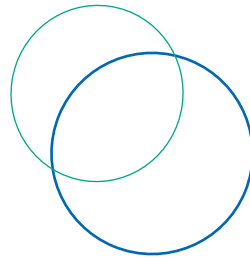
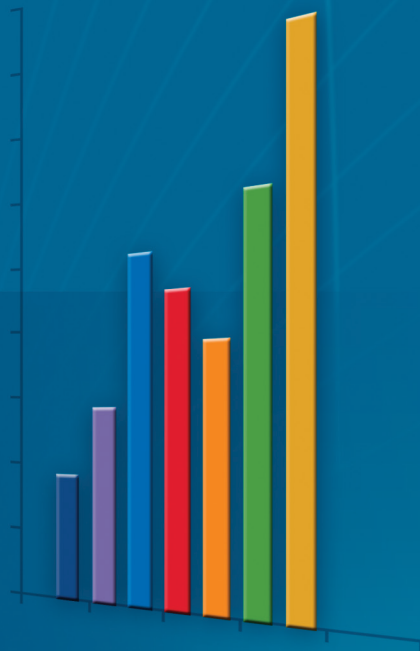
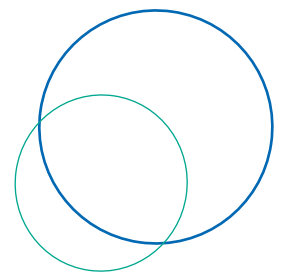


INSTITUT
DE LA STATISTIQUE
DU QUÉBEC

www.stat.gouv.qc.ca



Gestion de la qualité
Document de principes
sur la qualité dans les enquêtes



Pour tout renseignement concernant l'ISQ
et les données statistiques dont il dispose,
s'adresser à :

Institut de la statistique du Québec
200, chemin Sainte-Foy
Québec (Québec)
G1R 5T4
Téléphone : (418) 691-2411

ou

Téléphone : 1 800 463-4090
(sans frais d'appel au Canada et aux États-Unis)

Site Web : www.stat.gouv.qc.ca

Dépôt légal
Bibliothèque et Archives Canada
Bibliothèque et Archives nationales du Québec
3^e trimestre 2010
ISBN 978-2-550-59604-2 (version imprimée)
ISBN 978-2-550-59605-9 (PDF)

© Gouvernement du Québec
Institut de la statistique du Québec, 2010

Toute reproduction est interdite
sans l'autorisation du gouvernement du Québec.
www.stat.gouv.qc.ca/droits_auteur.htm

Septembre 2010

AVANT-PROPOS

La loi constitutive de l'Institut de la statistique du Québec (Institut) contient des dispositions claires quant aux obligations de qualité qu'il doit respecter. En effet, l'article 2 de la *Loi sur l'Institut de la statistique du Québec* stipule que « L'Institut a pour mission de fournir des informations statistiques qui soient fiables et objectives sur la situation du Québec quant à tous les aspects de la société québécoise pour lesquels de telles informations sont pertinentes » et qu'il « est le responsable de la réalisation de toutes les enquêtes statistiques d'intérêt général ». Par ailleurs, l'article 5 donne le mandat à l'Institut de traiter l'information qu'il produit « de façon à permettre des comparaisons à l'intérieur et à l'extérieur du Québec ». Pertinence, fiabilité et objectivité, de même que comparabilité sont donc les critères de base qui doivent guider la gestion de la qualité à l'Institut.

C'est pourquoi une politique générale, comportant des éléments relatifs à la gestion de la qualité des enquêtes statistiques, a été adoptée en novembre 2005. Par cette politique, l'Institut vise à systématiser et à améliorer les moyens déjà mis en place pour assurer la qualité de ses produits et services au bénéfice des utilisateurs. Le présent document s'inscrit dans le cadre intégré de gestion de la qualité que sous-tend la politique. Il vise le large public des clients, utilisateurs et partenaires de l'Institut qui, bien qu'ils connaissent les grands principes, ne sont pas des spécialistes des disciplines de pointe impliquées dans la réalisation d'enquêtes statistiques. Le *Document de principes sur la qualité dans les enquêtes* est donc conçu par l'Institut pour les informer de ses préoccupations générales en matière d'assurance qualité.

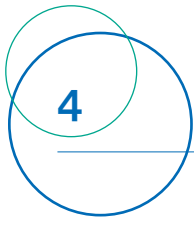
Le contenu du *Document de principes sur la qualité dans les enquêtes* est extrait du *Recueil de bonnes pratiques dans les enquêtes*, un outil de référence produit en avril 2009 à l'intention du personnel de l'Institut travaillant à la production d'enquêtes. Pour l'élaboration du recueil, un groupe de travail composé de professionnels de chacune des directions de l'Institut associées à de telles productions, s'est appuyé sur une recension des pratiques de référence reconnues par d'autres organismes statistiques, celles de Statistique Canada et de Statistics Finland notamment. Bon nombre des bonnes pratiques retenues dans le *Recueil* sont largement inspirées, parfois identiques, à celles choisies par ces deux organismes. Nous désirons les remercier de nous avoir permis d'utiliser leur matériel.

Enfin, l'Institut estime que la culture de la qualité doit s'incarner dans les comportements de tous au sein de l'organisme, ce qui inclut évidemment ses gestionnaires. Les travaux du groupe de travail sur la qualité dans les enquêtes ont donc été orientés et suivis par un comité composé de gestionnaires de l'Institut. Il importe donc de remercier tous ceux qui par leur leadership, leur rigueur et leur professionnalisme, ont contribué au dossier de la gestion de la qualité à l'Institut.

Le directeur général,



Stéphane Mercier



Cette publication a été réalisée
sous la coordination de :

France Lapointe

et la direction de :

Louise Bourque

Directrice générale adjointe aux services
à la production statistique

Les chapitres ont été produits par :

**Jimmy Baulne,
Lyne Bélanger,
André Breton,
Jocelyne Camirand,
Réjean Chevalier,
Richard Cloutier,
Marcel Godbout,**

**France Lapointe,
Hélène Lepage,
Marie-Thérèse Thibault,
Claude Vecerina,
Danielle Bilodeau,
Lyne Des Groseilliers
et Mario Haché**

Ont apporté leur précieuse collaboration :

Gaétane Dubé et **Sandra Gagnon** pour leur
participation aux activités du groupe de travail sur la
qualité dans les enquêtes

Clarisse Fréchette pour ses précieux commentaires
sur les chapitres 5, 6 et 7

Danielle Laplante, Gabrielle Tardif et **Esther Frève**
pour l'édition du document, la mise en page et la
révision linguistique

Pour tout renseignement concernant
le contenu de ce rapport s'adresser à :

Direction de la méthodologie et de la qualité
Institut de la statistique du Québec
200, chemin Sainte-Foy, 3^e étage
Québec (Québec) G1R 5T4

Téléphone : 418 691-2410

Télécopieur : 418 643-4129

Site Web : www.stat.gouv.qc.ca

TABLE DES MATIÈRES

INTRODUCTION	7	CHAPITRE 10	
CHAPITRE 1		Estimation.....	39
Objectifs, utilisation, utilisateurs, démarrage et planification de l'enquête	13	CHAPITRE 11	
CHAPITRE 2		Désaisonnalisation et estimation de la tendance-cycle	41
Concepts, variables et classifications.....	15	CHAPITRE 12	
CHAPITRE 3		Évaluation de la qualité	45
Base de sondage, population visée et couverture	19	CHAPITRE 13	
CHAPITRE 4		Contrôle de divulgation	47
Échantillonnage	21	CHAPITRE 14	
CHAPITRE 5		Analyse et présentation des données.....	51
Méthodes de collecte des données et élaboration du questionnaire et des instruments de collecte.....	23	CHAPITRE 15	
CHAPITRE 6		Diffusion des données	53
Opérations de collecte et de saisie des données.....	29	CHAPITRE 16	
CHAPITRE 7		Documentation et emmagasinage	55
Réponse et non-réponse	31	CHAPITRE 17	
CHAPITRE 8		Utilisation des données administratives	57
Validation et imputation.....	33	ANNEXE A	
CHAPITRE 9		Les méthodes de collecte.....	59
Pondération.....	37	BIBLIOGRAPHIE	63

INTRODUCTION

Contexte général

La gestion de la qualité

La *Politique générale en matière de gestion de la qualité de l'Institut de la statistique du Québec*¹ prévoit l'élaboration d'un *Recueil de bonnes pratiques dans les enquêtes* permettant l'harmonisation des façons de faire lors des phases de conception et de réalisation. On entend ici par « enquête » toute activité visant à recueillir ou à acquérir des données à des fins statistiques. Cela inclut les recensements, les enquêtes par sondage et la production de statistiques à partir de données provenant de dossiers administratifs. La constitution et le maintien d'un fichier administratif à des fins statistiques n'entrent pas dans cette catégorie, seule l'exploitation d'un tel fichier à des fins statistiques appartient au domaine des enquêtes.

Le *Recueil de bonnes pratiques* constitue un des mécanismes internes concourant à la qualité des processus, condition indispensable à la qualité des produits et services. Diffusé au personnel de l'Institut au printemps 2009, il constitue un guide à l'intention des équipes de travail responsables de la production d'enquêtes.

Cependant, le désir de l'Institut de s'adresser à un plus large public formé de clients, utilisateurs et partenaires a mené à la conception du *Document de principes sur la qualité dans les enquêtes*. Celui-ci est constitué des sections du *Recueil de bonnes pratiques dans les enquêtes* dont le contenu est moins spécialisé. Ainsi, le présent document est conçu par l'Institut pour informer ses clients, ses utilisateurs et ses partenaires de ses préoccupations générales en matière d'assurance qualité à chacune des étapes de réalisation de ses enquêtes statistiques.

La définition de la qualité

Comme de nombreux organismes statistiques, l'Institut définit la qualité d'un produit par l'ensemble des caractéristiques qui influent sur sa capacité à satisfaire un besoin donné, à permettre un usage prévu. L'Institut retient six dimensions comme critères de la qualité : la *pertinence*, la *fiabilité* et l'*objectivité*, la *comparabilité*, l'*actualité*, l'*intelligibilité* et l'*accessibilité*. Le tableau 1 résume la définition et les orientations générales d'assurance de la qualité pour chacune de ces dimensions.

Pendant la réalisation d'un projet d'enquête, les dimensions constituent des cibles à atteindre pour assurer la qualité de l'information statistique issue de cette enquête. Après sa réalisation, elles représentent les critères qui permettent d'évaluer la qualité de l'information statistique produite. En cours de réalisation, les décisions prises en ce qui a trait aux façons de faire et à leur mise en œuvre doivent tenir compte de toutes les dimensions de la qualité. L'assurance de la qualité d'un produit d'enquête dépend donc du travail effectué par l'équipe de réalisation et par l'ensemble du personnel impliqué, et ce, à toutes les étapes de réalisation de l'enquête. L'objectif ultime est d'assurer la pertinence, la fiabilité et l'objectivité, la comparabilité, l'actualité, l'intelligibilité et l'accessibilité de l'information statistique tirée de l'enquête.

1. Institut de la statistique du Québec, *Politique générale en matière de gestion de la qualité de l'Institut de la statistique du Québec*, gouvernement du Québec, janvier 2006.

Certains objectifs de qualité peuvent cependant s'opposer. Par exemple, il est difficile de maximiser à la fois l'actualité et la fiabilité et l'objectivité de l'information statistique. Ainsi, l'orientation générale retenue par l'Institut en matière de qualité de ses produits est la suivante : pour un niveau donné de ressources (humaines, matérielles et financières), on doit chercher à atteindre le meilleur équilibre entre les différentes dimensions de la qualité en vue de satisfaire le mieux possible les besoins des utilisateurs et permettre les utilisations prévues.

Tableau 1
Les dimensions de la qualité

Dimension de la qualité	Orientation en matière d'assurance qualité
Pertinence : concordance de l'information statistique par rapport aux besoins qu'elle doit combler.	Assurer que les données, les analyses et les synthèses produites par l'Institut répondent de la manière la plus complète possible aux besoins des utilisateurs.
Fiabilité et objectivité : conformité de l'information statistique à la réalité qu'elle doit décrire.	Garantir aux utilisateurs que l'information générée ou utilisée par l'Institut est la plus exacte possible; qu'elle est produite de façon à minimiser les erreurs typiques et à maximiser la rigueur scientifique.
Comparabilité : possibilité de la mise en parallèle ou de la combinaison d'une information statistique avec d'autres renseignements dans le temps, dans l'espace, entre domaines ou entre sources de données.	Assurer que l'information statistique s'appuie le plus possible sur des concepts, des classifications, des unités, des outils de mesure, des cadres conceptuels et d'intégration conformes aux standards internationaux, en laissant place aux particularités ou aux innovations qui sont dûment justifiées.
Actualité : disponibilité de l'information statistique en temps opportun et au moment prévu.	Assurer que l'information est produite et diffusée dans le délai le plus court possible ou à la fréquence appropriée. Assurer que le moment où elle devient disponible est conforme à l'échéance préalablement annoncée ou convenue.
Intelligibilité : facilité de compréhension de l'information statistique par les différents utilisateurs.	Fournir aux utilisateurs, de la façon la plus claire, mais aussi la mieux adaptée possible, les explications nécessaires à l'interprétation et à l'utilisation adéquates de l'information statistique.
Accessibilité : facilité avec laquelle les utilisateurs peuvent obtenir l'information statistique.	Faire en sorte que les conditions d'obtention de l'information statistique par les utilisateurs soient les plus simples et les plus conviviales possible (sous divers formats et selon divers modes d'accès).

Sources

Le contenu du *Document de principes sur la qualité dans les enquêtes* origine du *Recueil des bonnes pratiques dans les enquêtes* qui réunit des façons de faire reconnues par plusieurs organismes statistiques. Elles se veulent des suggestions, des pratiques recommandées; bon nombre d'entre elles sont couramment utilisées à l'Institut. En effet, l'élaboration du recueil s'appuie d'une part sur un *Rapport de constats des pratiques actuelles en matière de qualité dans les enquêtes* et d'autre part sur un Rapport de recension des pratiques de référence préparés par les auteurs du présent document. La recension de ces pratiques inclut notamment celles trouvées dans les lignes directrices concernant la qualité de Statistique Canada et de Statistics Finland. Des documents de l'Office for National Statistics et de Statistics New Zealand portant sur le sujet ont aussi été considérés dans cette recension. La section « Bibliographie » fournit une liste détaillée des documents consultés.

Portée et limites du document de principes

Parce que l'Institut souhaite informer au mieux le large public de ses clients, utilisateurs et partenaires, quels que soient leurs champs d'expertise et leurs besoins de renseignements, nous avons choisi de consigner dans le présent document les grands principes guidant le travail d'assurance qualité plutôt que les pratiques recommandées détaillées qui sont de niveau plus opérationnel. Le *Document de principes* ne prétend donc pas définir comment les enquêtes statistiques doivent être réalisées. D'autres sources peuvent être consultées à cet effet. Le lecteur intéressé à en savoir davantage à ce sujet est invité à consulter l'une ou l'autre des références fournies dans la bibliographie à la fin du document ou encore les experts en la matière de l'Institut.

Il importe aussi de souligner que le *Document de principes sur la qualité dans les enquêtes* ne couvre pas l'ensemble des produits de l'Institut. Les produits d'analyse, de modélisation ou les produits de diffusion de renseignements statistiques ne sont ni directement ni complètement couverts par le présent document. Il n'est pas non plus question ici des orientations générales en matière de qualité pour la constitution et le maintien de bases ou de registres de données. Cependant, quelques sections du document comportent des éléments pouvant aider à l'assurance de la qualité de ces produits. Les sections portant respectivement sur l'analyse et la présentation des données et sur la diffusion en sont des exemples.

Mentionnons enfin que la méthodologie d'enquête n'a cessé d'évoluer au cours des dernières décennies et il en sera de même dans le futur. Suivant ses valeurs de gestion et pour remplir sa mission, l'Institut devra suivre ce courant. Si bien que les principes devant guider le travail d'assurance qualité vont changer. Au fil du temps et à mesure que les façons de faire de l'Institut seront modifiées au bénéfice de ses utilisateurs, le *Document de principes sur la qualité dans les enquêtes* rendra compte de l'évolution de l'Institut. Il s'agit donc d'une première édition de ce document.

Structure du document

Les principes sont présentés selon un ordre logique correspondant au découpage d'une enquête en fonction des principaux processus ou des étapes qu'elle comporte, de sa conception à sa mise en œuvre. Le travail effectué à chaque étape de réalisation contribue à la qualité d'ensemble de l'enquête qui est multidimensionnelle. Cependant, au moment de concevoir et de réaliser une étape particulière, la préoccupation de l'Institut peut davantage se situer sur le plan de certaines dimensions de la qualité. Le tableau 2 présente le découpage d'une enquête en ses différentes étapes et les principales dimensions de la qualité ciblées lors de la réalisation de chacune.

Tableau 2

Les étapes de réalisation et les dimensions de la qualité de l'information statistique

Étapes de réalisation	Pertinence	Fiabilité et objectivité	Comparabilité	Actualité	Intelligibilité	Accessibilité
1. Objectifs, utilisation, utilisateurs (démarrage et planification de l'enquête)	✓	✓				
2. Concepts, variables et classifications	✓	✓	✓			
3. Base de sondage, population visée et couverture	✓	✓	✓	✓		
4. Échantillonnage		✓	✓			
5. Méthodes de collecte des données et élaboration du questionnaire et des instruments de collecte		✓	✓	✓		
6. Opérations de collecte et de saisie des données		✓		✓		
7. Réponse et non-réponse		✓				
8. Validation et imputation	✓	✓	✓	✓		
9. Pondération		✓	✓			
10. Estimation		✓	✓			
11. Désaisonnalisation et estimation de la tendance-cycle		✓	✓			
12. Évaluation de la qualité	✓	✓	✓	✓	✓	
13. Contrôle de divulgation	✓					✓
14. Analyse et présentation des données	✓	✓	✓		✓	✓
15. Diffusion des données	✓	✓	✓	✓	✓	✓
16. Documentation et emmagasinage			✓	✓	✓	✓
17. Utilisation de données administratives	✓	✓	✓	✓		

En plus de l'introduction, le document compte 17 sections principales. Une section a été constituée pour chacune des 16 étapes types de réalisation d'une enquête et une autre est consacrée à l'utilisation de données administratives qui comporte certaines particularités. Enfin, on trouve à la fin du document une section présentant les références bibliographiques consultées pour la préparation du document.

Les sections principales ont été conçues de sorte qu'on peut les consulter de façon indépendante. Une courte introduction commence chacune des sections. Elle fournit un aperçu global du travail fait à l'étape de réalisation traitée. On y établit aussi le lien avec les dimensions de la qualité. Dans chaque chapitre, la partie « **définitions** » qui suit l'introduction regroupe les éclaircissements nécessaires à la bonne compréhension de certains termes utilisés. Enfin, les « **principes** » sont énoncés. Il va sans dire que ceux-ci s'adressent au lecteur au fait de la production d'enquêtes statistiques et lui permettent de mieux saisir la nature des objectifs de qualité poursuivis par l'Institut.

CHAPITRE 1 – OBJECTIFS, UTILISATION, UTILISATEURS, DÉMARRAGE ET PLANIFICATION DE L'ENQUÊTE

Tout projet d'enquête commence par une phase de démarrage puis de planification qui permet, entre autres, de préciser les objectifs et l'utilisation de l'enquête ainsi que les besoins des utilisateurs.

Or, à l'étape de la définition des objectifs, utilisations et utilisateurs d'une enquête, c'est d'abord la dimension de la *pertinence* et, dans une moindre mesure, celles de la *fiabilité* et de l'*objectivité* des produits de l'Institut dont il est question. La *pertinence* de l'information statistique est définie par l'Institut comme sa concordance par rapport aux besoins qu'elle doit combler. Assurer la *pertinence* de l'information statistique, c'est faire en sorte que les données, les analyses et les synthèses que l'Institut produit répondent de la manière la plus complète possible aux besoins que les utilisateurs institutionnels et autres jugent les plus importants. Donc, assurer la pertinence d'une enquête, c'est s'assurer qu'elle répond aux besoins des utilisateurs, qu'elle est conforme à la mission et au mandat de l'Institut et qu'elle répond à ses orientations stratégiques.

Quant à la dimension *fiabilité* et *objectivité*, elle fait référence à la conformité de l'information statistique à la réalité qu'elle doit décrire. Pour l'Institut, assurer la *fiabilité* et l'*objectivité* de l'information statistique, c'est pouvoir garantir aux utilisateurs que l'information générée ou utilisée est la plus exacte possible, qu'elle a été produite de manière à minimiser les erreurs typiques qui entachent normalement les données et de façon à maximiser la rigueur scientifique dans les interprétations, analyses et synthèses qui en sont faites. Dans toute enquête, une définition adéquate des objectifs, utilisations et utilisateurs est donc une étape primordiale du processus visant à assurer l'exactitude de l'information statistique produite et sa conformité à la réalité qu'elle doit décrire.

Définitions

Dans *Le Petit Robert*, un objectif se définit de la façon suivante : « *But précis que se propose l'action* ». Dans le contexte d'enquêtes statistiques, les objectifs de l'enquête sont définis à partir des besoins pour lesquels on requiert l'information. Ils sont établis en fonction de la problématique de la gouvernance, de la recherche ou des hypothèses qui donnent naissance au besoin en information.

L'utilisation est l'application attendue des résultats. Elle peut viser différents usages, dont la prise de décision, la veille, la production de statistiques officielles, etc. Pour les enquêtes récurrentes, l'utilisation des résultats peut évoluer au fil du temps.

Les utilisateurs sont ceux qui se servent de l'information. Ils sont de toutes natures : organismes publics (ministères, organismes, etc.), groupes de recherche (universités, instituts scientifiques), organismes privés (entreprises, associations d'employeurs, associations professionnelles, syndicats, etc.) ou individus (professeurs, chercheurs, étudiants ou grand public). Dans certains cas, les clients et les principaux utilisateurs, qu'ils participent ou non au financement de l'enquête, peuvent être plus directement concernés par l'information à recueillir.

La phase de démarrage d'une enquête statistique consiste à déterminer les objectifs généraux, les principales hypothèses et contraintes qui caractérisent le projet ainsi qu'une estimation très large des dépenses et des échéanciers prévisibles. Cette phase exploratoire se retrouve au début du processus d'une enquête statistique, lors des discussions et des échanges avec les clients et utilisateurs principaux. Par la suite, une phase de planification de l'enquête vient notamment préciser, dans un énoncé de contenu puis, éventuellement, dans un plan de projet, les objectifs spécifiques et les éléments de qualité attendue et autres paramètres connus.

Principes

- i) La pertinence d'une enquête se décide lors de la phase de démarrage, en établissant le lien entre les besoins des clients, des utilisateurs et l'intérêt de l'Institut pour une telle enquête. Il va sans dire que cet intérêt dépend de la mission et du mandat de l'Institut, des balises qu'il s'est données en matière de production statistique et de ses moyens. De ces besoins et de l'intérêt de l'Institut découleront des objectifs. Une définition claire des objectifs et de l'utilisation d'une enquête est une condition préalable à la phase de planification de celle-ci qui se fera en collaboration avec les parties prenantes à l'enquête (clients, utilisateurs et équipe de réalisation).
- ii) À la phase de démarrage, les échanges avec les clients et utilisateurs permettent de définir les grands objectifs de l'enquête. Lorsque ces échanges sont concluants, la phase de planification peut commencer. À cette étape, il est important de déterminer des objectifs spécifiques et d'en faire un énoncé de contenu clair en indiquant les hypothèses à mettre à l'épreuve et les exigences de données spécifiques. Le consensus au sein de l'équipe de travail de même que chez les clients et utilisateurs sur l'utilisation et les objectifs précis d'une enquête facilite sa planification.
- iii) Dans certains cas, la production de statistiques est requise par la loi. Les objectifs et l'utilisation de ces statistiques sont donc définis par une loi. À titre d'exemple, la loi constitutive de l'Institut l'oblige à produire le bilan démographique et la comparaison de la rémunération globale des salariés de l'État avec celle des autres salariés québécois. Dans d'autres cas, les objectifs, l'utilisation et les utilisateurs découlent de grandes orientations gouvernementales. Parfois, ils sont définis ou mis sur pied par des comités d'orientation ou des tables de concertation.
- iv) Il importe de s'assurer que les étapes de définition des objectifs, de l'utilisation et des utilisateurs de même que les phases de démarrage et de planification de l'enquête ne sont pas escamotées. Ultiment, la pertinence même de l'enquête en dépend, ainsi que la fiabilité et l'objectivité des résultats qui en découleront.

CHAPITRE 2 – CONCEPTS, VARIABLES ET CLASSIFICATIONS

L'essentiel du cadre conceptuel et opérationnel d'une production statistique est constitué de l'ensemble des concepts, des variables et des classifications (nomenclatures). Leurs définitions, préférablement normalisées, doivent respecter les utilisations prévues des données.

Les choix des concepts, variables et classifications sont cruciaux pour assurer la *pertinence* de l'information statistique. En effet, ces choix sont d'abord faits parce qu'ils permettent de satisfaire aux objectifs de l'enquête : les concepts, variables et classifications doivent répondre aux besoins que l'enquête doit combler et permettre les utilisations prévues.

Quant à la *comparabilité* de l'enquête dans le temps ou avec d'autres enquêtes, elle repose en grande partie sur l'utilisation de concepts, variables et classifications communes et normalisées.

Enfin, il importe à cette étape de réalisation de l'enquête de préciser les concepts et classifications choisis. Quant aux variables déterminées pour mesurer ces concepts, elles doivent être basées sur des questions, et plus généralement sur le recueil de renseignements, faciles à comprendre par les répondants et par les intervieweurs, le cas échéant. La clarté des formulations utilisées est de mise afin de minimiser les erreurs de réponse, de non-réponse et toute autre erreur de mesure (ex. : erreur de rappels). La dimension *fiabilité et objectivité* de l'information statistique produite à l'issue de l'enquête en dépend.

Définitions

Les concepts, les variables et les classifications (nomenclatures) forment ensemble l'essentiel du cadre conceptuel et opérationnel d'une production statistique.

Les concepts sont utilisés pour définir le sujet d'enquête et d'analyse, les unités statistiques à décrire ou la population visée par l'enquête. De fait, ils renvoient à des caractéristiques ou à des attributs généraux d'une unité statistique ou d'un groupe d'unités statistiques semblables.

Les variables sont les indicateurs utilisés pour mesurer les concepts. Leurs définitions, ainsi que celles des concepts, doivent être précises, sans équivoque, en prévision de l'analyse qui sera faite des résultats de l'enquête.

On désigne parfois par « variables dérivées » ou « indices » des indicateurs complexes utilisés pour mesurer certains concepts. Par exemple, on peut construire un indice en combinant les réponses à plusieurs questions dans une enquête ou encore en ayant recours à un calcul basé sur plusieurs renseignements recueillis auprès d'un répondant. La construction de variables dérivées et d'indices requiert souvent un traitement plus important des données de base. Plusieurs indices complexes comportent des difficultés spécifiques dont font état des publications scientifiques particulières. Dans le présent document, la question de la construction d'indices est toutefois abordée de façon générale. Ainsi, comme dans le cas des variables simples, il est important de faire un choix judicieux des variables dérivées et indices pour s'assurer de mesurer au mieux les concepts choisis pour répondre aux besoins de l'enquête. En plus, un soin particulier doit être apporté aux traitements des données nécessaires à la production de tels indices, le souci étant toujours de minimiser les erreurs dans les données produites. Pour le reste du document, le terme général « variable » servira à désigner tant les variables simples, les variables dérivées que les indices. Les principes devant servir à la définition des variables s'appliquent en effet de façon générale à toutes ces formes de mesure des concepts.

La classification, dans sa forme la plus simple, consiste en une structure hiérarchisée de noms de groupes et de sous-groupes auxquels correspondent des identificateurs (codes numériques ou alpha-numériques). Les classifications comprennent ainsi les définitions des groupes (noms, codes et description de chaque groupe) et une présentation des concepts de base et des principes et critères de classification. Les classifications peuvent être catégorisées comme suit :

- les classifications décrivant la population, les personnes et leurs conditions de vie (classifications démographiques);
- les classifications décrivant l'activité économique et l'économie nationale;
- les classifications géographiques et les classifications pour les statistiques environnementales;
- les classifications spéciales, telles que la classification des maladies, la classification sur la criminologie, la classification des activités de la culture et des communications, etc.

Les critères de classification choisis pour regrouper de façon systématique les données recueillies dans une enquête doivent être pertinents aux besoins de l'analyse.

Principes

- i) Afin de pouvoir tirer de bonnes conclusions à partir d'un ensemble de données, il est extrêmement important que les utilisateurs aient une connaissance des concepts qui sous-tendent les données, c'est-à-dire qu'ils doivent savoir ce que les données devraient mesurer. L'Institut doit informer les utilisateurs en cette matière.
- ii) La classification fait partie intégrante des statistiques produites dans une enquête. Les classifications standards sont donc des instruments clés de la statistique officielle à des fins de comparabilité notamment.
- iii) Bien que l'utilisation de définitions harmonisées de concepts, de variables et de classifications aide les utilisateurs à comparer et à intégrer les données, il peut être nécessaire de modifier ces définitions dans le but de respecter les utilisations prévues des données.
- iv) Les définitions des concepts, des variables et des classifications doivent être soigneusement documentées. Par ailleurs, il faut noter toute différence par rapport aux normes ou aux définitions utilisées pour produire les statistiques. Cette documentation est particulièrement importante pour les utilisateurs qui désirent se servir des données pour un usage autre que celui prévu au plan initial.
- v) La classification doit être systématique et doit permettre de classer les unités de façon cohérente en utilisant un critère accepté. Les groupes formés doivent être sans équivoque, exhaustifs et exclusifs; la catégorie « autres » ne doit contenir que très peu d'unités relativement aux autres catégories.

- vi) Les variables et les classifications utilisées pour mesurer un concept doivent également tenir compte de facteurs comme :
- la facilité d'obtenir l'information désirée (ex. : la facilité de compréhension du répondant, la clarté du langage utilisé dans le questionnaire, etc.);
 - le fardeau imposé au répondant;
 - la méthode de collecte;
 - le traitement des données;
 - la présence d'information dans les dossiers administratifs;
 - les coûts associés à la collecte et au traitement.
- vii) Il faut changer ou modifier les variables ou classifications devenues désuètes avec le temps.

CHAPITRE 3 – BASE DE SONDAGE, POPULATION VISÉE ET COUVERTURE

Que ce soit dans le cas d'enquêtes par recensement ou par échantillonnage, la couverture, la complétude, l'actualité, le contenu en information auxiliaire et l'exactitude de la base de sondage sont des facteurs critiques de son adéquation.

Lors du choix d'une base de sondage pour une enquête et de l'évaluation de la couverture de la population visée qu'elle offre, c'est de la dimension *fiabilité et objectivité* dont on se préoccupe d'abord. Toutefois, le travail sur la qualité comporte également ici des éléments relatifs aux dimensions de *comparabilité*, d'*actualité* (la base de sondage devant être le plus près possible de la période de référence) et également de *pertinence* (le plus en lien possible avec la population visée).

Définitions

Une base de sondage est soit une liste, soit un document ou un système qui délimite et définit les éléments d'une population visée (ou population cible) et permet d'y avoir accès. La population visée correspond à l'ensemble d'éléments sur lequel on désire recueillir des renseignements et produire des estimations.

Parfois, pour diverses raisons, dont celle de minimiser les coûts de collecte, il est convenable d'exclure certaines unités de l'enquête. On définit alors la population enquêtée qui correspond à la population visée moins les unités exclues. La population enquêtée (ou échantillonnée) correspond donc en pratique à la population à partir de laquelle on sélectionne l'échantillon.

Le plan de sondage fait partie intégrante du plan global de l'enquête. Il est essentiellement constitué des éléments suivants : le plan d'échantillonnage, les méthodes d'estimation et les méthodes de mesure de la précision. Le plan d'échantillonnage précise comment l'échantillon est tiré de la base de sondage.

La couverture désigne la proportion des éléments de la population visée, qui se retrouvent dans la base de sondage.

Le potentiel d'inférence, la pertinence et l'exactitude des résultats d'une enquête statistique dépendent de l'utilisation d'une base de sondage adéquate.

Principes

- i) La population visée par l'enquête devrait être définie de telle sorte que l'inclusion de chaque unité dans la population peut être déterminée sans ambiguïté. La définition doit donc préciser les critères d'inclusion dans la population.
- ii) Une enquête requiert une base de sondage qui couvre toutes les unités de la population visée une et une seule fois, avec chaque unité clairement distincte des autres. Idéalement, chaque unité est accompagnée des renseignements pertinents qui pourront être utilisés pour le plan de sondage ou le contact ou l'estimation.
- iii) Un ensemble de bases de sondage de référence en fonction du type de population visée à joindre devrait être disponible.

- iv) La base de sondage doit être conforme à la population visée par l'enquête et contenir une sous-couverture et une surcouverture minimales (celles-ci devraient faire l'objet d'une évaluation).
- v) L'information contenue sur la base de sondage doit être mise à jour régulièrement afin de tenir compte des unités inactives (décès) et des nouvelles unités (naissances) ou de tout autre changement observé dans la population visée ou en ce qui concerne l'information auxiliaire.
- vi) Les caractéristiques des unités de la base de sondage doivent être de grande qualité, car elles ont un impact majeur sur tous les processus de l'enquête.
- vii) Lorsqu'une base de sondage est constituée à l'aide de plusieurs sources différentes, il est primordial de s'assurer de la qualité des variables d'appariement. Cela permet d'éviter les problèmes de couverture pouvant, par exemple, être liés à la création de doublons ou encore à des pertes d'unités au moment de l'appariement.
- viii) Il est essentiel d'évaluer la proximité des unités de la base de sondage avec les unités que l'on souhaite mesurer dans une enquête, et aussi de noter les différences et similitudes existant dans leur définition.
- ix) Le rapport de l'enquête devrait inclure des tableaux présentant des statistiques relatives aux principales variables auxiliaires se trouvant dans la base de sondage.

CHAPITRE 4 – ÉCHANTILLONNAGE

L'échantillonnage est la sélection d'un ensemble d'unités tirées d'une population enquêtée (ou échantillonnée). C'est sur cet ensemble d'unités, appelé échantillon, que seront réalisées les activités de collecte des données. Dans le but de réduire les coûts, cette méthode est souvent utilisée plutôt que de faire un recensement complet des unités de la population. Un échantillon choisi adéquatement permettra d'obtenir des estimations suffisamment précises des paramètres à l'étude.

L'échantillonnage étant fortement associé à la précision des paramètres estimés, il va sans dire que la dimension *fiabilité et objectivité* est la principale dimension de la qualité touchée à l'étape de l'échantillonnage. Certains principes de référence en matière d'échantillonnage permettent aussi l'assurance de la *comparabilité* entre les réalisations d'une enquête récurrente notamment.

Définitions

Le choix de la méthode d'échantillonnage dépend d'un certain nombre de facteurs dont les caractéristiques de la population, la disponibilité d'une base de sondage appropriée, la disponibilité de variables auxiliaires, le degré de précision souhaité, le mode de collecte des données et les budgets prévus. Tous ces facteurs influent directement sur la qualité des données.

L'échantillonnage probabiliste suppose une sélection aléatoire d'un échantillon à partir de la base de sondage de manière que toutes les unités de la population enquêtée (ou échantillonnée) aient une probabilité connue et positive d'être choisies. L'échantillonnage probabiliste permet d'inférer à la population et de calculer la précision associée à chacun des paramètres estimés.

Les méthodes d'échantillonnage utilisées pour produire des statistiques officielles sont en général probabilistes. Cependant, pour des raisons d'économies ou à cause d'une grande difficulté à obtenir la coopération des répondants, il arrive parfois qu'on choisisse d'utiliser un échantillon de convenance ou échantillon obtenu par quotas, qui sont deux exemples d'échantillons obtenus par une méthode d'échantillonnage non probabiliste. L'utilisation d'une telle méthode comporte cependant des limites importantes (biais potentiels, inférence restreinte, mesures de précision quasi inexistantes).

Principes

- i) Idéalement, les méthodes d'échantillonnage utilisées ne doivent pas reposer sur un modèle qui limite l'usage des statistiques, puisque les statistiques officielles ont besoin d'être utilisées avec confiance dans un large éventail d'usages. Ainsi, à moins de raisons exceptionnelles, utiliser l'échantillonnage probabiliste pour choisir un échantillon de la population enquêtée plutôt qu'une méthode non probabiliste. Toute décision de recourir à une méthode non probabiliste doit être dûment justifiée par le responsable du projet.
- ii) La méthode d'échantillonnage choisie doit permettre d'atteindre les objectifs de la façon la plus adéquate possible tout en respectant les contraintes associées à l'enquête et en limitant le fardeau des répondants au plus bas niveau.
- iii) Les méthodes recourant à l'utilisation de variables auxiliaires peuvent permettre d'améliorer la précision des estimateurs. Le choix du plan de sondage doit tenir compte de l'accessibilité à cette information.

CHAPITRE 5 – MÉTHODES DE COLLECTE DES DONNÉES ET ÉLABORATION DU QUESTIONNAIRE ET DES INSTRUMENTS DE COLLECTE

MÉTHODE DE COLLECTE

La méthode de collecte de données est en lien avec plusieurs aspects d'une enquête : le plan de sondage, le calendrier de réalisation, la qualité des résultats tirés de l'enquête, etc. Ainsi, le choix de la méthode de collecte n'est pas une décision isolée dans la planification d'une enquête, puisqu'il influence l'ensemble du processus. Par exemple, ce choix a un impact sur l'élaboration du questionnaire, sur la qualité et la quantité de données à recueillir, sur les mesures mises en place pour en assurer la sécurité et la confidentialité, et sur les coûts et le calendrier de réalisation de la collecte.

La méthode de collecte doit notamment prendre en considération le fardeau du répondant. Il importe d'élaborer une stratégie de collecte qui facilite le travail du répondant tant en ce qui touche sa compréhension des renseignements recherchés que de l'accès au questionnaire d'enquête. Cela peut, par exemple, vouloir dire qu'on prépare des aides visuelles accompagnant le questionnaire d'enquête ou encore qu'on considère un mode mixte de collecte, au choix du répondant. En outre, la méthode de collecte doit aussi prendre en compte le fardeau de l'intervieweur lorsque ce mode d'administration du questionnaire est choisi.

Il va de soi que la méthode de collecte des données doit assurer qu'elles sont le plus possible exemptes de biais (ex. : les biais de réponse, les biais de non-réponse, les biais de rappels, etc.), puisque ces données sont celles qui servent à produire l'information statistique désirée. La principale dimension de la qualité de l'information statistique en lien avec cette étape est donc la dimension *fiabilité et objectivité*. Par ailleurs, toutes les méthodes de collecte des données ne sont pas équivalentes quant à l'impact sur la durée de la période de collecte. C'est pourquoi le choix de la méthode de collecte est une étape déterminante de la dimension *actualité* de l'information statistique. Enfin, dans un contexte d'enquêtes répétées dans le temps ou d'autres formes de comparaison d'enquêtes, la méthode de collecte des données peut avoir un impact important sur la *comparabilité* de l'information statistique produite.

Définitions

La méthode de collecte se définit principalement à l'aide de deux paramètres : le mode de collecte et le mode d'administration du questionnaire. Il existe quatre modes de collecte distincts : téléphonique (assisté par ordinateur ou non), postal, en face à face (assisté par ordinateur ou non) et par voies électroniques (incluant, par exemple, le recours au Web et les échanges sécurisés de fichiers). De même, il existe deux modes d'administration du questionnaire : les questionnaires administrés par un intervieweur et les questionnaires autoadministrés où le répondant remplit lui-même le questionnaire. Généralement, les questionnaires autoadministrés sont employés lors des enquêtes postales et des enquêtes utilisant le Web, alors que les enquêtes téléphoniques et les enquêtes en face à face utilisent des questionnaires administrés par un intervieweur. Dans les deux modes d'administration du questionnaire, la technologie d'acquisition des données peut être électronique (entrevue assistée par ordinateur ou questionnaire électronique) ou il peut s'agir de questionnaire papier.

Par ailleurs, il existe deux principaux types d'enquêtes : les enquêtes longitudinales où les mêmes unités sont interrogées à plusieurs reprises et les enquêtes transversales pour lesquelles un échantillon indépendant est choisi à chaque occurrence. Finalement, les enquêtes s'effectuent auprès de deux types d'unités : les ménages (ou les personnes) et les entreprises.

MÉTHODE DE COLLECTE

Principes

- i) Ce sont les objectifs de l'enquête qui guident le choix de la méthode de collecte, pas l'inverse.
- ii) Plusieurs facteurs sont à considérer dans le choix de la méthode de collecte. Parmi ceux-ci, notons l'échéance à laquelle les résultats doivent être disponibles, le budget associé à l'enquête, le type d'information à recueillir, le type de population à sonder, le fardeau du répondant et la comparabilité avec d'autres enquêtes.
- iii) La durée de la période de collecte dépend beaucoup du mode de collecte retenu. Par exemple, les enquêtes téléphoniques sont en général plus rapides que les enquêtes en face à face ou les enquêtes postales (Voir, par exemple, Lyberg, L. et autres (1997) ou Groves, R. M. et autres (2002) à ce sujet).
- iv) La durée de la période de collecte est également influencée par la quantité totale d'information à obtenir. Ainsi, la quantité d'information à recueillir auprès de chacune des unités d'enquête et le nombre d'unités échantillonnées sont des facteurs déterminants de la durée de la collecte.
- v) La durée de la période de collecte et le taux de réponse sont souvent directement liés : plus on alloue de temps et de ressources aux opérations de collecte, meilleur est le taux de réponse. Le taux de non-réponse toléré est donc toujours un compromis.

Une règle informelle généralement acceptée veut que seulement deux des trois conditions suivantes puissent être réalisées simultanément dans une enquête :

- Coûts faibles;
- Taux de réponse élevé;
- Calendrier de réalisation court.

Pour les enquêtes à caractère facultatif, les différents modes de collecte produisent habituellement des taux de réponse différents. Par exemple, on obtient en général des taux de réponse plus faibles avec une collecte postale et une collecte utilisant le Web. (Voir Lyberg, L. et autres (1997) pour en savoir plus sur le sujet.)

- vi) Pour maximiser le taux de réponse et la qualité des données, il est important de considérer que les questionnaires administrés par téléphone devraient durer moins de 20 minutes. Exceptionnellement, pour des sujets intéressants particulièrement les répondants visés, il est possible de prolonger le temps d'entrevue. On devrait toutefois, en toutes circonstances, éviter de dépasser 30 minutes.

De façon générale, les entrevues en face à face peuvent se permettre d'être plus longues. Enfin, pour les questionnaires postaux, plus il y a de questions, plus le taux de réponse a tendance à diminuer. (Voir, par exemple, Statistics Finland, (2002), p. 55.)

- vii) En principe, les collectes utilisant le Web devraient être réservées aux enquêtes auprès des entreprises. En effet, dans l'état actuel des choses, les enquêtes auprès des ménages utilisant le Web font face à trois écueils de taille soit l'absence de base de sondage adéquate et des problèmes de couverture et de non-réponse importants. Même dans le cas d'enquêtes auprès des entreprises, on devrait prendre en considération la capacité des entreprises à répondre adéquatement au questionnaire administré par Internet et la qualité des bases de sondage disponibles.

La collecte utilisant le Web peut être indiquée dans un contexte de mode mixte d'enquête afin de permettre au répondant d'opter pour le mode de collecte qui lui convient le mieux. Une préoccupation particulière devrait alors être accordée au type de population visée.

Puisque ce mode de collecte implique le recours à un questionnaire autoadministré, la conception d'une telle enquête doit se faire avec le souci constant de faciliter la tâche du répondant pour fournir les renseignements recherchés. Il faudrait en plus prévoir alors un soutien technique pour les répondants.

Enfin, il serait pertinent de prévoir de recueillir des mesures du déroulement de la collecte pour les questionnaires administrés par Internet (ex. : indicateurs de suivi de la collecte). De telles mesures permettent de rendre compte de la qualité des renseignements recueillis selon ce mode de collecte. À plus long terme, une telle approche devrait également permettre à l'Institut une amélioration de la qualité d'enquêtes réalisées à l'aide du Web.

Afin de compléter cette énumération de principes, l'annexe A offre un aperçu des principaux avantages et désavantages de chaque méthode de collecte en fonction du type d'enquête.

ÉLABORATION DU QUESTIONNAIRE ET DES INSTRUMENTS DE COLLECTE

Le questionnaire joue un rôle clé dans le cadre du processus de collecte de renseignements. Son influence sur la qualité des données recueillies, les coûts, le comportement des répondants, la prestation de l'intervieweur et les relations avec les répondants s'avère déterminante. Un bon questionnaire comporte des énoncés clairs et faciles à comprendre et des instructions précises pour le répondant ou l'intervieweur; il est aussi facile d'accès.

Plutôt qu'un questionnaire à proprement parler, certaines enquêtes ont recours à d'autres types d'instruments de collecte, comme des applications de saisie de données. Ceux-ci prennent alors la forme d'une série d'instructions à l'intervieweur pour la collecte des renseignements pertinents à la mesure des concepts d'intérêt pour l'enquête. Comme le questionnaire, ce type d'instrument est adéquat s'il comporte des énoncés clairs et faciles à comprendre pour l'intervieweur, qui lui permettent de faire une « mesure » fiable et objective des variables désirées.

L'influence directe du questionnaire ou des autres instruments de collecte sur le contrôle des erreurs de non-réponse, de réponse ou des autres erreurs de mesure rend compte de l'importance de cette étape de réalisation pour assurer la qualité, en termes de *fiabilité* et d'*objectivité* de l'information statistique produite. La dimension d'*actualité* fait également partie des principales cibles de qualité à cette étape de réalisation de l'enquête, puisque la période de référence des questions soumises au répondant ou des « mesures » prises à son sujet est un des paramètres importants du questionnaire. De même, tout comme pour la méthode de collecte, un souci particulier est apporté ici à la dimension de *comparabilité* de l'information statistique lorsqu'on élabore un questionnaire ou un instrument de collecte pour une enquête répétée dans le temps ou pour une enquête qui doit être comparée avec d'autres. L'utilisation de questions et de mesures identiques est alors de rigueur.

Définitions

Le questionnaire consiste en une série de questions qu'on formule en vue de recueillir des renseignements auprès d'un répondant. On désigne parfois le questionnaire sous le terme « instrument ». Celui-ci convient également pour désigner du matériel de collecte ne se présentant pas sous la forme d'une série de questions. Il existe en effet des situations où des mesures d'intérêt pour l'enquête ne prennent pas cette forme : qu'on pense, par exemple, à des mesures anthropométriques dans une enquête de santé ou encore à une description détaillée des emplois au sein d'une entreprise dans une étude portant sur la rémunération comparée. L'instrument de collecte se présente alors comme une série d'instructions précises aux intervieweurs pour la collecte des renseignements pertinents aux calculs des variables et indicateurs d'intérêt. Une enquête peut comporter plusieurs questionnaires ou instruments de collecte.

À moins de précisions particulières, dans les paragraphes de cette section du document et dans ceux de la section 6 portant sur les opérations de collecte des données, les termes « questionnaire » et « question » seront utilisés pour désigner les instruments de collecte de toute nature et leurs éléments.

ÉLABORATION DU QUESTIONNAIRE

Principes

- viii) Le questionnaire est conçu en fonction des besoins des utilisateurs de données en matière statistique, des exigences administratives de l'organisation de l'enquête, des exigences du traitement des données et de la nature et des caractéristiques de la population visée par l'enquête.
- ix) Un questionnaire efficace :
 - Impose un faible fardeau de réponse;
 - Est convivial pour les intervieweurs et les répondants;
 - Comporte des questions (ou instructions aux intervieweurs) pertinentes, claires, concises, exemptes d'ambiguïté et permet de recueillir l'information avec efficacité et un minimum d'erreurs tout en facilitant le codage, la saisie et en minimisant les tâches de validation et d'imputation requises à l'issue de la collecte;
 - Permet de recueillir des renseignements quant au déroulement de la collecte lorsqu'il n'y a pas de contact direct avec le répondant (ex. : collecte postale ou collecte utilisant le Web).
- x) Les questions devraient être formulées de telle sorte que le répondant comprend facilement comment y répondre, et qu'il connaît les réponses. Dans le cas d'un instrument se présentant sous la forme d'une série de questions, les instructions aux intervieweurs sur les renseignements à recueillir devraient être formulées de telle sorte que l'intervieweur et le répondant comprennent facilement de quoi il s'agit, et que le répondant soit à même de fournir ces renseignements.
- xi) Le questionnaire devrait être conçu en collaboration avec des experts en questionnaire et des experts en contenu ou ceux qui seront les utilisateurs principaux des données.
- xii) Le début du questionnaire devrait contenir une section qui met le répondant en contexte et assure qu'on est bien en présence de l'unité échantillonnée. Plus de détails sont présentés à l'étape 6.

- xiii) Lorsqu'il y a lieu, le répondant d'une enquête téléphonique devrait être choisi aléatoirement dans le ménage. Le questionnaire doit alors comporter toutes les questions utiles à cette sélection aléatoire. Plus de détails sont présentés à l'étape 6.

CHAPITRE 6 – OPÉRATIONS DE COLLECTE ET DE SAISIE DES DONNÉES

L'incidence des opérations de collecte et de saisie des données sur la qualité de l'information statistique tirée d'une enquête est directe et essentielle, car les données recueillies sont les principaux intrants d'une production statistique d'enquête. Le soin qu'on met à la collecte, la saisie et la codification des données recueillies est donc une condition nécessaire à la qualité d'une enquête.

Les opérations de collecte doivent se faire en ayant le souci constant de faciliter le travail des répondants et des intervieweurs afin de minimiser les risques d'erreurs dans les données recueillies. Le pré-test est un moyen approprié pour s'assurer que cet objectif est atteint. Cela veut notamment dire qu'on facilite la compréhension des questions soumises aux répondants et des instructions pour remplir le questionnaire. Il importe aussi de bien former les intervieweurs pour la collecte des données en leur donnant des instructions claires et précises pour susciter la collaboration des répondants, minimiser la non-réponse à des questions et assurer le respect du plan de sondage. Il faut aussi fournir aux intervieweurs tout le matériel nécessaire pour faciliter leur tâche. Les opérations de collecte doivent être suivies et contrôlées afin de maximiser la qualité des données recueillies. Enfin, les activités de saisie, de validation et de codification des données doivent être judicieusement planifiées et exécutées.

Comme pour la section précédente du document, à moins de précisions particulières, dans les paragraphes qui suivent, les termes « questionnaire » et « question » seront utilisés pour désigner les instruments de collecte de toute nature et leurs éléments (voir la section 5 – Méthode de collecte des données et élaboration des questionnaires et instruments de collecte pour des précisions sur des instruments de collecte n'ayant pas la forme d'un questionnaire).

Comme il s'agit de collecte de données primaires, c'est-à-dire de données recueillies directement « sur le terrain », à même leur source, les opérations de collecte et de saisie de ces données doivent assurer qu'elles sont le plus possible exemptes d'erreurs (ex. : les biais de non-réponse, de réponse, d'autres erreurs de mesure comme les biais de rappel). C'est pourquoi la principale dimension de la qualité en lien avec cette étape est la dimension *fiabilité et objectivité*. Aussi, comme les données recueillies doivent être les plus « fraîches » possible et que le temps nécessaire à la réalisation des opérations de collecte, de saisie et de codification des données est un facteur déterminant de l'échéancier global d'un projet d'enquête, l'*actualité* est une dimension importante à considérer dans la mise en œuvre de cette étape de réalisation.

Définitions

La collecte de données renvoie à tout processus dont l'objectif est d'acquérir ou de faciliter l'acquisition de données. On procède à la collecte en demandant et en obtenant des données pertinentes auprès de ménages, de personnes, d'entreprises ou d'établissements par le truchement d'un mécanisme adéquat (questionnaires ou autres instruments d'enquête et méthode de collecte appropriés).

La saisie des données renvoie à tout processus qui convertit l'information communiquée par un répondant en un format électronique convenant à une utilisation dans des processus subséquents. Parfois, les données sont saisies immédiatement au moment de la collecte (ex. : collecte téléphonique assistée par ordinateur), alors que dans d'autres cas, il faut procéder à une opération *a posteriori* pour saisir les données.

La codification des données renvoie à tout processus qui convertit l'information communiquée par un répondant sous la forme d'une réponse à une question en une information numérique pouvant être analysée. Par exemple, à la codification, on remplace les réponses à une question ouverte saisies par l'intervieweur en catégories identifiées par des codes.

Principes

- i) Les instruments et les stratégies de collecte doivent être élaborés de façon à limiter au minimum le fardeau du répondant.
- ii) Les instruments et les stratégies de collecte doivent être prétestés avant d'amorcer la période de collecte de données.
- iii) La volonté de participation des répondants est étroitement liée à la mise en place de stratégies capables de susciter leur intérêt, de les mettre en confiance et de les inciter à collaborer à l'enquête.
- iv) La stratégie de collecte doit prévoir une procédure pour s'assurer que les unités enquêtées sont bien celles qui sont échantillonnées.
- v) Les opérations de collecte, de saisie et de codification des données doivent inclure des procédures de contrôle de la qualité afin d'assurer la fiabilité et la cohérence des données recueillies.
- vi) Des méthodes de suivis et de contrôle de la qualité doivent être mises en place à toutes les étapes de la collecte (collecte, saisie, codification et validation des données).
- vii) Des procédures pour minimiser la non-réponse doivent être mises en place. Ceci est en fait un des enjeux principaux des opérations de collecte de données.
- viii) En vertu de l'article 25 de la *Loi sur l'Institut de la statistique du Québec*, l'Institut a l'obligation légale de garantir la confidentialité des données transmises par les répondants dans le cadre d'une enquête.
- ix) Des conditions (physiques, matérielles, etc.) optimales de collecte doivent être mises en place afin de créer un environnement propice au travail efficace et rigoureux des intervieweurs.

CHAPITRE 7 – RÉPONSE ET NON-RÉPONSE

La non-réponse influence les résultats de l'enquête. Même lorsqu'on peut supposer que les répondants et les non-répondants sont similaires en ce qui a trait aux variables étudiées et aux variables auxiliaires, la non-réponse fait augmenter la variance à cause de la réduction de la taille d'échantillon. Mais comme dans la plupart des cas, les non-répondants diffèrent des répondants, il en résulte une source additionnelle d'erreur. La non-réponse est alors particulièrement nuisible, puisqu'elle peut causer des biais dans les résultats de l'enquête.

L'étape de réponse et non-réponse est étroitement liée à la dimension *fiabilité et objectivité* de la qualité des données du fait que la non-réponse peut avoir un impact direct sur la qualité des produits statistiques.

Définition

La non-réponse est définie comme le fait que les données demandées ne sont pas obtenues pour certaines des unités échantillonnées. Ce phénomène est présent dans tous les types d'enquêtes, les recensements ou celles basées sur un échantillon, même lorsqu'il s'agit d'enquêtes à caractère obligatoire.

La non-réponse peut être divisée en deux catégories : la non-réponse totale et la non-réponse partielle. La non-réponse totale survient lorsque, pour une unité échantillonnée, les données sont manquantes pour toutes les questions (ou mesures). Celle-ci est en général divisée en trois catégories : les cas « non contactés », les « refus » et les « autres raisons » de non-réponse. Quant à la non-réponse partielle, elle survient lorsque certaines réponses ou parties du questionnaire sont manquantes ou incomplètes.

Pour qu'une unité échantillonnée soit classée comme répondante, le degré de non-réponse partielle ne doit pas excéder un certain seuil au-delà duquel l'unité sera considérée comme non répondante. Ce seuil est à déterminer selon les besoins de l'enquête et peut être influencé par le caractère essentiel de certaines questions.

Principes

- i) Le principe le plus important est de produire des statistiques non biaisées, de la meilleure qualité possible. Pour ce faire, le taux de réponse devrait être élevé et la variance petite. Ces buts peuvent être atteints en portant une attention particulière à la stratégie de collecte, au prétest des instruments de mesure et à l'exécution et la supervision de la collecte de données, ceci étant fait en respectant les contraintes touchant les ressources financières, les ressources humaines et l'échéancier.
- ii) On obtient un bon taux de réponse en s'assurant d'un degré de qualité élevé durant toutes les étapes de planification et de mise en œuvre de l'enquête. Au besoin, on peut invoquer l'article 11 de la *Loi sur l'Institut de la statistique du Québec* pour déclarer l'enquête obligatoire de façon à obtenir le taux de réponse souhaité.
- iii) Une attention particulière devrait être portée à l'organisation de la collecte des données du fait qu'elle constitue le lien principal entre le répondant et l'Institut.

- iv) L'obtention d'un taux de réponse élevé se fait en maintenant les trois catégories de non-réponse (« non contactés », « refus », « autres raisons ») à des niveaux faibles. De façon générale, une organisation qui souhaite réduire la non-réponse à une enquête en fin de période de collecte devrait s'appliquer en priorité aux unités des catégories « non contactés » et « autres raisons », car il est plus facile de convertir ces dernières en unités répondantes. Bien qu'il soit plus difficile de convertir les unités de la catégorie « refus » en répondants, l'utilisation d'approches adaptées (techniques de conversion des refus), combinées à l'habileté de certains intervieweurs, permettront de diminuer cette catégorie de non-réponse. Certaines interventions particulières peuvent également parfois être envisagées pour certains types de non-réponse dans le contexte spécifique d'une enquête obligatoire.
- v) Les utilisateurs de données doivent avoir accès à un rapport précis décrivant la collecte de données, les propriétés des données finales et une description détaillée de la non-réponse accompagnée par des analyses décrivant les effets de la non-réponse.
- vi) L'impact de la non-réponse devrait toujours être réduit au minimum par l'utilisation de méthodes statistiques appropriées d'ajustement de données.
- vii) Le degré des efforts à fournir pour obtenir une réponse d'une unité non répondante ou pour élaborer des techniques de rajustement de la non-réponse est susceptible d'être influencé par des contraintes de budget et de temps, et par leur incidence sur la qualité générale et le risque de biais potentiel.

CHAPITRE 8 – VALIDATION ET IMPUTATION

VALIDATION

La validation vise à détecter les erreurs dans les données (données manquantes, invalides ou incohérentes, unités susceptibles d'avoir fourni des renseignements erronés) afin d'éviter des distorsions significatives dans les résultats de l'enquête. Tout le processus de validation doit donc se faire en ayant en tête les impacts des données individuelles sur les estimations produites; la validation doit assurer que les estimations soient de la meilleure qualité possible.

Toutes les activités réalisées à l'étape de la validation visent d'abord à accroître la dimension *fiabilité et objectivité* de la qualité des données. Cependant, la planification et la réalisation adéquate de la validation peuvent grandement contribuer à augmenter d'autres dimensions de la qualité des données. Puisque cette étape demande en général passablement de temps et de ressources, une bonne planification des travaux de validation peut notamment améliorer l'*actualité* des données et ainsi éviter des pertes importantes de qualité pouvant compromettre la *pertinence* de certaines données.

Définitions

La validation (ou vérification) des données englobe l'ensemble des activités visant à contrôler la qualité des données et à les rendre les plus justes possible en tant que valeurs individuelles et mutuellement compatibles avec les autres données recueillies. À cette étape, certaines corrections sont apportées aux données, des corrections manuelles par exemple.

La validation commence à un niveau très fin, pour chaque réponse fournie par chaque unité (micro-validation) et se poursuit jusqu'au niveau le plus agrégé lorsqu'on vérifie une estimation par exemple (macrovalidation). La compatibilité entre les niveaux est aussi vérifiée (cohérence).

La validation s'effectue à toutes les étapes d'acquisition des données dans une enquête : de la planification de la collecte (ex. : insertion de validations automatiques dans un questionnaire informatisé) en passant par les vérifications des intervieweurs en cours de collecte et les rappels de validation qui sont alors faits, à la constitution d'un fichier final en préparation de l'estimation jusqu'à l'analyse des données.

Principes

- i) La validation vise trois objectifs (Granquist, 1984) : servir de fondement à l'amélioration future de l'instrument d'enquête, fournir des renseignements sur la qualité des données et « nettoyer » les données. Ce troisième objectif occulte souvent les deux autres qui n'en sont pas moins fondamentaux. En effet, lorsqu'elle sert les deux premiers objectifs, la validation fournit des mesures de qualité de l'enquête et permet de dégager des avenues d'amélioration pour les répétitions futures de l'enquête si elle est récurrente ou pour d'autres enquêtes de même nature.

- ii) Les données devraient être validées et corrigées pour leur importance globale, c'est-à-dire pour leur impact sur les estimations à produire. Des efforts considérables devraient être mis dans l'analyse et la prévention des problèmes majeurs, alors que les problèmes mineurs devraient être résolus à l'aide de solutions automatiques.
- iii) La validation est un processus coûteux et complexe qui doit être efficace pour engendrer la qualité désirée; il devrait être soigneusement planifié et contrôlé. En particulier, il importe d'avoir le point de vue du chargé de projet sur les données qui sont les plus importantes à valider (par exemple, celles qui servent à estimer les paramètres principaux de l'enquête) et de bien définir le type de validations à effectuer et le meilleur moment pour les faire. À l'intérieur d'une équipe de réalisation de projet, les responsabilités de chacun à l'égard de la validation doivent être adéquatement définies afin d'éviter les doublons.
- iv) On doit se méfier de la survérification qui survient lorsque la validation est laborieuse et qu'elle a une incidence négligeable sur les résultats. Elle peut entraîner de graves biais, en particulier lorsque des données sont remplacées par d'autres obtenues des modèles implicites aux contrôles effectués.

IMPUTATION

L'imputation peut être utilisée pour remplacer des données manquantes ou corriger des données qui n'ont pas pu l'être par les rappels de validation auprès des répondants, lors de validations automatiques ou par des corrections manuelles pouvant être déduites de l'étape de validation.

Il importe de choisir une méthode d'imputation appropriée, car certaines méthodes ne préservent pas les liens entre les variables (cohérence interne) et peuvent engendrer des distorsions pour ce qui est des distributions sous-jacentes. Par conséquent, il faut tenir compte de l'imputation au cours de la production des estimations et des estimations de variances connexes.

Les opérations effectuées lors de cette étape de réalisation permettent principalement d'accroître la dimension *fiabilité et objectivité* des données. La dimension *comparabilité* est également souvent au cœur des préoccupations au moment de réaliser la validation et l'imputation.

Définitions

L'imputation est le processus par lequel une valeur manquante, invalide ou incohérente pour une variable est remplacée par une autre valeur. Cette valeur, appelée valeur imputée, se doit d'être la plus près possible de la valeur réelle, mais inconnue, de la variable, et respecter la plausibilité et la cohérence interne. L'imputation est généralement employée pour corriger la non-réponse partielle, mais peut aussi être appliquée pour la non-réponse totale. On parle alors d'imputation de masse.

Il existe deux types de méthodes d'imputation : l'imputation par modèle et l'imputation par enregistrement donneur. Dans l'imputation par modèle, la valeur imputée est obtenue au moyen d'un modèle ajusté aux données fournies par les répondants à l'enquête. Dans l'imputation par enregistrement donneur, la valeur imputée est empruntée d'un autre répondant.

Les méthodes d'imputation peuvent par la suite être classifiées en méthodes déterministes et en méthodes stochastiques. L'imputation logique qui implique que la valeur imputée est déduite à l'aide d'un lien entre des valeurs connues, compte parmi les méthodes déterministes. Il en est de même de l'imputation historique, de l'imputation par la moyenne, par le ratio, par la régression et par le plus proche voisin. D'autres méthodes sont dites stochastiques parce qu'elles comprennent une composante aléatoire. La méthode du *hot deck aléatoire* fait partie de cette catégorie, tout comme la méthode du plus proche voisin avec sélection aléatoire, la méthode d'imputation par le ratio avec résidus aléatoires ou toute autre méthode avec résidus aléatoires.

Enfin, les méthodes d'imputation peuvent être divisées en méthodes d'imputation simples ou multiples. Dans l'imputation simple, la valeur à corriger est remplacée par une seule valeur, alors que dans l'imputation multiple, on la remplace par plusieurs valeurs.

Principes

- v) Les bons processus d'imputation sont informatisés, objectifs, reproductibles et efficaces.
- vi) De bonnes données auxiliaires, pertinentes et de qualité, sont nécessaires pour l'imputation.
- vii) Une imputation efficace permet d'assurer que les enregistrements imputés respectent la cohérence interne. Autrement dit, les valeurs imputées pour un enregistrement ne doivent pas être contradictoires entre elles et doivent être cohérentes avec les valeurs des réponses au questionnaire.
- viii) Il est recommandé d'apporter des changements à un nombre minimal de champs afin de s'assurer que l'enregistrement complet rencontre chacune des règles de validation. (Voir Fellegi et Holt, 1976.)
- ix) L'imputation vise à réduire le biais introduit par le fait que toutes les valeurs souhaitées n'ont pas été observées. Toutefois, cela se fait au prix d'une augmentation de la variance.
- x) Comme la validation, l'imputation fournit des pistes à suivre pour l'évaluation de la qualité des données d'une enquête. Par exemple, le taux d'imputation pour une question est un indicateur de qualité des estimations qui en découlent.

CHAPITRE 9 – PONDÉRATION

La pondération a pour but de permettre l'inférence adéquate des résultats observés à la population visée. Le poids accordé (variable de pondération) à une unité enquêtée reflète en quelque sorte le nombre d'unités de la population visée que celle-ci « représente ».

Dans le cas d'une enquête basée sur un échantillon, la variable de pondération est fonction du plan de sondage utilisé et, le cas échéant, des corrections apportées pour améliorer la précision ou pour réduire les biais causés par la non-réponse, les erreurs de couverture ou les erreurs de mesure. Quant au cas des enquêtes par recensement, qui n'échappent pas au phénomène de non-réponse, aux erreurs de mesure et dont la base de sondage peut également comporter des problèmes de couverture de la population visée, la variable de pondération est fonction des ajustements faits pour réduire les effets néfastes de ces écueils.

Il va sans dire que la pondération contribue à assurer la dimension *fiabilité et objectivité* des statistiques produites. Cependant, la pondération a aussi à voir avec la dimension *comparabilité* de la qualité. En particulier, elle permet d'assurer la cohérence des estimations produites à différents cycles d'une enquête récurrente ou encore de celles tirées de différentes enquêtes portant sur la même population visée.

Définitions

S'il s'agit d'une enquête par échantillon, la forme la plus simple de la variable de pondération est définie par l'inverse de la probabilité de sélection des unités dans l'échantillon. Ce poids, appelé *poids d'échantillonnage*, représente le nombre de fois où l'on doit reproduire chaque unité échantillonnée afin de représenter la population visée complète. Pour les recensements, la probabilité de sélection des unités est égale à 1, pour toutes les unités. Le poids d'échantillonnage est donc aussi égal à 1 pour chaque unité.

Pour contrebalancer les effets néfastes de la non-réponse, tant pour les enquêtes par échantillon que pour les recensements, il est fréquent d'utiliser la repondération. Cette procédure consiste à rajuster les poids d'échantillonnage des unités répondantes à l'aide de facteurs de correction afin qu'elles représentent l'ensemble de la population enquêtée (ou échantillonnée). Le principe de base du calcul de ces facteurs utilise l'inverse de la probabilité qu'ont les unités de répondre.

L'information auxiliaire peut aussi rendre possible l'amélioration du processus de pondération. Le calage aux marges, dont la forme la plus simple est appelée *poststratification* ou simplement *ajustement à la population*, permet de rajuster les poids d'échantillonnage au moyen de multiplicateurs appelés les facteurs de calage. Ces facteurs font correspondre les estimations aux totaux de population connus.

Les poids construits pour faire une inférence adéquate à la population visée peuvent se présenter sous la forme de *poids* dits *populationnels* ou sous celle de *poids échantillonnaires* (ou *poids à moyenne un*). La somme des poids populationnels correspond à la taille de la population visée; celle des poids échantillonnaires correspond à la taille de l'échantillon des répondants.

Principes

- i) En principe, les probabilités de sélection utilisées dans la procédure d'échantillonnage devraient être connues et avoir une valeur positive pour chacune des unités de la population enquêtée (ou échantillonnée).
- ii) Il est recommandé de baser la repondération sur des variables de la base de sondage associées à la fois à la non-réponse et aux variables étudiées par l'enquête. La recherche de l'équilibre entre la réduction désirée du biais dû à la non-réponse et l'augmentation de la variabilité des estimations associée à une plus grande variance des poids est alors de mise.
- iii) Les techniques de calage aux marges et de repondération permettent d'améliorer les poids d'échantillonnage et de réduire les biais possibles causés par les erreurs de base de sondage, la non-réponse et les erreurs de mesure.
- iv) La recherche et l'utilisation de données auxiliaires dans la construction des poids, la repondération pour la non-réponse et le calage aux marges font partie des bonnes pratiques statistiques.

CHAPITRE 10 – ESTIMATION

L'objectif principal poursuivi lors de la réalisation d'une enquête statistique est de pouvoir établir des constats clairs pour l'ensemble de la population visée par l'enquête, et ce, même si une portion seulement de la population a été consultée. L'estimation permet de faire ce passage adéquatement.

Cette étape de réalisation d'une enquête, qui est liée à la production des résultats, touche directement la dimension *fiabilité et objectivité*. La *comparabilité* est une dimension de la qualité qui est également visée lors de l'étape d'estimation, en particulier si on est dans un contexte d'enquête récurrente.

Définitions

Les principaux paramètres d'intérêt lorsqu'on fait de la statistique descriptive sont habituellement les suivants : totaux, moyennes, proportions, ratios et percentiles. Ceux-ci sont généralement inconnus pour l'ensemble de la population, ce qui implique qu'il est nécessaire d'en calculer une approximation. L'estimation est l'étape importante où l'on procède à l'approximation des paramètres inconnus de la population avec le plus de fiabilité possible au moyen d'une partie seulement de la population, celle comprise dans l'échantillon.

Même dans le cas d'enquête par recensement, lorsqu'il y a présence de non-réponse, il est nécessaire de procéder à l'étape d'estimation; l'échantillon est alors constitué des répondants seulement.

La méthode ou l'algorithme de calcul utilisé pour produire l'estimation est appelée « estimateur ». Cet estimateur devrait prendre en compte le recours à la repondération, au calage aux marges et à l'utilisation de données auxiliaires. La fiabilité d'une estimation est évaluée en calculant l'erreur-type ou le coefficient de variation de l'estimateur utilisé. L'intervalle de confiance pour un paramètre peut être calculé en utilisant soit l'erreur-type de l'estimation, soit la distribution statistique de l'estimateur choisi.

Principes

- i) De façon générale, il importe d'utiliser une pondération lors de l'étape d'estimation afin de pouvoir inférer les résultats obtenus à la population visée.
- ii) On devrait choisir des estimateurs qui produisent le plus faible biais et la plus faible variance d'échantillonnage comme ceux du type Horvitz-Thompson, les estimateurs par la régression généralisée et les estimateurs par calage.
- iii) L'utilisation de données auxiliaires fortement corrélées aux variables à estimer permet de réduire la variance des estimations.
- iv) L'estimateur Horvitz-Thompson est l'estimateur le plus simple à utiliser lorsqu'aucune donnée auxiliaire n'est disponible. Cet estimateur pondère les données à l'aide de l'inverse de la probabilité de sélection des unités échantillonnées, qu'on appelle poids d'échantillonnage.
- v) Il est important de tenir compte de la non-réponse lors de l'étape d'estimation. Notamment, lorsque des données auxiliaires sont disponibles, il est souhaitable d'utiliser cette information pour procéder à une repondération afin de corriger les biais dus à la non-réponse.

- vi) Il peut également être pertinent d'utiliser un estimateur basé sur un calage aux marges ou sur une poststratification (ajustement à la population). Ces procédures rajustent les poids d'échantillonnage afin de faire correspondre les estimations aux totaux de population connus.
- vii) En général, lorsqu'on travaille sur des petits sous-groupes de la population, les estimateurs usuels basés sur le plan d'échantillonnage ne sont pas suffisamment précis. Ainsi, des estimateurs spécifiques pour des petits domaines devraient être utilisés. Des techniques statistiques ou des estimateurs « synthétiques » peuvent également être utilisés pour produire des estimations sur des éléments de données non recueillis directement.
- viii) L'information relative à la précision des estimations produites est essentielle pour évaluer leur fiabilité et devrait être systématiquement rapportée aux utilisateurs.

CHAPITRE 11 – DÉSAISONNALISATION ET ESTIMATION DE LA TENDANCE-CYCLE

La réalisation d'une enquête à des intervalles de temps réguliers (hebdomadaire, mensuel, trimestriel ou annuel) mène à la constitution d'une série chronologique, c'est-à-dire un ensemble d'estimations séquentielles produit par un processus stochastique. Ces séries peuvent être décomposées en plusieurs composantes : la tendance, le cycle, la saisonnalité, les effets de calendrier et la partie irrégulière. L'estimation et l'étude de ces différentes composantes permettent de mieux comprendre les fluctuations et l'évolution des séries chronologiques.

Pour plusieurs séries, les mouvements saisonniers de grande ampleur rendent difficile l'identification des autres variations. Il est alors souhaitable d'éliminer les variations produites par des phénomènes qui reviennent systématiquement à la même période de l'année (saisonnalité), et de mettre en lumière les mouvements sous-jacents à moyen et long terme (tendance et cycle) ainsi que les événements spéciaux (irréguliers). La désaisonnalisation consiste justement à estimer des facteurs saisonniers et à les appliquer à une série chronologique afin d'en supprimer les variations saisonnières. Cette opération et la prise en compte des effets de calendrier rendent les mois comparables entre eux. De plus, elles font ressortir les mouvements intéressants et critiques dans les séries comme la fin et le début des cycles économiques.

À l'étape de désaisonnalisation et d'estimation de la tendance-cycle, le travail de qualité porte essentiellement sur les dimensions *fiabilité* et *objectivité* et *comparabilité*.

Définitions

Tendance et cycle

Ces deux composantes représentent l'évolution fondamentale de la série. La tendance représente un mouvement sous-jacent à long terme provenant de phénomènes généraux de croissance ou de décroissance (population, activité économique). Le cycle est une oscillation extra annuelle de longueur variable provenant d'une variation conjoncturelle propre aux fluctuations économiques (expansion, récession).

En pratique, puisque le nombre d'observations dans une série chronologique est insuffisant pour distinguer la tendance du cycle, une seule série chronologique combinant la tendance et le cycle est produite. Cette série est appelée tendance-cycle.

Saisonnalité

La composante saisonnière est constituée des fluctuations intra annuelles qui se produisent plus ou moins régulièrement d'une année à l'autre. Ces fluctuations sont principalement causées par des facteurs climatiques (la température, la neige) et institutionnels (ex. : la fête de Noël, les vacances d'été).

Effets de calendrier

Les effets de calendrier sont provoqués par les jours ouvrables et les fêtes mobiles. L'effet des jours ouvrables apparaît lorsque l'activité mesurée varie en fonction du jour de la semaine. Puisque le nombre de fois qu'un jour de la semaine est présent dans un mois est variable d'une année à l'autre, des fluctuations indésirables seront introduites dans la série. L'effet des fêtes mobiles apparaît lorsqu'une fête passe d'un mois à l'autre (ex. : la fête de Pâques).

Irréguliers

Les irréguliers sont composés par les fluctuations inexplicables par les autres composantes de la série (la tendance-cycle, la saisonnalité et les effets de calendrier). Cette composante représente des événements accidentels ou fortuits passagers (ex. : une grève, un hiver exceptionnellement froid) et des erreurs de mesure du phénomène.

Série désaisonnalisée

Une série désaisonnalisée comprend la composante tendance-cycle ainsi que la composante irrégulière. D'une manière équivalente, la série désaisonnalisée peut être vue comme la série originale à laquelle la composante saisonnière et les effets de calendrier ont été retranchés.

Principes

- i) Puisque le nombre d'observations dans une série chronologique est fréquemment insuffisant pour distinguer la tendance du cycle, on produit en général une seule série chronologique combinant la tendance et le cycle appelée tendance-cycle.
- ii) Pour produire des séries désaisonnalisées, on doit estimer la composante saisonnière et les effets de calendriers et ensuite les enlever, de sorte qu'il reste seulement une tendance-cycle et la variation irrégulière.
- iii) Avant qu'une série soit désaisonnalisée, elle doit être ajustée pour les jours ouvrables et les congés fériés et pour n'importe quelle déviance observée causée par des actions industrielles ou d'autres circonstances exceptionnelles.
- iv) Il convient de désaisonnaliser la série chronologique uniquement lorsqu'elle est manifestement influencée par des facteurs saisonniers, et que son caractère saisonnier peut être décelé avec certitude.
- v) Une méthode de désaisonnalisation efficace produit une série dénuée de saisonnalité, qui est beaucoup plus lisse que la série originale.
- vi) Les révisions apportées aux estimations désaisonnalisées devraient être minimales, lorsque d'autres points de données sont ajoutés à la fin de la série. Certaines méthodes (X11ARMMI, X11ARMMI/88 et X-12-ARIMA) peuvent être utilisées précisément dans le but de réduire l'importance des révisions.
- vii) La fréquence des révisions doit être réduite au minimum. Bien qu'il y ait une révision des valeurs passées de la série chaque fois qu'une donnée est ajoutée, les révisions doivent être effectuées uniquement si elles améliorent les estimations, c'est-à-dire si les estimations révisées sont beaucoup plus près des estimations définitives. Il est de mise qu'un organisme statistique établisse un calendrier de révision de façon à mieux informer les utilisateurs des modifications à venir.

- viii) Chaque fois que des données désaisonnalisées se rapportant à la même activité économique sont publiées, les méthodes de désaisonnalisation employées doivent être coordonnées et tout doit être mis en œuvre afin que les séries connexes soient traitées avec cohérence.
- ix) Lorsqu'on produit des estimations de la tendance-cycle afin d'apporter un complément à la série désaisonnée, il ne faut pas oublier qu'elles visent à rendre compte du stade du cycle économique et, à ce titre, doivent refléter avec exactitude la direction du mouvement, l'amplitude du cycle et, tout particulièrement, les dates de point de retournement.
- x) Les estimations de la tendance-cycle doivent être cohérentes avec les estimations des données désaisonnalisées publiées. Par exemple, si ces dernières estimations deviennent définitives dans la base de données après trois mois, l'estimateur de la tendance-cycle doit être appliqué aux estimations désaisonnalisées figurant dans la base et dans la publication.

CHAPITRE 12 – ÉVALUATION DE LA QUALITÉ

À l'étape d'évaluation de la qualité, on s'intéresse autant à la qualité d'ensemble de l'enquête, qu'on appellera « macroévaluation », qu'à la qualité des données qui en découle, désignée par le terme « microévaluation ».

Il importe de rappeler que la qualité des données est l'amalgame de toutes les caractéristiques qui les rendent aptes à satisfaire aux besoins explicitement énoncés ou implicites déterminés lors de la planification de l'enquête. Cette étape de réalisation d'une enquête est donc étroitement liée à la première, « Objectifs, utilisations, utilisateurs (démarrage et planification de l'enquête) ».

Les renseignements obtenus lors de l'évaluation de la qualité doivent d'abord permettre aux utilisateurs de mieux étayer leur interprétation des résultats d'enquête. Ils constituent aussi une source d'information précieuse pour l'équipe de réalisation afin d'améliorer l'enquête ou des enquêtes semblables. Enfin, l'évaluation de la qualité d'une enquête peut aider l'Institut à améliorer un programme d'enquêtes. Le personnel de l'Institut travaillant à sa fonction d'évaluation de programme prendra notamment en compte l'évaluation de la qualité de plusieurs enquêtes appartenant à un même programme pour tirer des conclusions sur la qualité du programme. Les dimensions de la qualité seront alors toutes considérées, mais on adoptera en ce cas la perspective plus large d'un ensemble de projets pour faire l'évaluation. Les travaux d'évaluation de la qualité d'une enquête et ceux réalisés dans le cadre d'une évaluation de programme sont en ce sens complémentaires.

Les données doivent au moins faire l'objet d'une évaluation de leur qualité avant leur diffusion officielle afin de prévenir la circulation de données erronées et des interprétations abusives. Au besoin, des études particulières sur certaines sources d'erreurs peuvent être entreprises. Elles peuvent éventuellement faire suite à la diffusion officielle des données si ces sources d'erreurs ne mettent pas en péril leur usage. On entreprendra par exemple parfois « une révision des processus » en profondeur pour une enquête récurrente dont la qualité a été jugée adéquate au moment de la diffusion, mais pour laquelle on désire dégager des possibilités d'amélioration. La *Politique sur l'évaluation de programmes et sur la révision de processus* (à venir) fait état d'évaluation de la qualité de cette nature. Cela n'est cependant pas l'objet du présent document.

On comprendra qu'à cette étape, le spectre presque entier des dimensions de la qualité est une préoccupation. L'évaluation portant tantôt sur la dimension *pertinence*, tantôt sur la *fiabilité* et l'*objectivité*, la *comparabilité* ou sur l'*actualité* de l'enquête en général et des données qui en découlent ensuite. Enfin, le but premier de l'évaluation de la qualité est d'améliorer l'*intelligibilité* de l'enquête et des données au bénéfice de tous les utilisateurs potentiels.

Définitions

L'évaluation de la qualité de l'enquête (macroévaluation) est la mesure de l'atteinte des objectifs de qualité et de l'adéquation des données qu'on en tire par rapport à l'usage visé. Elle désigne le processus visant à déterminer si le produit fini satisfait aux objectifs initiaux de l'activité statistique. Elle est habituellement faite à partir d'une description écrite des procédures et processus de mise en application de l'enquête; elle sert notamment à vérifier si les activités réalisées l'ont été de façon efficace pour répondre aux besoins déterminés.

L'évaluation de la qualité des données produites (microévaluation) se fait, entre autres, en examinant les différentes sources d'erreur dont elles peuvent être entachées. Il peut notamment s'agir d'erreurs dues à l'échantillonnage ou d'erreurs non dues à l'échantillonnage. Ces dernières regroupent les erreurs de couverture et de classement, les erreurs attribuables à la non-réponse, les erreurs de mesure et les erreurs de traitement incluant les erreurs de modèles. Par ailleurs, la détermination de tous les facteurs pouvant limiter l'utilisation de ces données fait également partie de la microévaluation.

Principes

- i) Il importe que les utilisateurs puissent déterminer dans quelle mesure les erreurs présentes dans les données sont des entraves à leur usage. Il appartient à l'Institut d'effectuer les évaluations nécessaires de la qualité des données et de communiquer aux utilisateurs les résultats de ces évaluations sous une forme pratique.
- ii) L'Institut peut se servir des évaluations de la qualité pour améliorer la qualité de la prochaine occurrence d'une enquête et d'autres enquêtes similaires.
- iii) L'actualité des résultats de l'évaluation de la qualité des données est tout aussi importante que l'actualité des données. Idéalement, les résultats de l'évaluation de la qualité doivent être suffisamment valables et actuels pour améliorer les données produites. Lorsque cela s'avère impossible, les résultats de l'évaluation doivent à tout le moins être suffisamment actuels pour aider les utilisateurs à interpréter et à analyser correctement les données. Dans le cas d'une enquête récurrente, les évaluations doivent être suffisamment actuelles pour permettre au personnel de concevoir la prochaine répétition de l'enquête.
- iv) Des ressources nécessaires devraient être prévues pour s'assurer que les critères de qualité visés sont atteints. L'utilisation de ressources pour atteindre des niveaux de qualité supérieurs aux critères fixés n'est pas une bonne utilisation des ressources.

CHAPITRE 13 – CONTRÔLE DE DIVULGATION

En vertu de la loi, l'Institut de la statistique du Québec est tenu de protéger la confidentialité des renseignements fournis par les répondants. Plus les répondants auront confiance en l'Institut pour protéger les renseignements qu'ils lui fournissent, plus ils seront enclins à participer à ses enquêtes, et plus la qualité de l'information statistique diffusée augmentera. La satisfaction des répondants en termes de protection des renseignements personnels passe obligatoirement par l'étape du contrôle de la divulgation.

La dimension centrale de la qualité mise en cause à cette étape est l'*accessibilité*, car le contrôle de la divulgation a un impact direct sur la disponibilité de l'information statistique à diffuser. On peut par contre y intégrer également la dimension *pertinence*, car toutes ces opérations doivent s'effectuer avec le souci du respect des attentes et des objectifs des utilisateurs.

Définitions

Le contrôle de la divulgation désigne les mesures visant à protéger les renseignements personnels recueillis par l'Institut, dans le respect des exigences en matière de confidentialité. La protection des renseignements s'effectue sur deux plans : les mesures de sécurité durant le processus de collecte de données et le traitement statistique, et le contrôle de la divulgation à effectuer pour permettre la diffusion des renseignements statistiques. À ce sujet, on distingue deux types de renseignements statistiques : les fichiers de microdonnées et les tableaux de résultats.

Le mode d'accès aux fichiers de microdonnées (Centre d'accès aux données, accès à distance, locaux d'un organisme partenaire) et le type de fichiers à communiquer (fichier maître, fichier à grande diffusion, fichier aux fins d'analyse ou de recherche externe) sont des éléments qui influent sur les mesures de contrôle de divulgation à appliquer aux fichiers de microdonnées préalablement à leur communication. Quant aux tableaux, le type de statistiques produites (ex. : fréquence, total) et le type de fichier de microdonnées servant à les produire (fichier maître, fichier à grande diffusion, fichier aux fins d'analyse ou de recherche externe) sont les principaux éléments qui influent sur les mesures de contrôle de divulgation à leur appliquer avant leur diffusion.

En général, les mesures de contrôle de la divulgation à appliquer avant la communication des renseignements statistiques ont une incidence défavorable sur la quantité d'information qui est rendue disponible. Le contrôle de la divulgation doit donc assurer la confidentialité des renseignements tout en préservant le plus possible l'utilité de l'information produite. Ainsi, l'équilibre entre la confidentialité et la disponibilité des renseignements statistiques est un concept de haute importance dans l'étude du contrôle de la divulgation.

La protection des renseignements personnels représente un principe central de la production des statistiques officielles, et c'est un élément constituant de l'image externe des autorités publiques.

Principes

- i) Les principes qui sous-tendent les activités de contrôle de la divulgation sont régis par les dispositions de la *Loi sur l'Institut de la statistique du Québec*. Celles-ci sont extrêmement rigoureuses en matière de confidentialité. En effet, l'article 25 de la loi stipule que :

« Le directeur général, les fonctionnaires et toute autre personne dont les services sont utilisés par le directeur général dans l'exercice de ses fonctions ne peuvent révéler ni faire révéler, par quelque moyen que ce soit, des renseignements obtenus en vertu de la présente loi si ces révélations permettent de rattacher un renseignement à une personne, à une entreprise, à un organisme ou à une association en particulier. »

Cependant, la loi prévoit que certains types de renseignements confidentiels peuvent être diffusés à la discrétion du directeur général. Par exemple, malgré l'article 25 :

- un renseignement peut être révélé lorsqu'un consentement écrit préalable a été obtenu de la personne, de l'entreprise, de l'organisme ou de l'association concernée;
- le directeur général peut autoriser la communication d'un renseignement sous forme d'un index ou d'une liste de personnes morales, d'entreprises, d'associations ou d'établissements;
- le directeur général peut également autoriser la communication de renseignements qui ont un caractère public en vertu d'une loi.

Le cadre normatif de l'Institut comporte les politiques, procédures, lignes directrices, directives, normes, etc., en matière de confidentialité auxquelles les employés doivent se référer.

- ii) L'objectif premier de l'application de la loi et du cadre normatif qui en découle consiste à s'assurer qu'aucun renseignement confidentiel ne puisse être révélé, ni même inféré dans une fourchette restreinte, à quiconque non autorisé, et cela, à aucune étape de la production statistique incluant la diffusion.
- iii) Les renseignements confidentiels collectés pour la production de statistiques officielles ne devraient être communiqués à des tiers sous forme de fichiers de microdonnées que pour des besoins de recherche statistique, sous une forme qui ne permette pas d'identifier directement ou indirectement un individu ou une entreprise, et ce, conformément au cadre normatif. Les méthodes de contrôle de divulgation sont variées et dépendent du contexte de communication et du produit communiqué. Une politique a été établie pour encadrer la communication de fichiers de microdonnées à des tiers (*Politique d'accès aux microdonnées d'enquêtes individus ou ménages*).

De plus, une autre politique, accompagnée de lignes directrices, (*Politique relative à la confidentialité des tableaux de résultats pour diffusion*) vise la diffusion de tableaux.

Ces politiques devraient être appliquées avant toute communication de fichiers de microdonnées ou diffusion de tableaux de résultats.

- iv) Il est primordial d'informer les utilisateurs potentiels des données de l'impact possible des mesures de contrôle de divulgation sur les produits diffusés, et ce, dès le début d'un projet.
- v) Il faut s'assurer que la collecte et la transmission des données se font de façon sécurisée.
- vi) Il faut aussi s'assurer de la sécurité physique des lieux où se trouvent les données.

- vii) Toute personne accédant à des données confidentielles doit signer une déclaration de discrétion, soit à titre d'employé de l'Institut, soit à titre de chercheur qui exploite un fichier de microdonnées, c'est-à-dire à titre de personne dont les services sont retenus par l'Institut. Dans le cas d'un chercheur, les droits devraient être accordés pour une période définie et les participants à la recherche clairement identifiés.
- viii) Les objectifs de recherche des utilisateurs doivent être considérés lors de l'application des méthodes de contrôle de divulgation. À titre d'exemple, il peut être préférable de regrouper des valeurs plutôt que de supprimer des données.
- ix) Les éléments du cadre normatif en matière de confidentialité des données devraient être révisés périodiquement pour tenir compte des nouveaux développements et des nouvelles technologies.
- x) Ne communiquer des fichiers de microdonnées qu'à des fins statistiques seulement. Ainsi, toute demande d'accès à des fichiers de microdonnées devant servir à la prise de décision administrative ou à d'autres activités similaires devrait être refusée.
- xi) Il ne faut pas oublier que, pour un organisme statistique, l'apparence d'une divulgation peut parfois être aussi néfaste qu'un cas réel de divulgation. Ainsi, la façon dont le public perçoit la vigilance avec laquelle l'Institut protège la confidentialité de ses données est à tout le moins aussi importante que les mesures réelles prises pour empêcher la divulgation des données des répondants.

CHAPITRE 14 – ANALYSE ET PRÉSENTATION DES DONNÉES

Dans la réalisation d'une enquête, l'étape de l'analyse et de la présentation des données couvre des activités variées. La qualité des processus mis en place à cette étape influe grandement sur la qualité de l'ensemble de l'enquête.

La plupart des dimensions de la qualité sont touchées par la manière dont on procède lors de l'analyse des données et de la présentation des résultats de cette analyse, en particulier la *fiabilité* et l'*objectivité*, la *comparabilité*, l'*intelligibilité* et l'*accessibilité*. L'analyse des données et la présentation des résultats permettent d'établir un lien direct avec les utilisateurs. La qualité de l'analyse et une méthode de présentation et de communication efficace des résultats ont des répercussions substantielles sur l'utilité de l'enquête dans son ensemble, donc sur sa *pertinence* au sens large.

Définitions

L'analyse des données désigne le processus de transformation des données brutes en information utilisable, par exemple, sous la forme d'un recueil de tableaux statistiques, d'un rapport descriptif ou d'un article publiable. Les principales étapes du processus d'analyse consistent à identifier un problème, à poser des questions éloquentes, à étayer les réponses aux questions par l'examen et l'interprétation des données ainsi qu'à communiquer le message aux lecteurs. L'analyse, la synthèse des résultats et l'interprétation de leur signification visent à fournir des réponses claires aux questions qui ont motivé l'enquête.

L'analyse consiste habituellement dans l'examen de tableaux d'estimations ou de figures de diverses mesures statistiques (totaux, moyennes, proportions, ratios, percentiles, etc.) et dans l'examen de tests statistiques usuels. Elle inclut, à l'occasion, des analyses plus poussées comme les analyses multivariées. L'inférence statistique est de mise lorsqu'on teste des hypothèses ou qu'on étudie des liens entre les caractéristiques.

L'analyse peut porter sur deux types principaux de résultats : les résultats descriptifs (résultats portant sur la population finie étudiée au moment de la collecte des données) et les résultats analytiques (résultats portant sur une population qui s'étend souvent au-delà de la population visée par l'enquête appelée superpopulation). En outre, l'analyse peut porter sur les résultats de l'enquête seulement ou elle peut comprendre une comparaison entre les résultats de l'enquête et ceux d'autres enquêtes ou sources de données.

L'analyse des données joue un rôle important dans le processus d'élaboration et de révision des enquêtes. Les premières analyses peuvent également servir à mettre en évidence des problèmes dans les données de base ayant échappé à la validation. L'analyse permet de cerner les problèmes liés à la qualité des données et de proposer des améliorations à apporter au processus d'enquête. Elle est essentielle pour comprendre les résultats des études pilotes et des éditions antérieures des enquêtes, pour concevoir les enquêtes ou pour planifier de nouvelles activités statistiques, pour cerner les lacunes en matière de données et formuler des objectifs relativement à la qualité.

La présentation des résultats de l'analyse des données est une étape très importante : elle représente l'aboutissement des étapes précédentes. Elle montre l'utilité des données et comment elles peuvent être exploitées plus largement. L'information statistique brute que sont les microdonnées statistiques analysées peut être illustrée par deux méthodes : par une présentation textuelle ou à l'aide de tableaux ou de figures.

Principes

Pour l'analyse

- i) L'analyse est le principal moyen permettant d'obtenir de l'information à partir des données.
- ii) Il faut déterminer le niveau d'analyse souhaité (tableaux de données, analyse détaillée ou synthèse) de façon à répondre de la manière la plus complète possible aux besoins des utilisateurs.
- iii) Pour être efficace, l'analyste doit connaître le public cible, les enjeux et les problématiques lorsqu'il détermine les sujets et les façons convenables de présenter les résultats.
- iv) L'étude du contexte de l'analyse permet à l'analyste de choisir les sources de données et les méthodes statistiques adéquates.
- v) Il faut tenir compte du plan d'échantillonnage lors de l'analyse des données afin de faire une inférence adéquate à la population visée.

Pour la présentation des résultats

- vi) Les lignes directrices du *Cahier des normes en matière d'édition, d'impression et d'identification visuelle* de l'Institut doivent être suivies. Ces normes sont des pratiques essentielles à une présentation soignée des manuscrits et à un aspect visuel uniforme des publications de l'Institut.
- vii) Il importe de bien doser le rôle du texte, des tableaux et des figures dans la présentation.
- viii) La conception du texte et de la présentation des tableaux et des figures doit s'adapter au public et au moyen de diffusion choisi (rapport spécialisé, bulletin statistique pour un large public, diaporama, etc.).
- ix) La présentation générale doit contribuer à la clarté des données contenues dans les tableaux et prévenir les erreurs d'interprétation.
- x) Tous les éléments visuels qui illustrent une publication peuvent, malgré leur diversité, être appelés « figures ». Ils comprennent des données essentielles qu'un texte ne saurait rendre de façon aussi simple, claire et efficace.
- xi) Les figures doivent fournir un portrait clair, précis, fidèle et objectif de la réalité du phénomène étudié. C'est la responsabilité de l'analyste de s'assurer que la figure choisie possède ces qualités.
- xii) Plusieurs formes de figures existent. Le type de figure utilisé pour l'analyse (diagramme à secteurs, diagramme à colonnes, diagramme à barres, graphique linéaire) dépend des données à représenter et du message à transmettre.
- xiii) Toutes les conclusions présentées dans une étude analytique doivent être appuyées par les données soumises à l'analyse statistique.

CHAPITRE 15 – DIFFUSION DES DONNÉES

La diffusion consiste à rendre publics les produits statistiques, à les mettre à la disposition des utilisateurs par différents moyens, qu'ils soient électroniques ou plus traditionnels.

Toutes les dimensions de la qualité entrent en ligne de compte lors de la diffusion, puisque les utilisateurs doivent obtenir une information qui est *pertinente*, *fiable* et *objective*, *comparable* avec d'autres renseignements similaires, *actuelle* et *intelligible*. Quant à la dimension d'*accessibilité*, elle est implicite lorsqu'il est question de diffusion.

Définitions

La diffusion consiste à mettre à la disposition des utilisateurs l'information obtenue dans le cadre d'une activité statistique. Toute statistique officielle produite ou compilée par l'Institut devrait être diffusée au bénéfice du public, et elle devrait être rendue disponible dans le respect des engagements pris.

La diffusion d'information statistique comprend l'analyse des besoins des clients et les conseils donnés à ces derniers pour qu'ils puissent trouver la source adéquate d'information, ainsi que les produits et services statistiques proprement dits, qu'ils soient fournis gratuitement ou non.

Divers modes de diffusion peuvent être envisagés, comme les supports électroniques accessibles sur Internet, les publications sur papier, les fichiers de microdonnées d'enquêtes, les renseignements statistiques transmis par téléphone, par télécopieur ou par courriel en réponse à une demande spéciale d'information adressée à l'Institut, les allocutions prononcées en public, de même que les entrevues accordées à la télévision, à la radio ou à la presse écrite. On peut également inclure la parution du *calendrier de diffusion*, les communiqués et conférences de presse. Le *Cadre intégré de diffusion* (à venir) présente les renseignements pertinents concernant tous les moyens dont dispose l'Institut quant à la diffusion.

Principes

- i) Pour une enquête, la diffusion doit se faire en conformité avec le *Cadre intégré de diffusion* (à venir). Celui-ci tiendra compte de tous les moyens dont dispose l'Institut quant à la diffusion, du contexte de protection des renseignements personnels et de sécurité propre à l'Institut, de la nature de l'information statistique à diffuser, du type d'utilisateurs visés, des mécanismes de diffusion en place et des pratiques reconnues en matière de diffusion.
- ii) Les besoins des utilisateurs doivent être étudiés avec attention. Il convient effectivement de prendre en considération les attentes de tous les utilisateurs potentiels, en se basant, entre autres, sur leurs commentaires.
- iii) Il est souhaitable, quand une entente est prise avec un tiers pour la réalisation d'une enquête statistique, que les parties impliquées se mettent d'accord à l'avance sur les questions de diffusion et de communication.

- iv) Il importe que le personnel impliqué dans la production d'information statistique soit en mesure de préparer les produits selon les normes et les moyens en vigueur à l'Institut, de manière à faciliter le travail des différents intervenants (les responsables des communications et des technologies de l'information). Le personnel doit aussi savoir à qui s'adresser lorsqu'il est question de diffusion.
- v) On doit toujours veiller à ce que l'information satisfasse aux exigences en matière de protection de la confidentialité et de sécurité. En outre, il est impératif de maintenir des règles de sécurité strictes lors de la préparation et de la diffusion d'information statistique.
- vi) Il convient de s'assurer de réutiliser au maximum les outils informatiques et processus en place à l'Institut, ceci dans un souci d'efficacité, d'économie et d'intégration.
- vii) Il convient de s'assurer que l'information communiquée aux utilisateurs est exacte et fiable, qu'elle est intelligible, facilement accessible, utilisable, objective et actuelle. Il importe aussi que l'information diffusée soit transmise à l'utilisateur dans un format facilement lisible et qui favorise sa compréhension.
- viii) Le public doit être informé de la disponibilité d'une nouvelle information statistique ou d'une nouvelle enquête. D'ailleurs, lorsque c'est possible, ceci doit avoir été au préalable annoncé dans le *calendrier de diffusion*.
- ix) L'information diffusée sur Internet ou sous une autre forme (ex. : publications papier), doit être disponible au moment de l'annonce de sa diffusion. Cette annonce peut se faire par la voie de communiqués de presse ou par le biais du site Web de l'Institut. Les employés de l'Institut susceptibles de recevoir des demandes de renseignements doivent connaître, idéalement à l'avance, les nouveaux produits et savoir à quelle personne diriger les utilisateurs pour des précisions.
- x) En principe, le public doit pouvoir obtenir de l'information récente à propos des statistiques qu'il consulte. Les renseignements présentant cette information sont appelés « métadonnées ». Dans la mesure du possible, la diffusion des statistiques doit notamment être accompagnée d'une description de celles-ci, offrant aux utilisateurs une garantie de la haute qualité du matériel.
- xi) Les réserves appropriées doivent être fournies lorsqu'une donnée sujette à l'erreur (imprécision ou possibilité de biais) est diffusée. Se référer à la *Politique de diffusion des informations statistiques et des métadonnées de l'Institut* (à venir).
- xii) Il est souhaitable que la diffusion de données révisées tende vers un équilibre entre le besoin des utilisateurs d'avoir les meilleures estimations possibles et l'incertitude causée par des révisions fréquentes.
- xiii) Il convient de voir à ce que le tarif exigé, lors de demandes de renseignements *ad hoc* par exemple, soit adéquat.
- xiv) Dans la mesure du possible, on doit tirer parti des progrès technologiques afin de permettre aux utilisateurs de traiter les statistiques de façon rentable et efficace dans leur milieu de travail.

CHAPITRE 16 – DOCUMENTATION ET EMMAGASINAGE

Dans le cadre d'une enquête statistique, il est important que soient amassés tous les documents qui peuvent être utiles à la fois aux utilisateurs et aux producteurs de données. L'emmagasiner s'avère la suite logique de cette opération, puisqu'il consiste en la conservation à plus long terme des données et des documents recueillis.

Cette section du document présente des principes reconnus concernant la documentation et l'emmagasiner; ce sont donc des guides de choix pour les équipes de réalisation des enquêtes. Cependant, l'Institut prévoit préciser ses propres pratiques en la matière à plus ou moins court terme et ainsi mieux outiller son personnel pour tout ce qui concerne la documentation et l'emmagasiner. Notamment, la *Politique sur la documentation interne* (à venir) définira de façon plus détaillée les éléments de la documentation minimale qui doit être produite et conservée pour l'ensemble des projets réalisés par l'Institut.

Comme la documentation et l'emmagasiner font appel à la diffusion d'information, ils font référence aux mêmes dimensions de la qualité, soit la *comparabilité*, en permettant la mise en parallèle avec d'autres documents similaires, l'*actualité*, l'*intelligibilité*, puisqu'elles favorisent la compréhension des données, et l'*accessibilité*.

Définitions

Dans *Le Petit Robert*, la documentation est définie de la façon suivante : « *Ensemble de documents relatifs à une question. Réunir de la documentation sur un sujet. Documentation riche, variée.* ». Dans le cas d'une enquête statistique, la documentation est un « ensemble de documents relatifs à un projet particulier, aux concepts, aux définitions et aux méthodes utilisées (collecte, traitement et analyse) pour la production des données ».

Dans ce contexte, la documentation s'adresse à deux clientèles :

- Aux **utilisateurs**, pour qui la documentation permet de déterminer les statistiques dont ils ont besoin, de les analyser et de les interpréter correctement. La documentation permet éventuellement de savoir si l'information est disponible sur d'autres supports.
- Aux **producteurs**, pour qui la documentation permet d'assurer la pérennité de l'enquête, par la transmission des connaissances aux nouveaux employés qui prendront le relais. La documentation permet aussi de maintenir, d'améliorer et de concevoir des méthodes de travail, des processus de production et des systèmes d'information, afin de produire des statistiques de façon sûre et efficace et d'en assurer la qualité. Enfin, elle permet un partage des connaissances qui peut s'avérer profitable pour d'autres processus similaires.

La *Politique sur la documentation interne* (à venir) peut être consultée pour certains détails sur la documentation minimale qui doit faire partie d'une production de l'Institut.

Quant à l'emmagasiner, il consiste à conserver les données dans le but de maintenir notre capacité à les traiter et à étudier leur évolution dans le temps. Un autre objectif poursuivi par l'emmagasiner est de servir de mémoire collective, d'accumulation de connaissances, afin de garantir la conservation du patrimoine statistique.

Principes

- i) Le niveau de détail de la documentation est fonction de plusieurs facteurs, le budget total du projet notamment. Il varie aussi selon le public cible et l'éventail d'utilisations potentielles de l'information statistique produite.
- ii) La documentation doit contenir l'ensemble des renseignements qui permettent de répondre à des questionnements ou à des besoins de renseignements complémentaires aux données statistiques, et qui proviennent à la fois des utilisateurs et des producteurs de ces données. Certains éléments de la documentation s'adressent en particulier aux utilisateurs, d'autres visent d'abord les producteurs de données. La correspondance officielle avec les clients et utilisateurs principaux de l'enquête appartient à cette dernière catégorie.
- iii) L'importance de la documentation prend tout son sens dans les projets ou les parties de projet potentiellement plus à risque.
- iv) La documentation doit être actuelle, bien structurée (organisée selon un système de classement), facilement accessible, concise et précise.
- v) La documentation doit être dépourvue d'ambiguïté et appropriée selon les usages.
- vi) Toute direction qui participe à la réalisation d'une enquête doit emmagasiner, selon sa méthode de classement, la documentation ou les fichiers de données qu'elle a produits pour ladite enquête, et ce, afin de permettre un repérage plus facile de l'information recherchée.
- vii) Il est important que la description des données soit de bonne qualité si on veut être en mesure d'utiliser et d'interpréter celles-ci correctement.
- viii) La documentation d'une enquête ainsi que les éléments d'information (fichiers) doivent être accompagnés d'un calendrier de conservation. Ce calendrier repose sur la nature des données, sur les obligations légales de l'Institut et, dans certains cas, sur les exigences de la Commission d'accès à l'information.

CHAPITRE 17 – UTILISATION DES DONNÉES ADMINISTRATIVES

Bien que l'Institut réalise annuellement plusieurs enquêtes, l'exploitation de fichiers de données administratives à des fins statistiques représente une partie importante de sa production. Que ce soit pour la réalisation de compilations de données ou pour une utilisation à titre de complément aux enquêtes, l'Institut exploite plusieurs fichiers de données administratives.

À l'Institut, les productions statistiques ayant recours aux données administratives sont cependant de natures diverses, chacune ayant des caractéristiques particulières. Il est donc prévu que des travaux supplémentaires soient entrepris sur le sujet afin de mieux encadrer la réalisation de telles productions. En attendant, la présente section du document offre une recension des principes reconnus en matière d'utilisation de données administratives.

Cette étape fait appel aux mêmes dimensions de la qualité que celles qui sont liées aux étapes associées au traitement des données que sont la validation et l'imputation ou l'estimation, par exemple. Le travail consiste donc en grande partie à assurer la dimension *fiabilité et objectivité*, *l'actualité*, la *pertinence* et la *comparabilité* de l'information statistique.

Définition

Les dossiers administratifs sont constitués aux fins de l'administration de divers programmes non statistiques. Parce qu'ils sont collectés pour des raisons autres que la production de statistiques officielles, les dossiers administratifs peuvent être utilisés sans augmenter le fardeau des répondants. De plus, les coûts associés à la production de statistiques à partir des dossiers administratifs sont habituellement inférieurs aux coûts occasionnés par la collecte directe de données. Par ailleurs, certains ensembles de dossiers administratifs sont également appelés « registres administratifs ».

En matière de production statistique, les dossiers administratifs servent (i) aux bases de sondage, à titre de base ou de supplément d'une base existante, (ii) au remplacement de la collecte de données, (iii) à la vérification et à l'imputation, (iv) à la tabulation directe, (v) à l'estimation et (vi) à l'évaluation de l'enquête, ce qui comprend la confrontation des statistiques produites à partir de données recueillies auprès de répondants avec celles tirées de fichiers administratifs.

Le recours aux dossiers administratifs peut soulever des questions relativement à la protection de la confidentialité de l'information du domaine public. Ces questions sont encore plus importantes lorsque les dossiers administratifs sont couplés avec d'autres sources d'information.

Un élément primordial dans l'utilisation de dossiers administratifs est la présence d'un identifiant unique pour chaque unité individuelle. Cet identifiant unique rend possible l'appariement des dossiers administratifs ou de certaines enquêtes possédant le même identifiant. Au Québec, chaque domaine possède son propre identifiant unique; à titre d'exemples, la santé utilise le numéro d'assurance maladie (NAM), l'éducation utilise le code permanent de l'élève, alors que les entreprises sont identifiées par leur numéro d'entreprise du Québec (NEQ). Contrairement à certains pays européens, il n'existe pas au Québec de code unique attribué à chaque individu disponible dans toutes les sources de données, ni même de correspondance entre les identifiants uniques des différents domaines. Pour ces raisons, le jumelage à l'aide d'un identifiant unique n'est possible qu'à l'intérieur d'un même domaine.

Principes

- i) Utiliser les dossiers administratifs chaque fois qu'un tel recours constitue une solution de rechange rentable à la collecte directe de données. Cette solution est attrayante, car elle évite les coûts inhérents à la collecte de données et n'augmente pas le fardeau de réponse, pourvu que la couverture et le cadre conceptuel des dossiers administratifs soient compatibles avec la population visée et que les données qu'ils contiennent soient jugées de qualité suffisante pour produire l'information statistique désirée.
- ii) Selon l'objectif à atteindre, il est souvent avantageux de combiner des données administratives avec des données provenant d'une autre source, une enquête par exemple.
- iii) Le couplage des enregistrements doit être conforme à la *Politique de communication de fichiers, de renseignements personnels et confidentiels à des fins de comparaison, de couplage et d'appariement* de l'Institut. En particulier, toutes les demandes de couplage d'enregistrements doivent être approuvées par le directeur général.
- iv) Toute décision relative à l'utilisation des dossiers administratifs doit être précédée d'une évaluation de ces dossiers sur le plan de la couverture, du contenu, des concepts et des définitions, des procédures d'assurance et de contrôle de la qualité mises en place par le programme administratif pour en assurer la qualité, de la fréquence des données, de la rapidité de l'Institut à recevoir les données et de la stabilité du programme administratif au fil du temps.
- v) Il va de soi que le coût associé à l'obtention des dossiers administratifs est également un facteur clé à prendre en compte dans la décision de les utiliser ou non. Cependant, ce coût doit être mis en relation avec l'actualité des données. Ainsi, un arbitrage entre les coûts et l'actualité des données est recommandé.
- vi) Aux fins de l'utilisation de dossiers administratifs, l'Institut doit parfois réaliser certaines étapes du processus d'enquête, décrites dans les sections précédentes, de façon à compenser les différences au chapitre des objectifs de qualité de l'organisme responsable du programme administratif à l'origine du recueil des données. Ainsi, des étapes comme la saisie ou la validation de certaines données, recueillies par l'organisme responsable mais inutilisées aux fins administratives, doivent parfois être réalisées par l'Institut avant l'exploitation des dossiers administratifs.

ANNEXE A – LES MÉTHODES DE COLLECTE

L'objectif de cette annexe est de donner un aperçu des principaux avantages et désavantages propres à chacune des méthodes de collecte. Plusieurs ouvrages ont été rédigés sur ce sujet; les références inscrites à la fin de l'annexe en forment un échantillon qui peut être consulté pour obtenir de plus amples détails.

Les résultats qui sont présentés dans cette annexe sont le fruit des recherches scientifiques effectuées sur le sujet. Bien qu'il existe quantité de travaux de recherche liés aux méthodes de collecte, il est possible qu'aucun ne traite de façon spécifique de la comparaison de deux méthodes de collecte quant à un critère bien précis. Par exemple, il existe peu de recherches scientifiques comparant les enquêtes par voies électroniques avec les autres types d'enquête.

La structure de la présente annexe permet d'abord de décrire l'information propre à chacun des quatre principaux modes de collecte, des deux modes d'administration du questionnaire et des deux modes d'acquisition des données, paramètres qui déterminent la méthode de collecte. Un bref aperçu des méthodes de collecte mixte, dont la fréquence d'utilisation ne cesse d'augmenter, achève ensuite cette section. Finalement, l'annexe se termine en distinguant quelques particularités propres à chacun des types d'enquêtes existant, soit les enquêtes transversales et les enquêtes longitudinales.

Modes de collecte

Enquête en face à face

De façon générale, les enquêtes en face à face, qu'elles soient assistées par ordinateur ou non :

- sont les plus coûteuses,
- sont celles dont la logistique opérationnelle est la plus complexe,
- requièrent une plus grande organisation des ressources que les enquêtes téléphoniques et postales,
- génèrent de meilleurs taux de réponse.

Dans les enquêtes auprès des ménages, c'est un mode de collecte souvent utilisé en conjonction avec un plan de sondage par grappes, ce qui permet d'assurer une bonne couverture de la population visée.

Comme le mode de collecte en face à face est l'alternative aux enquêtes téléphoniques lorsqu'on a recours à un intervieweur pour administrer le questionnaire, ces deux combinaisons mode de collecte / mode d'administration du questionnaire sont souvent comparées.

Ainsi, pour les enquêtes en face à face, contrairement aux enquêtes téléphoniques, il est possible d'utiliser une aide visuelle au moment de l'entrevue. Par contre, il y est plus risqué d'observer un effet « dû » à l'intervieweur, car le niveau d'interaction de l'intervieweur avec le répondant est généralement plus fort que celui existant dans les enquêtes téléphoniques.

Enquête téléphonique

Le mode de collecte téléphonique est fréquemment utilisé, puisqu'il s'avère le choix le moins coûteux parmi les modes impliquant un intervieweur. De plus, c'est un mode de collecte qui permet habituellement d'obtenir des taux de réponse relativement élevés (supérieurs à ceux obtenus dans les enquêtes postales ou par voie électronique).

Le mode de collecte téléphonique est aussi souvent employé comme mode de collecte complémentaire pour effectuer des opérations de dépistage préalable à la collecte originale, et ce, peu importe le mode de collecte utilisé lors de celle-ci.

Enquête postale

Le mode de collecte postal s'avère moins coûteux que les modes de collecte téléphoniques et en face à face (la comparaison des coûts avec le mode de collecte par voies électroniques n'est pas aussi déterminée). Par contre, c'est un mode de collecte qui nécessite des efforts importants de relance des répondants afin de maximiser le taux de réponse. Malgré ces efforts, le mode de collecte postal, à l'exception du mode de collecte par voies électroniques, s'avère celui présentant les taux de réponse les plus bas.

C'est un mode de collecte qui est souvent utilisé avec des populations spécialisées ou spécifiques, puisqu'il nécessite l'emploi d'une base de sondage contenant les coordonnées postales de tous ces éléments.

Il peut être plus difficile de distinguer les unités inadmissibles dans ce mode de collecte par rapport aux autres modes.

Dans ce mode de collecte, les questionnaires sont autoadministrés et l'acquisition des données se fait par papier. Des opérations de saisies, de codifications et de validations des données sont donc nécessaires à la suite de la collecte. Ces opérations peuvent parfois être passablement laborieuses.

Enquête par voies électroniques

Au moment d'écrire ces lignes, c'est un mode de collecte qui est encore en émergence. Il se veut une solution de rechange assistée par ordinateur à l'enquête postale. Comme il existe peu ou pas de bases de sondage pour ce mode (sauf peut-être pour certaines populations spécialisées ou spécifiques), son utilisation semble s'avérer plus prometteuse comme mode de collecte complémentaire.

Pour l'instant, il semble que ce soit le mode de collecte présentant les taux de réponse les plus bas.

Modes d'administration du questionnaire

Par un intervieweur

L'administration du questionnaire par l'intervieweur permet un meilleur « contrôle » du déroulement de l'entrevue, offre la possibilité de clarifier des questions ou des réponses, ou de motiver les répondants à fournir une réponse précise et complète. D'ailleurs, par leur habileté à répondre aux interrogations des répondants, les intervieweurs peuvent contribuer à hausser les taux de collaboration et de réponse et ainsi influencer sur les caractéristiques de l'erreur de non-réponse de l'enquête.

Par contre, l'utilisation d'un intervieweur (et plus particulièrement dans les enquêtes en face à face) peut influencer négativement sur les réponses fournies, notamment pour les questions sensibles. De même, l'administration du questionnaire par un intervieweur augmente généralement les effets de désirabilité sociale comparativement aux questionnaires autoadministrés.

Autoadministré

L'utilisation des questionnaires autoadministrés constitue le meilleur choix pour les questions sensibles et intimes. De plus, ces questionnaires permettent généralement d'atténuer les effets de désirabilité sociale. Les effets liés à l'ordre des questions ou des choix de réponses sont moins susceptibles de se présenter dans ce type de questionnaires (en comparaison, notamment, avec les questionnaires administrés par un intervieweur au téléphone).

Par contre, le recours à des questionnaires autoadministrés peut avoir des implications plus sérieuses sur les biais dus à la non-réponse totale, car, contrairement aux méthodes avec intervieweurs, le répondant peut consulter l'ensemble du questionnaire avant de choisir de répondre ou non à l'enquête. Dans les méthodes avec intervieweur, le sujet est énoncé au début dans des termes plutôt vagues (par exemple, une enquête sur la santé et la vie sociale), et le répondant ne peut se faire une idée précise du questionnaire avant de répondre.

Généralement, la présence de non-réponse partielle est aussi plus importante dans les questionnaires autoadministrés que dans les questionnaires administrés par un intervieweur.

Modes d'acquisition des données

Collecte assistée par ordinateur

Les entrevues assistées par ordinateur offrent énormément d'avantages par rapport aux entrevues réalisées avec un questionnaire papier. Par exemple, il est possible de faire une codification automatique des réponses, de présenter un questionnaire avec des cheminements compliqués, d'inclure des données auxiliaires bénéfiques à l'entrevue et de procéder à certaines validations automatiques en cours d'entrevue assurant une correction immédiate.

Collecte papier

Le questionnaire papier permet la présentation de choix de réponse à plusieurs catégories et offre aussi une bonne solution pour poser des questions intimes ou dites sensibles. Toutefois, puisque aucune correction interactive n'est faite au moment de l'entrevue, le questionnaire papier peut générer plus d'erreurs qu'une collecte assistée par ordinateur.

Méthodes de collecte mixtes

L'utilisation de plus en plus fréquente de méthodes de collecte mixtes, c'est-à-dire de méthodes utilisant une combinaison de méthodes de collecte, s'explique par le fait que ces méthodes peuvent être efficaces pour maximiser les taux de réponse tout en réduisant les coûts. Souvent, elles permettent d'obtenir un meilleur équilibre entre les coûts et la qualité. Par exemple, on voit de plus en plus d'enquêtes longitudinales qui utilisent un mode de collecte en face à face dans le ou les premiers cycles avant de se tourner vers un mode de collecte moins coûteux dans les cycles subséquents.

Types d'enquêtes

Les enquêtes longitudinales peuvent permettre des économies. En effet, le même échantillon est utilisé pour chacun des cycles de l'enquête de même qu'au moins une partie des instruments de mesure. Toutefois, ces enquêtes présentent le danger que le répondant « s'habitue » au questionnaire et change son comportement ou ses réponses pour réduire le fardeau de réponse (« panel effect »).

Références

- GROVES, R. M. (1989). *Survey errors and survey costs*. John Wiley and Sons Inc., New York.
- GROVES, R. M., F. J. FOWLER, M.P. COUPER, J. M. LEPKOWSKI, E. SINGER et R. TOURANGEAU (2004). *Survey methodology*. John Wiley and Sons Inc., New York.
- GROVES, R. M. et R. L. KAHN (1979). *Surveys by telephone*. Academic Press, New York.
- LYBERG, L., P. BIEMER, M. COLLINS, E. DE LEEUW, C. DIPPO, N. SCHWARZ et D. TREWIN (éd.) (1997). *Survey Measurement and Process Quality*. Wiley, New York.

BIBLIOGRAPHIE

- BENVENISTE, C. (2005). « Savoir compter, savoir conter. Épisode n° 3 – Des graphiques et des tableaux », INSEE, *Courrier des statistiques*, n° 115, septembre 2005, p. 49-53.
- COCHRAN, W. G. (1977). *Sampling Techniques*, New York, John Wiley & Sons, 428 p.
- COX, B. G., D. A. BINDER, B. NANJAMME CHINNAPPA, A. CHRISTIANSON, M. J. COLLEDGE ET P. S. SCOTT (éd) (1995). *Business Survey Methods*, New York, John Wiley & Sons, 732 p.
- FELLEGI, I. P. ET D. HOLT (1976). « A systematic approach to automatic edit and imputation », *Journal of the American Statistical Association*, n° 71, p. 17-35.
- GRANQUIST, L. (1995). « Improving the Traditional Editing Process », dans Cox B. G. et coll (éd.), *Business Survey Methods*, New York, John Wiley & Sons, p. 385-403.
- GRANQUIST, L. (1984). « On the Role of Editing », *Statistical Review*, n° 2, p. 105-118.
- GRANQUIST, L. ET J.G. KOVAR (1997). « Editing of survey data : how much is enough? », dans Lyberg, L. et coll. (éd.), *Survey Measurement and Process Quality*, New York, John Wiley & Sons, p. 415-435.
- GROVES, R. M., D. A. DILLMAN, J. L. ELTINGE ET R. J. A. LITTLE (2002). *Survey Nonresponse*, New York, John Wiley & Sons, 500 p.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (Juin 2009). *Politique d'information au répondant*, document interne.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (2006). *Politique d'accès aux microdonnées d'enquêtes individus ou ménages*, document interne.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (septembre 2006). *Rapport de recension des pratiques de référence en matière de qualité dans les enquêtes*, document interne.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (juin 2006). *Rapport de constats des pratiques actuelles en matière de qualité dans les enquêtes*, document interne.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (janvier 2006). *Le cadre intégré de gestion de la qualité de l'Institut de la statistique du Québec – Document d'orientations générales*, Québec, Gouvernement du Québec, 20 p.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (janvier 2006). *Politique générale en matière de gestion de la qualité de l'Institut de la statistique du Québec*, Québec, Gouvernement du Québec, 15 p.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (2004). *Cahier des normes en matière d'édition, d'impression et d'identification visuelle*, document interne.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (2004). *Politique relative à la confidentialité des tableaux de résultats pour diffusion*, document interne.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (2004). *Politique de communication des fichiers de renseignements personnels et confidentiels à des fins de comparaison, de couplage et d'appariement*, document interne.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (2004). *Politique linguistique de l'Institut de la statistique du Québec*, document interne.

- INSTITUT DE LA STATISTIQUE DU QUÉBEC (1998). *Loi sur l'Institut de la statistique du Québec*, Québec, Gouvernement du Québec.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (à venir). *Cadre intégré de diffusion*.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (à venir). *Politique sur l'évaluation de programmes et sur la révision des processus*.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (à venir). *Politique de diffusion des informations statistiques et des métadonnées*.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (à venir). *Politique sur la documentation interne*.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (à venir). *Procédures et directives relatives au suivi qualitatif des entrevues*.
- KISH, L. (1965). *Survey Sampling*, New York, John Wiley & Sons, 643 p.
- LATOUCHE, M. ET J.-M. BERTHELOT (1992). « Use of a score function to prioritize and limit recontacts in editing business surveys », *Journal of Official Statistics*, n° 8, p. 389-400.
- LOHR, S. L. (1999). *Sampling : Design and Analysis*, Californie, Duxbury Press, 494 p.
- LYBERG, L., P. BIEMER, M. COLLINS, E. DE LEEUW, C. DIPPO, N. SCHWARZ ET D. TREWIN (éd) (1997). *Survey Measurement and Process Quality*, New York, John Wiley & Sons, 777 p.
- OFFICE FOR NATIONAL STATISTICS (2007). *Guidelines for measuring statistical quality*, London, Government Statistical Service, 82 p.
- OFFICE FOR NATIONAL STATISTICS (1997). *Statistical Quality Checklist*, London, Government Statistical Service (Crown), 52 p.
- STATISTIQUE CANADA (DIVISION DES MÉTHODES D'ENQUÊTES SOCIALES) (2003). *Méthodes et pratiques d'enquête*, Ottawa, Ministère de l'Industrie, n°12-587-XPF, 422 p.
- STATISTIQUE CANADA (octobre 2003). *Lignes directrices concernant la qualité*, Ottawa, Ministère de l'Industrie, n°12-539-XIF, 122 p.
- STATISTICS FINLAND (2002). *Quality Guidelines for Official Statistics*, Helsinki, Statistics Finland, Handbooks 43b, 132 p.
- STATISTICS NEW ZEALAND (2006). *Statistics New Zealand Graphics Guidelines*, [En ligne]: <http://www.stats.govt.nz/about-us/policies-and-guidelines/data-use/graphic-guidelines-standards.htm> (page consultée le 20 novembre 2006).
- STATISTICS NEW ZEALAND (Août 1998). « Protocols for Official Statistics ». Dans : Development of National Statistical Systems. *Site de la United Nations Statistics Division*, [En ligne] : <http://unstat.un.org/unsd/dnss/gp/searchgp.aspx> (page consultée le 13 avril 2010).
- TEMAM, D. (2005). « Savoir compter, savoir conter. Épisode n° 4 – Bien préparer, voilà le secret », INSEE, *Courrier des statistiques*, n° 116, décembre 2005, p. 39-40.
- TEMAM, D. (2004). « Savoir compter, savoir conter. Épisode n° 1 – Des chiffres : point trop n'en faut », INSEE, *Courrier des statistiques*, n° 112, décembre 2004, p. 55-57.
- TREMBLAY, A. (1991). *Sondages : histoire, pratique et analyse*, Boucherville, Gaëtan Morin éditeur, 492 p.

La loi constitutive de l'Institut de la statistique du Québec (Institut) contient des dispositions claires quant aux obligations de qualité qu'il doit respecter. Ainsi, en novembre 2005, l'Institut a adopté une politique générale comportant des éléments relatifs à la gestion de la qualité des enquêtes statistiques. Le présent document vise le large public des clients, utilisateurs et partenaires de l'Institut qui, bien qu'ils connaissent les grands principes, ne sont pas des spécialistes des disciplines de pointe impliquées dans la réalisation d'enquêtes statistiques. Le *Document de principes sur la qualité dans les enquêtes* est conçu par l'Institut pour les informer de ses préoccupations générales en matière d'assurance qualité.

Les principes sont présentés selon un ordre logique correspondant au découpage d'une enquête en fonction des principaux processus ou des étapes qu'elle comporte, de sa conception à sa mise en œuvre. En plus d'une mise en contexte, le document compte 17 sections : une pour chacune des 16 étapes types de réalisation d'une enquête et une autre consacrée à l'utilisation de données administratives qui comporte certaines particularités.

Au fil du temps et à mesure que les façons de faire de l'Institut seront modifiées au bénéfice de ses utilisateurs, le *Document de principes sur la qualité dans les enquêtes* rendra compte de l'évolution des orientations prises par l'Institut.