particulière à la façon dont les régions géographiques sont définies dans les dossiers administratifs ou déterminées à partir de tels dossiers. Voici quelques exemples de sources possibles de problèmes que l'on doit évaluer et régler:

- absence d'adresse ou de code postal;
- adresse ou code postal non à jour;
- adresse ou code postal ne correspondant pas à l'emplacement pour lequel on souhaite obtenir les renseignements (par exemple un lieu de travail au lieu d'une résidence ou vice versa, ou encore l'adresse d'un comptable);
- discordance entre l'adresse ou le code postal fourni et la région géographique visée.

8.6 MÉTHODES D'ESTIMATION

Si les données administratives fournies représentent un échantillon statistique de l'univers à l'étude, on doit vérifier la pertinence de l'échantillonnage en comparant si possible les estimations avec des estimations correspondantes produites à partir des données des fichiers de l'univers, ou encore avec des estimations obtenues d'autres sources. On peut aussi appliquer d'autres méthodes de vérification comme celles qui sont décrites au chapitre 5.

Il faudra souvent recourir à des méthodes d'ajustement quand les dossiers administratifs présentent des lacunes connues qui peuvent limiter leur utilité aux fins statistiques. Il peut s'agir de simples mises à l'échelle ou d'étalonnage à divers niveaux d'agrégation ou encore d'élaboration de modèles d'estimation qui rattachent les variables intéressantes aux variables pour les dossiers administratifs.

8.7 DOCUMENTATION

On doit résumer la documentation et l'information fournies à l'utilisateur et établir un système de renvoi approprié. Cette documentation doit être mise à jour et révisée périodiquement. On doit consigner tout changement dans les politiques, l'interprétation, les méthodes et les procédures.

8.8 ÉVALUATION DE LA QUALITÉ

L'évaluation de la qualité des données en entrée constitue un élément important de l'utilisation des données administratives, étant donné qu'au départ, la collecte et le regroupement des donnés n'ont pas été faites à des fins statistiques. Il faut à tout le moins examiner soigneusement les instruments et méthodes de collecte, les méthodes de contrôle qualitatif et de vérification, les concepts sous-jacents et les définitions utilisées. Il conviendrait également, dans la mesure du possible, de comparer les données avec des macro-données ou des micro-données provenant d'autres sources.

Il y aurait lieu de documenter le contrôle de la qualité des données en entrée et d'inclure dans la publication une description ou le rapport définitif de l'activité.

8.9 COMMUNICATION SUIVIE AVEC LE FOURNISSEUR

Si les données administratives sont fournies dans le cadre d'un programme régulier, il viendra un temps où le fournisseur réexaminera les objectifs administratifs et voudra apporter des modifications au programme. L'utilisateur statistique doit être prêt à exploiter de telles occasions pour tenter de convaincre le fournisseur d'apporter des changements aux méthodes de collecte ou de traitement des données qui peuvent aider à la réalisation des objectifs du programme statistique, et faire valoir au fournisseur l'importance de ces objectifs.

Le fournisseur doit aussi être tenu informé des problèmes ayant découlé de l'utilisation des données puisque ceux-ci peuvent révéler l'existence de problèmes fondamentaux dans l'exploitation de son propre programme. On peut également lui faire parvenir un exemplaire des documents produits dans le cadre du programme statistique qui utilise ses données en entrée. De cette façon, le fournisseur se rend compte que ses efforts sont appréciés et apprend à connaître l'utilisation qui est faite de ses données. Une telle attitude permet aussi de faire mieux accepter les changements proposés par l'utilisateur. D'autre part, l'utilisateur lui-même peut ainsi mieux comprendre les activités de collecte et de traitement des données.

L'utilisateur et le fournisseur pourront parvenir dans certains cas à établir un programme de collecte ou de traitement conjoint en vertu duquel l'utilisateur s'engage par contrat à s'acquitter de fonctions précises (par exemple le codage par secteur d'activité, le contrôle de la qualité) pour le compte du fournisseur, en raison de sa plus grande expérience de ce genre de fonctions. Un représentant de l'utilisateur peut aussi faire partie des comités de planification ou de direction mis sur pied par le fournisseur.

BIBLIOGRAPHIE

AUGER, E., "Postal Code Accuracy: Final Report," Statistics Canada, March 1981, mimeo.

AUGER, E. AND MUSSELY, A., "Comparaisons entre les données de Revenu Canada et d'autres sources", Statistics Canada, July 1984, mimeo.

DURKIN, T.M. AND SCHWARTZ, O., *"The SOI Quality Control Program"*, pp. 131-7 in Statistical Uses of Administrative Records, Vol. I (March 1984), *U.S. Internal Revenue Service.*

KOTEEN, G. S. AND GRAYSON, P. E., *"Quality of Occupation Information on Tax Returns"*, pp. 427-435 in Statistical Uses of Administrative Records Vol. II (July 1984), U.S. Internal Revenue Service.

LEYES, J. AND ROBINSON, L., "Quality Evaluation of Data Derived from Administrative Records: The Personal Income Tax Records", Paper to be presented at the meeting of the Statistical Society of Canada, Winnipeg, June 1985.

NORRIS, D.A. AND STANDISH, L.D., *A Technical Report on the Development of Migration Data from Taxation Records* (Ottawa, Statistics Canada, May 1983).

REDFERN, P., "A Study of the Future of the Census of Population: Alternative Approaches," Sept. 1983. A study commissioned by the Statistical Office of the EEC.

STATISTICS CANADA, "The Use of Administrative Files as Sources of Statistical Data in Relation to the Need for a Census of Population", March 9, 1983 (abridged edition) mimeo.

STATISTICS CANADA, *The Use of Administrative Social Data: Overview and Guide* (Ottawa, Statistics Canada, April 1985).

SWAIN, L., "Quality Assurance for Administrative Data" Statistics Canada, 1981, mimeo.

9. *ACTIVITÉS STATISTIQUES DÉRIVÉES*

INTRODUCTION

Les activités dérivées sont celles dans lesquelles on utilise des données recueillies par d'autres pour produire de nouvelles statistiques sur un sujet, un domaine ou une région géographique en particulier, grâce à un modèle ou à une structure, ainsi qu'à l'analyse, la réduction ou la transformation théorique de ces données. Des activités comme l'élaboration du système des comptes nationaux mettent en oeuvre des systèmes comptables fondés sur des modèles qui utilisent les données de Statistique Canada et des données d'autres sources pour compiler et estimer des mesures de certains phénomènes économiques et sociaux. D'autres activités ont recours à des techniques d'analyse statistique et économétrique qui sont appliquées à diverses séries de données provenant de Statistique Canada et d'autres sources en vue de la production d'estimations, d'indicateurs, d'indices, de prévisions, etc.

La qualité de telles activités statistiques dérivées dépend dans une large mesure de la qualité des données utilisées en entrée et donc de la mesure dans laquelle les objectifs des programmes de collecte des données répondent aux besoins et aux exigences de l'activité dérivée qui utilise ces données.

De plus les analyses qui sont effectuées et les méthodes qui sont utilisées pour produire les données dérivées peuvent servir à valider la qualité des données originales.

Quand des modifications sont apportées aux données originales à la suite d'un traitement (pour des raisons d'uniformité par exemple), il se peut que l'activité dérivée ait pour effet d'améliorer la qualité des données d'entrée (ou de réduire les erreurs). Ce fait reste cependant à être établi par une évaluation des méthodes utilisées dans le cadre de l'activité dérivée.

Si l'activité dérivée donne lieu à une modification des données originales, la qualité des données dérivées ne dépend plus simplement de la qualité des données d'entrée.

LIGNES DIRECTRICES CONCERNANT LA QUALITÉ

La responsabilité du maintien et de l'assurance de la qualité des activités dérivées incombe principalement aux chefs de ces activités, mais les chefs des activités statistiques de base ont aussi leur rôle à jouer.

9.1 DU POINT DE VUE DES ACTIVITÉS DÉRIVÉES

1. On doit soigneusement examiner les notions et définitions de base ainsi que les méthodes de collecte relatives aux données utilisées en entrée. Quand les données ne répondent pas entièrement aux exigences de l'activité dérivée, on doit expliquer toute différence ou lacune observée à l'intention des utilisateurs ou des évaluateurs des données dérivées, afin qu'ils connaissent les points forts et les faiblesses des données de base et la mesure dans laquelle celles-ci ont été corrigées. Autant que faire se peut, on doit décrire ou mesurer le biais introduit dans les données dérivées qui est attribuable à la différence des objectifs visés.

2. On doit s'efforcer de faire connaître les besoins particuliers de l'activité dérivée aux responsables de l'activité statistique de base de façon qu'ils apportent, si possible, les corrections appropriées à leurs programmes de collecte des données.

3. Quand l'activité dérivée consiste à comparer et à réunir des séries de données de sources diverses et à essayer de concilier les différences qui existent entre elles, le traitement peut aider à déterminer la validité des séries visées, car il soulève habituellement des questions sur l'exactitude des données ainsi comparées.

   Il est important que les analystes des activités dérivées et ceux des activités de base entretiennent des communications réciproques. Quand les divisions responsables des activités de base sont informées des travaux faits et de leurs résultats, elles peuvent améliorer au besoin les données originales. Cette communication permet également d'assurer que les comparaisons sont menées de manière adéquate et que les corrections apportées par le personnel responsable de l'activité dérivée sont justifiées.

4. Certaines activités dérivées utilisent des données réelles et des techniques d'estimation statistique pour produire les données. Elles peuvent utiliser une combinaison de données tirées d'enquêtes périodiques, des rapports entre variables dérivées d'enquêtes occasionnelles et des estimations fondées sur des modèles. Il importe d'établir la documentation pertinente pour expliquer les hypothèses formulées, les méthodes utilisées et la fiabilité relative des données dérivées produites.

5. On doit de temps à autre valider les hypothèses sur lesquelles reposent le modèle et la structure utilisés dans le cadre de l'activité dérivée ainsi que les méthodes utilisées.

6. Pour ce qui est de la planification, de l'analyse, de la documentation et de la diffusion, les lignes directrices énoncées au chapitres 1, 5, 6 et 7 peuvent aussi être appliquées aux activités dérivées. Quant à la documentation qui doit être présentée avec le mode de diffusion choisi, voici à nouveau quelles sont les lignes directrices à suivre dans le cas d'activités dérivées (se reporter à la section 7.3 du chapitre 7).

   La documentation fournie avec le mode de diffusion choisi doit:

   - expliquer au besoin le modèle ou la structure qui a été utilisé;
   - expliquer les notions et définitions de base appliquées;
   - expliquer la méthode de compilation;
   - mentionner toutes les sources de données;
   - déterminer la fiabilité des estimations;
   - formuler des mises en garde touchant l'utilisation des données aux fins d'analyse (par exemple, quand on sait que des données seront utilisées sous forme de séries chronologiques, il est important de faire ressortir l'incidence des changements dans les notions de base utilisées, des modifications du champ d'observation, des interruptions dans la série et de tous les autres facteurs qui peuvent nuire à la comparabilité des données dans le temps).

9.2 DU POINT DE VUE DES ACTIVITÉS DE BASE

1. Grâce à des consultations suivies, les chefs d'activité statistique doivent veiller à se tenir informés des objectifs des activités dérivées de Statistique Canada qui utilisent leurs données, en vue d'une meilleure compréhension de leurs besoins.

2. Les chefs des activités statistiques de base doivent tenir les chefs des activités statistiques dérivées informés de tous les changements apportés ou devant être apportés aux notions et définitions de base, au champ d'observation visé, et ainsi de suite.

3. Il est tout particulièrement important que les responsables d'activités dérivées puissent savoir à quoi s'en tenir au sujet de la fiabilité des données qui leur sont fournies. Toute faiblesse ou problème observé dans le cadre de l'exploitation de l'activité statistique de base doit être communiqué aux responsables des activités statistiques dérivées, car ces clients doivent connaître avec précision la nature de telles faiblesses pour pouvoir concevoir des mesures correctives efficaces.

4. Au moment de la revue des objectifs des programmes de collecte des données, on doit tenir compte des besoins des activités dérivées et modifier les programmes en conséquence dans la mesure où le permettent les contraintes financières, les ressources disponibles et les autres priorités de l'activité de base.

BIBLIOGRAPHIE

CENTRAL STATISTICAL OFFICE (UK), *National Accounts Statistics Sources and Methods* (HMSO, London, 1968).

DOMINION BUREAU OF STATISTICS, *National Accounts, Income and Expenditure by Quarters 1947-1957*, Ottawa, April 1959, pp. 49-50, Catalogue No. 13-511.

GOLDBERG, S.A., ET. AL., "The Canadian Quarterly National Accounts - A Critical Appraisal", pp 2-99 in S. Goldberg and P. Deane (eds.) *Studies in Short-Term National Accounts and Long Term Economic Growth* (London, Bowes and Bowes, 1965).

JOHNSON, A.G., "The Accuracy and Reliability of the Quarterly Australian National Accounts", Canberia, Australian Bureau of Statistics, 1982. Occasional Paper 1982/2.

KIRKHAM, P., "Some problems in devising measures of error for The National Accounts", *Bulletin of the International Statistical Institute*, XLVI (1975), pp. 188-198.

LAL, K., "Quality Assurance and The Input/Output Accounts", (Ottawa, Statistique Canada, 1983).

LECLERC, G. AND WELLS, S., "Problems of Determining and Measuring the Reliability of National Accounts: Canadian Experience", *International Association for Research in Income and Wealth*, Hungary Conference, 1973.

MOSER, SIR C., ET COLL., "Statistics Canada Methodology Review", chapitre V, "Quality Assurance in Surveys", mars 1980.

STATISTIQUE CANADA, "Revisions to the Income and Expenditure Accounts, 1953-1977", Ottawa, 1977.

WELLS, J.S., "Quality Assurance at Statistics Canada", document interne de Statistique Canada, janvier 1981.

STATISTICS CANADA LIBRARY
BIBLIOTHEQUE STATISTIQUE CANADA
1010324939

c. 3