Guide de l'utilisateur des microdonnées

ENQUÊTE SUR LE SERVICE TÉLÉPHONIQUE RÉSIDENTIEL

Décembre 2010

Statistique Canada
Statistics Canada

Canada

Table des matières

1.0 Introduction 5

2.0 Contexte 7

3.0 Objectifs 9

4.0 Concepts et définitions 11

5.0 Méthodologie de l'enquête 13
5.1 Modifications au plan de l'Enquête sur la population active pour l'Enquête sur le service téléphonique résidentiel 13
5.2 Taille de l'échantillon selon la province pour l'Enquête sur le service téléphonique residentiel 13

6.0 Collecte des données 15
6.1 Réalisation des interviews dans le cadre de l'Enquête sur la population active 15
6.2 Supervision et contrôle de qualité 15
6.3 Non-réponse à l'Enquête sur la population active 16
6.4 Modifications apportées à la collecte des données de l'Enquête sur le service téléphonique résidentiel 16
6.5 Non-réponse à l'Enquête sur le service téléphonique résidentiel 16

7.0 Traitement des données 17
7.1 Saisie des données 17
7.2 Vérification 17
7.3 Codage des questions ouvertes 17
7.4 Création de variables dérivées 17
7.5 Pondération 18
7.6 Suppression des renseignements confidentiels 18

8.0 Qualité des données 19
8.1 Taux de réponse 19
8.2 Erreurs relatives à l'enquête 19
8.2.1 Base de sondage 20
8.2.2 Collecte des données 20
8.2.3 Traitement des données 20
8.2.4 Non-réponse 21
8.2.5 Mesure de l'erreur d'échantillonnage 21

9.0 Lignes directrices pour la totalisation, l'analyse et la diffusion de données 23
9.1 Lignes directrices pour l'arrondissement d'estimations 23
9.2 Lignes directrices pour la pondération de l'échantillon en vue de la totalisation 24
9.3 Définitions de types d'estimations : catégoriques et quantitatives 24
9.3.1 Estimations catégoriques 24
9.3.2 Estimations quantitatives 24
9.3.3 Totalisation d'estimations catégoriques 25
9.3.4 Totalisation d'estimations quantitatives 25
9.4 Lignes directrices pour l'analyse statistique 25
9.5 Lignes directrices pour la diffusion de coefficients de variation 26
9.6 Seuils pour la diffusion des estimations pour l'Enquête sur le service téléphonique résidentiel 28

10.0 Tables de variabilité d'échantillonnage approximative 29
10.1 Comment utiliser les tables de coefficients de variation pour des estimations catégoriques 30
10.1.1 Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques 31
10.2 Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance 36
10.2.1 Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance 37
10.3 Comment utiliser les tables de coefficients de variation pour effectuer un test t 37
10.3.1 Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t 38
10.4 Coefficients de variation pour des estimations quantitatives 38
10.5 Tables de coefficients de variation 39
10.6 Estimation de variance pour les fichiers maître 39

11.0 Pondération 41
11.1 Procédures de pondération pour l'Enquête sur la population active 41
11.2 Procédures de pondération pour l'Enquête sur le service téléphonique résidentiel 42

1.0 Introduction

Statistique Canada a mené la plus récente Enquête sur le service téléphonique résidentiel (ESTR) en décembre 2010. Ce manuel a été produit pour faciliter la manipulation du fichier de microdonnées portant sur les résultats de l'enquête.

Toutes les questions concernant l'ensemble de données ou son utilisation devraient être adressées à :

Statistique Canada

Services à la clientèle
Division des enquêtes spéciales
Téléphone : 613-951-3321 ou appelez sans frais : 1 800 461-9050
Télécopieur : 613-951-4527
Courriel : des@statcan.ca

2.0 Contexte

L'enquête sur le service téléphonique résidentiel (ESTR) recueille les renseignements sur le service téléphonique en vigueur au Canada.

L'enquête a été menée sur une base semestrielle en 2000, 2001, 2002 (mai et novembre) et 2004 (mai et décembre). En 2003, elle n'a été menée qu'une seule fois, soit en mai. On a déterminé qu'il était possible d'obtenir suffisamment de renseignements en effectuant l'enquête une fois par année. C'est la raison pour laquelle, depuis 2005, l'ESTR est menée en décembre seulement. L'ESTR n'a pas été menée en 2009.

3.0 Objectifs

Il y a trois objectifs principaux. Ce sont :

1) la cueillette d'information sur l'étendue du service téléphonique résidentiel au Canada et produire cette information au niveau provincial,

2) la cueillette d'information sur les types de ménages qui n'ont pas de service téléphonique. et

3) la cueillette d'information sur les ménages qui ont un service téléphonique sans fil seulement.

De façon à s'assurer que l'enquête remplie ces objectifs, Statistique Canada a préparé un plan d'analyse qui reflète les besoins de données. Ce plan a été utilisé pour la conception du questionnaire et pour justifier les variables requises.

4.0 Concepts et définitions

Ce chapitre donne un aperçu des concepts et des définitions d'intérêt pour les utilisateurs.

Téléphones cellulaires actifs
Il s'agit de téléphones cellulaires et de téléphones mobiles comme le IPhone et le BlackBerry utilisés par les membres du ménage. Les téléphones utilisés à des fins commerciales sont inclus. Les téléphones sans fil sont exclus.

Service téléphonique avec fil
Il s'agit d'un service de téléphone résidentiel relié par un fil, comme un fil téléphonique, le câble ou un service par Internet, c'est-à-dire : un service téléphonique voix par IP avec un numéro unique - (VoIP en anglais). Sont exclus les services sans fil comme les téléphones cellulaires et les autres appareils mobiles.

Service téléphonique par une ligne terrestre
Il s'agit du service téléphonique classique qui est offert depuis des décennies par les compagnies de téléphone locales.

Service téléphonique d'un câblodistributeur
Les câblodistributeurs offrent un service téléphonique qui permet de recevoir et d'effectuer des appels au moyen du réseau de câblodistribution.

VoIP service (service téléphonique par Internet)
« D'un service par Internet », c'est-à-dire : un service téléphonique voix par IP avec un numéro unique - (« VoIP » en anglais) qui permet au client de recevoir et d'effectuer des appels sur Internet soit au moyen d'un appareil raccordé au téléphone ou au moyen d'un ordinateur.

5.0 Méthodologie de l'enquête

Comme l'Enquête sur le service téléphonique résidentiel (ESTR) a été menée en décembre 2010 auprès d'un sous-échantillon des logements inclus dans l'échantillon de l'Enquête sur la population active (EPA), son plan de sondage est donc étroitement lié à celui de l'EPA. Une description détaillée du plan de l'EPA figure dans la publication de Statistique Canada intitulée *Méthodologie de l'Enquête sur la population active du Canada,* N° 71-526-XPB au catalogue. Les sections 5.1 et 5.2 décrivent comment l'Enquête sur le service téléphonique résidentiel s'est écartée du plan de base de l'EPA de décembre 2010.

5.1 Modifications au plan de l'Enquête sur la population active pour l'Enquête sur le service téléphonique résidentiel

On a utilisé pour l'Enquête sur le service téléphonique résidentiel des groupes de renouvellement 1 et 2 pour l'Ontario et le Québec et groupes 1, 2 et 4 pour tous les autres provinces dans l'échantillon de l'EPA de décembre 2010. Pour les besoins de l'ESTR, on a établi la couverture de l'EPA au niveau des ménages. Cependant, contrairement à l'EPA, pour laquelle des données sont recueillies pour tous les membres admissibles d'un ménage, l'ESTR n'a donné lieu à la cueillette d'information qu'auprès d'un seul membre d'un ménage qui a déclaré pour tout le ménage.

5.2 Taille de l'échantillon selon la province pour l'Enquête sur le service téléphonique residentiel

Le tableau qui suit montre le nombre de ménages inclus dans les groupes de renouvellement échantillonnés de l'EPA qui étaient admissibles pour le supplément de l'Enquête sur le service téléphonique résidentiel.

Provinces	**Taille de l'échantillon**
Terre-Neuve-et-Labrador	912
Île-du-Prince-Édouard	614
Nouvelle-Écosse	1 400
Nouveau-Brunswick	1 346
Québec	3 054
Ontario	4 734
Manitoba	2 156
Saskatchewan	1 824
Alberta	2 426
Colombie-Britannique	2 831
Canada	**21 297**

6.0 Collecte des données

La collecte des données aux fins de l'Enquête sur la population active (EPA) a lieu tous les mois durant la semaine qui suit la semaine de référence de l'EPA. La semaine de référence correspond habituellement à la semaine où tombe le 15 du mois.

6.1 Réalisation des interviews dans le cadre de l'Enquête sur la population active

Les intervieweurs de Statistique Canada sont des employés embauchés et formés pour mener l'EPA et d'autres enquêtes auprès des ménages. Chaque mois, ils communiquent avec les logements de l'échantillon pour recueillir l'information requise au sujet de la population active. Chaque intervieweur entre en contact avec environ 75 logements par mois.

Dans le cas des logements nouvellement incorporés à l'échantillon, l'intervieweur mène habituellement une interview sur place assistée par ordinateur (IPAO). L'intervieweur recueille premièrement des renseignements socio-démographiques sur chaque membre du ménage, puis recueille des renseignements relatifs à la population active concernant tous les membres du ménage qui sont âgés de 15 ans et plus et ne sont pas membres des Forces armées. Pourvu qu'il y ait un appareil téléphonique dans le logement et que l'autorisation en ait été accordée, les interviews subséquentes sont réalisées par téléphone, par moyen d'interview téléphonique assistée par ordinateur (ITAO), provenant d'une unité centralisée là où les cas sont assignés au hasard aux intervieweurs. Ainsi, environ 85 % de l'ensemble des ménages sont interviewés par téléphone. Lors de ces interviews mensuelles subséquentes, l'intervieweur confirme les renseignements socio-démographiques obtenus le premier mois, puis il recueille les renseignements relatifs à la population active pour le mois courant.

Pour chaque logement, l'intervieweur obtient habituellement les renseignements relatifs à tous les membres du ménage auprès d'un membre bien informé du ménage. Appelée réponse « par procuration », cette façon de procéder est appliquée parce qu'il serait trop coûteux et fastidieux de faire plusieurs visites ou appels pour obtenir l'information directement auprès de chaque répondant. Environ 65 % des renseignements recueillis le sont de cette manière.

Lorsque, pendant la période de six mois où un logement fait habituellement partie de l'échantillon, un ménage entier déménage et est remplacé par un nouveau ménage, on recueille des renseignements sur le nouveau ménage pendant le reste de la période de six mois.

À la fin des interviews mensuelles menées pour l'EPA, les intervieweurs présentent le questionnaire d'enquête supplémentaire, le cas échéant, qu'il faut faire remplir à certains ou à la totalité des membres d'un ménage ce mois-là.

6.2 Supervision et contrôle de qualité

Tous les intervieweurs de l'EPA travaillent sous la supervision d'un groupe d'intervieweurs principaux. Ceux-ci ont pour responsabilité de s'assurer que les intervieweurs connaissent bien les concepts de l'EPA, les procédures qui y sont associées et ses nombreuses enquêtes supplémentaires. Ils ont aussi pour fonction de contrôler régulièrement le travail des intervieweurs et d'examiner les documents complétés par ces derniers. Les intervieweurs principaux sont, à leur tour, supervisés par les gestionnaires du programme de l'EPA, affectés dans chacun des bureaux régionaux de Statistique Canada.

6.3 Non-réponse à l'Enquête sur la population active

Les intervieweurs ont pour consigne de faire tous les efforts raisonnables pour obtenir des interviews aux fins de l'EPA avec des membres des ménages admissibles. Lorsqu'une personne refuse au départ de participer à l'EPA, le bureau régional envoie à l'adresse du logement une lettre soulignant l'importance que revêtent l'enquête et la collaboration du ménage. Cette lettre est suivie d'un deuxième appel ou d'une deuxième visite de l'intervieweur. Dans les cas où la visite ou l'appel de l'intervieweur survient à un moment inopportun, on fixe un rendez-vous pour une visite de rappel à un moment plus opportun. Lorsqu'il n'y a personne dans le logement, l'intervieweur rappelle à de nombreuses reprises. Quelles que soient les circonstances, on ne remplace jamais des logements échantillonnés par d'autres logements en cas de non-réponse au questionnaire de l'EPA.

Chaque mois, après qu'on a déployé tous les efforts voulus pour réaliser les interviews, il reste un petit nombre de ménages qui n'ont pas répondu. Pour les ménages non répondants et pour lesquels on a obtenu le mois précédent des renseignements aux fins de l'Enquête sur la population active, on reporte et on utilise ces renseignements à titre de données de l'enquête du mois courant. On ne collecte pas de données sur ces ménages dans le cadre d'une enquête supplémentaire.

6.4 Modifications apportées à la collecte des données de l'Enquête sur le service téléphonique résidentiel

Les questions de l'Enquête sur le service téléphonique résidentiel (ESTR) ont été posées à un membre informé du ménage. Après avoir terminé l'interview de l'Enquête sur la population active, l'intervieweur présentait l'ESTR et continuait l'interview avec la permission du répondant.

L'application de l'Enquête sur le service téléphonique résidentiel avait été programmée pour apparaître sur la liste des enquêtes à être remplies après que la composante démographique de l'EPA a été remplie. Toutes les composantes de l'ESTR qui n'avaient pas été remplies au moment de la transmission de l'EPA à un des bureaux régionaux de Statistique Canada, ont été laissées en blanc et transmises avec l'EPA.

6.5 Non-réponse à l'Enquête sur le service téléphonique résidentiel

Pour les ménages ayant répondu au questionnaire de l'EPA, l'étape suivante de la collecte des données consistait à faire remplir le questionnaire de l'Enquête sur le service téléphonique résidentiel. Au total, 21 297 ménages étaient admissibles pour l'enquête supplémentaire; on a achevé l'interview aux fins de l'Enquête sur le service téléphonique résidentiel pour 19 310 de ces ménages, ce qui a donné un taux de réponse de 90,7 %. Le chapitre 8.0 (Qualité des données) renferme des renseignements plus détaillés sur les taux de réponse.

7.0 Traitement des données

Le principal produit de l'Enquête sur le service téléphonique résidentiel (ESTR) est un fichier de microdonnées « épuré ». Ce chapitre présente un bref résumé des phases de traitement inhérentes à la production de ce fichier.

7.1 Saisie des données

Les intervieweurs saisissent directement les réponses aux questions de l'enquête au moment de l'interview à l'aide d'une version automatisée du questionnaire. L'emploi d'un questionnaire automatisé réduit les délais et coûts de traitement associés à la saisie des données, aux erreurs de transcription et à la transmission des données. On soumet les données des réponses à un cryptage pour en préserver le caractère confidentiel, puis on les transmet par modem au bureau régional concerné de Statistique Canada. Les données sont par la suite transmises à Ottawa, au moyen d'une ligne protégée, pour y être traitées davantage. Au total, on a saisi et transmis pour l'enquête 21 297 documents.

Une partie du contrôle se fait au moment de l'interview. Lorsque les renseignements introduits sont hors limites (trop faibles ou trop élevés) des valeurs attendues, ou qu'ils entrent en contradiction avec des renseignements introduits auparavant, l'intervieweur voit paraître à l'écran de l'ordinateur des messages lui demandant de modifier les renseignements. Cependant, pour certaines questions, l'intervieweur a la possibilité de passer outre aux contrôles et de sauter des questions si le répondant ne connaît pas la réponse ou refuse de répondre. Pour cette raison, on soumet les données des réponses à d'autres processus de vérification et d'imputation après réception au bureau central.

7.2 Vérification

La première étape du traitement d'enquête effectué au bureau central a été de remplacer les valeurs « hors limites » incluses dans le fichier de données par des blancs. Ce processus a été conçu pour faciliter les vérifications ultérieures.

Les erreurs dans le déroulement du questionnaire, où l'on a relevé des questions qui ne s'appliquaient pas au répondant (et auxquelles on n'aurait donc pas dû répondre) et qui renfermaient des réponses, constituaient le premier type d'erreurs traitées. Dans ces cas, une vérification par ordinateur a éliminé automatiquement les données superflues en suivant l'ordre du questionnaire dicté par les réponses à des questions antérieures et subséquentes, parfois.

Le second type d'erreurs traitées avait trait à un manque d'information dans les questions pour lesquelles le répondant aurait dû répondre. Pour ce type d'erreur, un code de non-réponse ou « non déclaré » était attribué au poste.

7.3 Codage des questions ouvertes

Aucun poste du questionnaire n'a été rempli par l'intervieweur sous forme de question ouverte. Une question partiellement ouverte faisait partie de l'enquête. Il s'agit de la réponse « Autre - Précisez » à la question RTS_Q09 (À l'exclusion des téléphones cellulaires, quel autre type de service téléphonique avez-vous pour votre ménage? S'agit-il...?). Au total, 49 réponses ont été inscrites dans cette catégorie, 22 ayant été codées manuellement à la bonne réponse.

7.4 Création de variables dérivées

Un certain nombre de données élémentaires incluses dans le fichier de microdonnées en combinant des postes sur le questionnaire pour faciliter l'analyse des données. La variable

dérivée CMA1, par exemple, est en fait une combinaison des régions métropolitaines de recensement (RMR) et des agglomérations de recensement (AR). Les AR ont été recodés à 0 alors que les RMR demeurent telles quelles. La variable dérivée « taille du lieu de résidence » (AREASIZE) a également été créée. Cette variable permet d'obtenir le code de la taille de la population, selon les définitions du Recensement de 2006 en matière de la taille de classe des centres de population dans la base de sondage de l'Enquête sur la population active (EPA).

7.5 Pondération

Le principe qui sous-tend une estimation pour un échantillon probabiliste comme celui de l'EPA veut que chacune des personnes incluses dans l'échantillon « représente », en plus d'elle-même, plusieurs autres personnes qui en sont exclues. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne incluse dans l'échantillon représente 50 membres de la population. Le même principe s'applique aussi aux ménages.

La phase de la pondération est une étape où l'on calcule ce nombre (ou poids) pour chaque enregistrement. Ce poids, qui figure dans le fichier de microdonnées, **doit** servir à calculer des estimations significatives à partir de l'enquête. Si, par exemple, le nombre de ménages ayant un numéro de téléphone ou plus, doit être estimé, cette opération s'effectue en sélectionnant les enregistrements se référant aux ménages inclus à l'intérieur de l'échantillon qui présentent cette caractéristique et en additionnant les poids inscrits dans ces enregistrements.

Le chapitre 11.0 renferme des détails au sujet de la méthode utilisée pour calculer ces poids.

7.6 Suppression des renseignements confidentiels

Il convient de souligner que les fichiers de microdonnées « à grande diffusion » décrits ci-dessus diffèrent sous un certain nombre d'aspects importants des fichiers « maîtres » de l'enquête que conserve Statistique Canada. Ces différences sont le résultat de mesures prises pour protéger l'anonymat des répondants à un questionnaire d'enquête. Les utilisateurs ayant besoin d'avoir accès à de l'information exclue des fichiers de microdonnées peuvent acheter des totalisations spéciales. Les estimations produites seront communiquées à l'utilisateur, sous réserve du respect des lignes directrices pour l'analyse et la diffusion dont le chapitre 9.0 de ce document fournit un aperçu.

Province — Suppression d'identificateurs géographiques

Le fichier maître de données de l'enquête inclut des identificateurs géographiques explicites pour la province et la région métropolitaine de recensement. Il est aussi possible d'obtenir, lorsque les tailles des échantillons le permettent, des estimations selon la catégorie de taille pour les centres de population (AREASIZE). Les fichiers d'enquête de microdonnées à grande diffusion ne renferment habituellement pas d'identificateurs géographiques inférieurs au niveau provincial. Comme l'ESTR a comme unités d'échantillon les ménages, les variables RMR et taille de centre de population sont sur le fichier de microdonnées à grande diffusion.

8.0 Qualité des données

8.1 Taux de réponse

Le tableau qui suit renferme un résumé des taux de réponse au questionnaire de l'Enquête sur la population active (EPA) et à celui de l'Enquête sur le service téléphonique résidentiel (ESTR) de décembre 2010.

Provinces	**Taux de réponse des ménages pour l'EPA complète*** **Décembre 2010**	**Taux de réponse des ménages pour les groupes de renouvellement de l'EPA (Ontario et Québec groupes 1 et 2 et les autres provinces 1, 2 et 4)***	**Taux de réponse des ménages pour l'ESTR****	**Ménages répondants à l'ESTR**
	%			
Terre-Neuve-et-Labrador	92,8	92,6	94,6	863
Île-du-Prince-Édouard	91,7	92,1	93,3	573
Nouvelle-Écosse	92,5	92,7	92,4	1 294
Nouveau-Brunswick	92,9	93,4	93,3	1 256
Québec	90,7	91,1	93,2	2 845
Ontario	87,3	88,2	90,1	4 263
Manitoba	90,6	90,7	91,7	1 978
Saskatchewan	90,6	90,9	93,3	1 701
Alberta	86,8	87,5	88,5	2 148
Colombie-Britannique	88,1	88,9	84,4	2 389
Canada	**89,3**	**90,0**	**90,7**	**19 310**

* Le taux de réponse à l'EPA est le nombre de ménages ayant répondu au questionnaire exprimé sous forme de pourcentage du nombre de ménages admissibles.

** Le taux de réponse à l'ESTR est le nombre de ménages ayant répondu au questionnaire de l'ESTR exprimé sous forme de pourcentage du nombre de ménages ayant répondu au questionnaire de l'EPA dans le cadre des groupes de renouvellement échantillonnés.

8.2 Erreurs relatives à l'enquête

Les estimations calculées à partir de cette enquête reposent sur un échantillon de ménages. Des estimations légèrement différentes auraient pu être obtenues si un recensement complet avait été effectué en reprenant le même questionnaire et en faisant appel aux mêmes intervieweurs, superviseurs, méthodes de traitement, etc. que ceux effectivement utilisés dans l'enquête. L'écart entre les estimations découlant de l'échantillon et celles que donnerait un dénombrement complet réalisé dans des conditions semblables est appelé erreur d'échantillonnage de l'estimation.

Des erreurs qui ne sont pas liées à l'échantillonnage peuvent se produire à presque toutes les étapes des opérations d'enquête. Les intervieweurs peuvent avoir mal compris les instructions, les répondants peuvent se tromper en répondant aux questions, les réponses peuvent être mal saisies sur le questionnaire et des erreurs peuvent survenir lors du traitement et de la totalisation

des données. Ces erreurs sont toutes des exemples d'erreurs non dues à l'échantillonnage.

Sur un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations calculées à partir de l'enquête. Toutefois, les erreurs systématiques contribuent à biaiser les estimations de l'enquête. Énormément de temps et d'efforts ont été consacrés à réduire les erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été prises à chacune des étapes du cycle de collecte et de traitement des données afin de contrôler la qualité des données. Ces mesures comprenaient le recours à des intervieweurs hautement qualifiés, une formation poussée des intervieweurs concernant les procédures d'enquête et du questionnaire, l'observation des intervieweurs en vue de cerner les problèmes liés à la conception du questionnaire ou à une mauvaise compréhension des instructions, des procédures visant à s'assurer que les erreurs de saisie des données étaient réduites au minimum ainsi que des vérifications de la qualité du codage et de contrôle ayant pour but d'attester la logique du traitement.

8.2.1 Base de sondage

Comme l'Enquête sur le service téléphonique résidentiel était une enquête supplémentaire à l'EPA, la base de sondage de l'EPA fut utilisée. Toute non-réponse à l'EPA avait un impact sur la base de sondage de l'ESTR. Comme la non-réponse à l'EPA est très basse (généralement moins de 5 %), cet impact a été minime. La qualité des variables d'échantillon dans la base était très élevée. L'échantillon de l'ESTR était composé de trois groupes de renouvellement de l'EPA. Aucun enregistrement n'a été éliminé dû à un numéro de groupe de renouvellement ou à toute autre variable d'échantillon manquant.

Il est à noter que la base de sondage de l'EPA exclut environ 2 % de tous les ménages dans les 10 provinces du Canada. La base de sondage de l'ESTR exclut donc la même proportion de ménages dans les mêmes régions géographiques. Il est peu probable que cette exclusion introduise un biais important dans les données de l'enquête.

Toutes les variables dans la base de sondage de l'EPA sont mises à jour tous les mois.

Certaines variables incluses dans la base de sondage jouent un rôle crucial relativement à l'application logicielle utilisée dans le cadre de l'enquête. Dans le cas d'une application d'interview assistée par ordinateur (IAO) de l'ESTR, par exemple, chaque enregistrement doit comporter un code de strate, de grappes et de groupe de renouvellement exacts. Ces variables sont toujours de très haute qualité à l'EPA à tous les mois.

Il arrive parfois que des enregistrements en double soient reçus. On a dénombré 256 enregistrements en double dans les données de la collecte du mois de décembre 2010.

8.2.2 Collecte des données

La formation des intervieweurs pour l'ESTR consistait à lire le Manuel des procédures, la pratique avec des cas simulés de l'ESTR sur ordinateur et une discussion avec l'intervieweur principal des questions de l'intervieweur avant le début de l'enquête. L'information fournie aux intervieweurs comprenait une description du contexte et des objectifs de l'enquête ainsi qu'un glossaire de terminologie et une série de questions et de réponses. Les intervieweurs ont recueilli l'information pour l'ESTR après avoir recueilli l'information de l'EPA. La période de collecte allait du 12 au 21 décembre 2010.

8.2.3 Traitement des données

Au cours du traitement des données de l'ESTR, 399 enregistrements n'ont pu être jumelés à un enregistrement équivalent sur l'EPA. Ils ont donc été codés comme étant

hors du champ d'enquête et ont été retirés de tout traitement additionnel. Lorsque des enregistrements de l'enquête supplémentaire ne peuvent être jumelés à des enregistrements de l'enquête hôte, ils doivent être supprimés car aucune pondération ne pourra être obtenue pour ceux-ci.

À l'inverse, 252 enregistrements de l'EPA auraient dû être jumelés à un enregistrement de l'ESTR mais ne l'ont pas été. Ces enregistrements ont été codés comme faisant partie du champ de l'enquête car ils étaient éligibles au supplément dans la base de sondage. Pour une raison quelconque ils n'avaient pas d'enregistrements de l'ESTR correspondants. Ils ont été traités comme non-réponse et ont été utilisés dans le processus de pondération pour faire l'ajustement pour la non-réponse.

Le traitement des données pour l'ESTR fût assez simple car il n'y avait que neuf questions dans l'application IAO. Tout enregistrement qui contenait une réponse « Refus » ou « Ne sait pas » aux questions RTS_Q01 et RTS_Q05, a été codé comme non-réponse. Il est à noter que 3 611 enregistrements ont été codés de la sorte. Comme les données ont été recueillies à l'aide d'un instrument IAO, la qualité des données avant traitement était très bonne. Très peu de changements ont été faits aux données lors du traitement.

8.2.4 Non-réponse

L'une des principales sources d'erreurs non dues à l'échantillonnage observées dans le cadre des enquêtes est l'effet de la non-réponse sur les résultats de l'enquête. L'étendue d'une non-réponse varie d'une non-réponse partielle (le fait de ne pas répondre simplement à une ou à des questions) à une non-réponse totale. S'il y a eu non-réponse totale, c'est parce que l'intervieweur a été incapable de communiquer avec le répondant, qu'aucun membre du ménage n'a pu fournir l'information demandée ou que le répondant a refusé de participer à l'enquête. Les non-réponses totales ont été traitées en ajustant le poids des ménages ayant répondu au questionnaire d'enquête de façon à le contrebalancer pour ceux n'ayant pas répondu.

Dans la plupart des cas, il y a eu non-réponse partielle au questionnaire d'enquête lorsque le répondant n'a pas compris ou a mal interprété une question, a refusé d'y répondre ou ne pouvait se rappeler l'information demandée.

La non-réponse par question est normalement très basse pour l'ESTR. Pour les questions RTS_Q01 à RTS_Q09 le taux de non-réponse a varié entre 0,088 % pour la question RTS_Q01 et 0,725 % pour la question RTS_Q08.

8.2.5 Mesure de l'erreur d'échantillonnage

Puisqu'il est inévitable que des estimations établies à partir d'une enquête-échantillon (ou par sondage) soient sujettes à une erreur d'échantillonnage, une saine pratique de la statistique exige que les chercheurs fournissent aux utilisateurs une certaine indication de l'importance de cette erreur d'échantillonnage. Cette section de la documentation renferme un aperçu des mesures de l'erreur d'échantillonnage dont Statistique Canada se sert couramment et dont le Bureau conseille vivement aux utilisateurs qui produisent des estimations à partir de ce fichier de microdonnées à employer également.

La base pour mesurer l'importance potentielle des erreurs d'échantillonnage est l'erreur-type des estimations calculées à partir des résultats d'une enquête.

En raison, cependant, de la diversité des estimations pouvant être produites à partir d'une enquête, l'erreur-type d'une estimation est habituellement exprimée en fonction de l'estimation à laquelle elle se rapporte. La mesure résultante, appelée coefficient de

variation (CV) d'une estimation, s'obtient en divisant l'erreur-type de l'estimation par l'estimation elle-même et s'exprime en pourcentage de l'estimation.

Par exemple, supposons que, d'après les résultats de l'enquête de novembre 2002, l'on estime que 1,5 % des ménages canadiens n'avaient pas de service téléphonique à leur résidence (RTS_Q01B = 1, Aucun) et l'on constate que l'erreur-type de cette estimation est de 0,00092. Le coefficient de variation de l'estimation est donc calculé comme suit :

$$\left(\frac{0,00092}{0,015} \right) X\ 100\ \% = 6,1\ \%$$

De plus amples renseignements sur le calcul du coefficient de variation, se trouvent au chapitre 10.0.

9.0 Lignes directrices pour la totalisation, l'analyse et la diffusion de données

Ce chapitre de la documentation renferme un aperçu des lignes directrices que doivent respecter les utilisateurs qui totalisent, analysent, publient ou autrement diffusent des données calculées à partir des fichiers de microdonnées de l'enquête. Ces lignes directrices devraient permettre aux utilisateurs de microdonnées de produire les mêmes chiffres que ceux produits par Statistique Canada, tout en étant en mesure d'obtenir des chiffres actuellement inédits de façon conforme à ces lignes directrices établies.

9.1 Lignes directrices pour l'arrondissement d'estimations

Afin que les estimations qui sont destinées à la publication ou à toute autre forme de diffusion qui sont calculées à partir de ces fichiers de microdonnées correspondent à celles produites par Statistique Canada, nous conseillons vivement aux utilisateurs de respecter les lignes directrices qui suivent en ce qui concerne l'arrondissement de telles estimations :

a) Les estimations dans le corps principal d'un tableau statistique doivent être arrondies à la centaine près à l'aide de la technique d'arrondissement normale. Selon cette technique, si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, le dernier chiffre à conserver est augmenté de 1. Par exemple, selon la technique d'arrondissement normale à la centaine près, si les deux derniers chiffres se situent entre 00 et 49, ils sont remplacés par 00 et le chiffre précédent (le chiffre des centaines) reste inchangé. Si les derniers chiffres se situent entre 50 et 99, ils sont remplacés par 00 et le chiffre précédent est augmenté de 1.

b) Les totaux partiels marginaux et les totaux marginaux des tableaux statistiques doivent être calculés à partir de leurs composantes non arrondies correspondantes, puis ensuite être arrondis à leur tour à la centaine près à l'aide de la technique d'arrondissement normale.

c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir de composantes non arrondies (c'est-à-dire des numérateurs et/ou des dénominateurs), puis être arrondis à leur tour à une décimale à l'aide de la technique d'arrondissement normale. Dans le cas d'un arrondissement normal à un seul chiffre, si le dernier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, le dernier chiffre à conserver est augmenté de 1.

d) Les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs composantes non arrondies correspondantes, puis être arrondies à leur tour à la centaine près (ou à la décimale près) à l'aide de la technique d'arrondissement normale.

e) Dans les cas, où, en raison de limitations d'ordre technique ou de toutes autres limites, une technique d'arrondissement autre que la technique normale est utilisée produisant des estimations à être publiées ou autrement diffusées différentes des estimations correspondantes publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces différences dans le ou les documents à publier ou à diffuser.

f) En aucun cas, les utilisateurs ne doivent publier ou autrement diffuser des estimations non arrondies. Des estimations non arrondies laissent entendre qu'elles sont plus précises qu'elles ne le sont en réalité.

9.2 Lignes directrices pour la pondération de l'échantillon en vue de la totalisation

Le plan d'échantillonnage utilisé pour l'Enquête sur le service téléphonique résidentiel (ESTR) n'était pas autopondéré. Lorsqu'ils produisent des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent appliquer le poids d'échantillonnage approprié.

Si les poids appropriés ne sont pas utilisés, les estimations calculées à partir des fichiers de microdonnées ne peuvent être considérées comme représentatives de la population visée par l'enquête et ne correspondront pas à celles produites par Statistique Canada.

Les utilisateurs devraient également prendre note que certains progiciels pourraient peut-être ne pas permettre la production d'estimations correspondant exactement à celles qu'offre Statistique Canada, en raison du mode de traitement du champ du poids par ces progiciels.

9.3 Définitions de types d'estimations : catégoriques et quantitatives

Avant de discuter de la façon dont on peut totaliser et analyser les données de l'ESTR, il est utile de décrire les deux principaux types d'estimations ponctuelles des caractéristiques de la population qui peuvent être produites à partir du fichier de microdonnées créé pour l'ESTR.

9.3.1 Estimations catégoriques

Les estimations catégoriques sont des estimations du nombre ou du pourcentage de membres de la population visée par l'enquête possédant certaines caractéristiques ou faisant partie d'une catégorie définie. Le nombre de ménages qui n'avaient pas le service téléphonique à leur résidence durant le mois de référence ou la proportion de ménages qui avaient deux lignes téléphoniques ou plus constituent des exemples de telles estimations. Une estimation du nombre de ménages possédant une certaine caractéristique peut aussi être désignée une estimation d'un agrégat.

Exemple de question catégorique :

Q : Y a-t-il quelqu'un dans votre ménage qui a actuellement un téléphone cellulaire actif? Veuillez inclure les téléphones cellulaires utilisés à des fins commerciales.
R : Oui / Non

9.3.2 Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes ou d'autres mesures d'une tendance centrale de quantités reposant sur certains ou sur tous les membres de la population visée par l'enquête. Elles comprennent aussi expressément des estimations de la forme $\hat{X}/\hat{Y}$ où $\hat{X}$ est une estimation de la quantité totale pour la population visée par l'enquête et $\hat{Y}$, est une estimation du nombre de ménages dans la population visée par l'enquête qui contribuent à cette quantité totale.

Un exemple d'estimation quantitative est le nombre moyen de téléphones cellulaires par ménage ayant déclaré au moins un service de téléphonie cellulaire actif. Le numérateur est une estimation du nombre total de téléphones cellulaires enregistrés à la question RTS_Q02A, et son dénominateur est le nombre de ménages ayant déclaré au moins un service de téléphonie cellulaire actif.

Exemple de question quantitative :

Q : Combien de téléphones cellulaires actifs les membres de votre ménage ont-ils?
R : |_|_| le nombre de téléphone cellulaire actif

9.3.3 Totalisation d'estimations catégoriques

On peut obtenir des estimations du nombre de gens possédant une certaine caractéristique à partir du fichier de microdonnées en additionnant les poids finals de tous les enregistrements possédant la ou les caractéristiques qui nous intéressent. On obtient des proportions et des rapports de la forme $\hat{X} / \hat{Y}$ en :

a) additionnant les poids finals des enregistrements présentant la caractéristique qui nous intéresse pour le numérateur $\left(\hat{X}\right)$,
b) additionnant les poids finals des enregistrements présentant la caractéristique qui nous intéresse pour le dénominateur $\left(\hat{Y}\right)$, puis en
c) divisant l'estimation a) par celle de b) $\left(\hat{X} / \hat{Y}\right)$.

9.3.4 Totalisation d'estimations quantitatives

On peut obtenir des estimations de quantités à partir du fichier de microdonnées en multipliant la valeur de la variable qui nous intéresse par le poids final de chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements qui nous intéressent. Pour obtenir, par exemple, une estimation du nombre total de téléphones cellulaires utilisés par les ménages uniquement à des fins commerciales, multipliez la valeur déclarée à la question RTS_Q04 (nombre de téléphones cellulaires utilisés uniquement à des fins commerciales) par le poids final de l'enregistrement, puis additionnez cette valeur pour tous les enregistrements où la variable RTS_Q01 = 1 (ménages ayant déclaré au moins un service de téléphonie cellulaire actif).

Pour obtenir une moyenne pondérée de la forme $\hat{X} / \hat{Y}$, le numérateur $\left(\hat{X}\right)$ est calculé comme une estimation quantitative et le dénominateur $\left(\hat{Y}\right)$ est calculé comme une estimation catégorique. Pour estimer, par exemple, le nombre moyen de téléphones cellulaires utilisés par les ménages uniquement à des fins commerciales,

a) estimez le nombre total de téléphones cellulaires utilisés par les ménages uniquement à des fins commerciales $\left(\hat{X}\right)$ tel qu'il est décrit ci-dessus,
b) estimez le nombre de ménages ayant déclaré au moins un service de téléphonie cellulaire actif $\left(\hat{Y}\right)$ incluses dans cette catégorie en additionnant les poids finals de tous les enregistrements où la variable RTS_Q01 = 1, puis
c) divisez l'estimation a) par l'estimation b) $\left(\hat{X} / \hat{Y}\right)$.

9.4 Lignes directrices pour l'analyse statistique

L'Enquête sur le service téléphonique résidentiel repose sur un plan d'échantillonnage complexe comportant une stratification, de multiples étapes de sélection ainsi que des probabilités inégales de sélection des répondants. L'utilisation des données provenant d'enquêtes aussi complexes présente des problèmes pour les analystes, parce que le plan d'enquête et les probabilités de

sélection influent sur les procédures d'estimation et de calcul de la variance qui devraient être utilisées. Il faut utiliser les poids de l'enquête pour que les estimations et les analyses des données de l'enquête soient exemptes de biais.

Bien que de nombreuses procédures d'analyse que l'on trouve à l'intérieur de progiciels statistiques permettent d'utiliser des poids, la signification ou la définition du poids inclus dans ces procédures peut différer de ce qui convient dans le contexte d'une enquête-échantillon, de telle sorte que dans bien des cas les estimations produites au moyen de ces progiciels sont correctes, mais que les variances calculées sont piètres. Les variances approximatives pour des estimations simples comme des totaux, des proportions et des rapports (pour des variables qualitatives) peuvent être calculées à partir des tables de variabilité d'échantillonnage approximative qui accompagnent les données.

Pour d'autres techniques d'analyse (de régression linéaire, de régression logistique et de l'analyse de variance, par exemple), il existe une méthode qui peut rendre les variances calculées par l'application des progiciels normalisés plus significatives, en intégrant les probabilités inégales de sélection. L'application de cette méthode entraîne une remise à l'échelle des poids de façon à ce que le poids moyen soit de 1.

Supposons, par exemple, qu'il faut effectuer l'analyse de tous les ménages du Québec. Les étapes à suivre pour remettre à l'échelle les poids sont les suivantes :

1) sélectionner tous les ménages du fichier qui ont déclaré PROVINCE = 24, Québec;
2) calculer le poids MOYEN pour ces enregistrements en additionnant les poids originaux des ménages établis à partir du fichier de microdonnées pour ces enregistrements puis diviser cette somme par le nombre de ménages ayant déclaré PROVINCE = 24;
3) pour chacun de ces enregistrements, calculer un poids REMIS À L'ÉCHELLE égal au poids original du ménage divisé par le poids MOYEN;
4) effectuer l'analyse portant sur ces ménages en utilisant le poids REMIS À L'ÉCHELLE.

Parce qu'on ne tient toujours compte ni de la stratification ni des grappes du plan d'échantillonnage, les estimations des variances calculées avec cette méthode risquent cependant d'être des sous-estimations.

Il faut connaître les détails du plan d'enquête pour calculer des estimations des variances plus précises. De tels détails ne peuvent être fournis dans le fichier de microdonnées en raison de la confidentialité. Statistique Canada peut, contre remboursement des frais, calculer des variances qui tiennent compte du plan complet d'échantillonnage pour beaucoup de statistiques.

9.5 Lignes directrices pour la diffusion de coefficients de variation

Avant de diffuser et/ou de publier toutes estimations établies à partir de l'Enquête sur le service téléphonique résidentiel, les utilisateurs devraient premièrement déterminer le niveau de qualité de cette estimation. Les niveaux de qualité sont *acceptable, médiocre* et *inacceptable*. Les erreurs d'échantillonnage et non dues à l'échantillonnage, dont il a été question au chapitre 8.0, influencent la qualité des données. Aux fins du présent document, cependant, on ne déterminera le niveau de qualité d'une estimation qu'à partir d'une erreur d'échantillonnage dont rend compte le coefficient de variation indiqué à l'intérieur du tableau qui figure ci-dessous. Les utilisateurs devraient néanmoins s'assurer de lire le chapitre 8.0 pour être plus pleinement informés des caractéristiques relatives à la qualité de ces données.

On devrait premièrement déterminer le nombre d'enregistrements de ménages retenus pour le calcul de l'estimation. Si ce nombre est inférieur à 30, il faudrait considérer l'estimation pondérée comme étant de qualité inacceptable.

Pour les estimations pondérées fondées sur les tailles d'échantillon de 30 ou plus, les utilisateurs devraient déterminer le coefficient de variation de l'estimation et suivre les lignes directrices relatives au niveau de qualité qui figurent ci-dessous. Celles-ci devraient être appliquées, pour la détermination du niveau de qualité d'une estimation, aux estimations pondérées arrondies.

On peut considérer qu'il est possible de divulguer toutes les estimations. Celles d'un niveau de qualité médiocre ou inacceptable doivent cependant être accompagnées d'une mise en garde pour avertir les utilisateurs subséquents.

Lignes directrices relatives au niveau de qualité de l'estimation

Niveau de qualité de l'estimation	**Lignes directrices**
1) Acceptable	Les estimations proviennent d'une taille d'échantillon de 30 ou plus, et présentent de faibles coefficients de variation, de l'ordre de 0,0 à 16,5 %. Aucune mise en garde n'est requise.
2) Médiocre	Les estimations proviennent d'une taille d'échantillon de 30 ou plus, et présentent des coefficients de variation élevés, de l'ordre de 16,6 à 33,3 %. Ces estimations devraient être signalées par la lettre E (ou un quelconque identificateur similaire). Elles devraient être accompagnées d'une mise en garde avertissant les utilisateurs subséquents des niveaux élevés d'erreur associés aux estimations.
3) Inacceptable	Les estimations proviennent d'une taille d'échantillon inférieure à 30, ou présentent des coefficients de variation très élevés, supérieurs à 33,3 %. Statistique Canada recommande de ne pas diffuser d'estimations de qualité inacceptable. Si un utilisateur choisit cependant de le faire, ces estimations devraient alors être signalées à l'aide de la lettre F (ou d'un quelconque identificateur similaire) et devraient être accompagnées de la mise en garde suivante : « Nous informons l'utilisateur que ces estimations (désignées avec la lettre F) ne respectent pas les normes de qualité de Statistique Canada. Les conclusions qui reposeront sur ces données ne seront pas fiables et seront très probablement invalides. »

9.6 Seuils pour la diffusion des estimations pour l'Enquête sur le service téléphonique résidentiel

Le tableau ci-dessous fournit une indication de la précision des estimations des prévisions démographiques ainsi que les seuils de diffusion associés aux trois niveaux de qualité de l'estimation présentés à la section précédente. Ces seuils proviennent des tables de coefficients de variation (CV) dont il sera question au chapitre 10.0.

Par exemple, d'après le tableau, la qualité d'une estimation pondérée de 10 000 ménages possédant une caractéristique donnée à Terre-Neuve-et-Labrador est médiocre.

Veuillez noter que ces seuils de diffusion correspondent aux estimations de chiffres de population seulement. Dans le cas d'estimations de rapports, les utilisateurs ne devraient pas utiliser la valeur du numérateur (ni le dénominateur) afin de trouver le niveau de qualité de l'estimation correspondant. La règle 4 à la section 10.1 ainsi que l'exemple 4 à la section 10.1.1 expliquent la bonne procédure à suivre dans le cas d'un rapport.

Provinces et régions	**CV acceptable 0,0 à 16,5 %**	**CV médiocre 16,6 à 33,3 %**	**CV inacceptable > 33,3 %**
Terre-Neuve-et-Labrador	10 200 et plus	2 600 à < 10 200	moins de 2 600
Île-du-Prince-Édouard	4 000 et plus	1 000 à < 4 000	moins de 1 000
Nouvelle-Écosse	14 200 et plus	3 600 à < 14 200	moins de 3 600
Nouveau-Brunswick	10 800 et plus	2 700 à < 10 800	moins de 2 700
Québec	94 100 et plus	23 600 à < 94 100	moins de 23 600
Ontario	85 000 et plus	21 100 à < 85 000	moins de 21 100
Manitoba	15 400 et plus	3 900 à < 15 400	moins de 3 900
Saskatchewan	13 300 et plus	3 400 à < 13 300	moins de 3 400
Alberta	37 700 et plus	9 500 à < 37 700	moins de 9 500
Colombie-Britannique	43 300 et plus	10 800 à < 43 300	moins de 10 800
Provinces de l'Atlantique	12 000 et plus	3 000 à < 12 000	moins de 3 000
Provinces des Prairies	30 000 et plus	7 400 à < 30 000	moins de 7 400
Canada	**57 100 et plus**	**14 100 à < 57 100**	**moins de 14 100**

10.0 Tables de variabilité d'échantillonnage approximative

Afin de fournir des coefficients de variation (CV) qui pourraient s'appliquer à une gamme étendue d'estimations catégoriques produites à partir de ce fichier de microdonnées et auxquels il serait facilement possible pour l'utilisateur d'avoir accès, un ensemble de tables de variabilité d'échantillonnage approximative a été produit. Ces tables de CV permettent à l'utilisateur d'obtenir un coefficient de variation approximatif fondé sur la taille de l'estimation calculée à partir des données de l'enquête.

Les coefficients de variation sont calculés à l'aide de la formule de la variance pour un échantillonnage aléatoire simple et en y incorporant un facteur qui reflète la nature du plan d'échantillonnage, qui est à plusieurs degrés et qui prévoit la formation de grappes. Ce facteur, appelé l'effet du plan, a été déterminé en calculant premièrement les effets du plan pour une gamme étendue de caractéristiques, puis en choisissant parmi ceux-ci une valeur modérée (habituellement le 75^{e} percentile) à utiliser à l'intérieur des tables de CV qui s'appliqueraient ensuite à l'ensemble entier des caractéristiques.

Le tableau ci-dessous indique la valeur modérée des effets du plan, ainsi que les tailles de l'échantillon et les chiffres de population selon la province qui ont été utilisés pour produire les tables de variabilité d'échantillonnage approximative de l'Enquête sur le service téléphonique résidentiel (ESTR).

Provinces et régions	**Effet du plan**	**Taille de l'échantillon**	**Population**
Terre-Neuve-et-Labrador	1,22	863	207 431
Île-du-Prince-Édouard	1,17	573	58 493
Nouvelle-Écosse	1,31	1 294	396 978
Nouveau-Brunswick	1,21	1 256	317 024
Québec	2,19	2 845	3 425 042
Ontario	1,96	4 263	5 121 451
Manitoba	1,83	1 978	470 186
Saskatchewan	1,56	1 701	410 903
Alberta	1,60	2 148	1 419 224
Colombie-Britannique	1,58	2 389	1 826 430
Provinces de l'Atlantique	1,35	3 986	979 927
Provinces des Prairies	2,10	5 827	2 300 312
Canada	**2,21**	**19 310**	**13 653 162**

Tous les coefficients de variation inclus dans les tables de variabilité d'échantillonnage approximative sont approximatifs et donc non officiels. Des estimations de la variance réelle pour des variables précises peuvent être obtenues auprès de Statistique Canada, contre remboursement des frais. Étant donné que le CV approximatif est une estimation prudente, l'utilisation de la variance réelle estimée pourrait faire passer l'estimation d'un niveau de qualité à un autre. Par exemple, une estimation *médiocre* pourrait devenir *acceptable* si elle était fondée sur le calcul du CV exact.

Rappelez-vous que : Si le nombre d'observations sur lesquelles une estimation est basée est inférieur à 30, l'estimation pondérée est très probablement inacceptable et Statistique Canada recommande de ne pas diffuser une telle estimation, quelle que soit la valeur du coefficient de variation.

10.1 Comment utiliser les tables de coefficients de variation pour des estimations catégoriques

Les règles qui suivent devraient permettre à l'utilisateur de déterminer les coefficients de variation approximatifs à partir des tables de variabilité d'échantillonnage approximative pour des estimations du nombre, de la proportion ou du pourcentage de membres de la population visée par l'enquête possédant une certaine caractéristique et pour des rapports et des différences entre de telles estimations.

Règle 1 : Estimations du nombre de ménages possédant une caractéristique donnée (agrégats)

Le coefficient de variation dépend uniquement de la taille de l'estimation elle-même. Dans la table de variabilité d'échantillonnage approximative pour la région géographique appropriée, repérez le nombre estimé dans la colonne la plus à gauche (intitulée « Numérateur du pourcentage ») et suivez les astérisques (le cas échéant) jusqu'au premier chiffre rencontré. Ce chiffre est le coefficient de variation approximatif.

Règle 2 : Estimations de proportions ou de pourcentages de ménages possédant une caractéristique donnée

Le coefficient de variation d'une proportion estimée ou d'un pourcentage estimé dépend à la fois de la taille de la proportion ou du pourcentage et de la taille du total sur lequel la proportion ou le pourcentage repose. Les proportions estimées ou les pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque la proportion ou que le pourcentage repose sur un sous-groupe de la population. La proportion, par exemple, de ménages qui n'avaient pas de service téléphonique dans leur résidence au cours de la période de référence est plus fiable que le nombre estimé de ménages qui n'avaient pas de service téléphonique dans leur résidence au cours de la période de référence. (Remarquez que dans les tables la valeur des coefficients de variation diminue lorsqu'on les lit de gauche à droite.)

Lorsque la proportion ou que le pourcentage repose sur la population totale de la région géographique visée par la table, le CV de la proportion ou du pourcentage est le même que le CV du numérateur de la proportion ou du pourcentage. Dans ce cas, la règle 1 peut être appliquée.

Lorsque la proportion ou que le pourcentage repose sur un sous-ensemble de la population totale (p. ex., comme ses membres d'une province ou d'une région métropolitaine de recensement particulière), on devrait faire référence à la proportion ou au pourcentage (dans le haut de la table) et au numérateur de la proportion ou du pourcentage (dans la colonne de gauche de la table). L'intersection de la rangée et de la colonne appropriées donne le coefficient de variation.

Règle 3 : Estimations de différences entre des agrégats ou des pourcentages

L'erreur-type d'une différence entre deux estimations est approximativement égale à la racine carrée de la somme des carrés de chaque erreur-type considérée séparément. C'est-à-dire que l'erreur-type d'une différence $\left(\hat{d} = \hat{X}_1 - \hat{X}_2\right)$ est :

$$\sigma_{\hat{d}} = \sqrt{\left(\hat{X}_1 \alpha_1\right)^2 + \left(\hat{X}_2 \alpha_2\right)^2}$$

où $\hat{X}_1$ est l'estimation 1, $\hat{X}_2$ est l'estimation 2 et α_1 et α_2 sont les coefficients de variation de $\hat{X}_1$ et $\hat{X}_2$ respectivement. Le coefficient de variation de $\hat{d}$ est donné par $\sigma_{\hat{d}} / \hat{d}$. Cette formule est exacte pour la différence entre des caractéristiques distinctes et non corrélées, mais

n'est autrement qu'approximative.

Règle 4 : Estimations de rapports

Dans le cas où le numérateur est un sous-ensemble du dénominateur, le rapport devrait être converti en un pourcentage et la règle 2 appliquée. Cela s'appliquerait, par exemple, au cas où le dénominateur est le nombre de ménages qui n'avaient pas de service téléphonique dans leur résidence au cours de la période de référence et le numérateur, le nombre de ménages qui n'avait pas de service téléphonique dans leur résidence au cours de la période de référence parce qu'ils n'en avaient pas les moyens.

Dans le cas où le numérateur n'est pas un sous-ensemble du dénominateur, comme dans l'exemple du rapport du nombre de ménages au Québec avec un téléphone cellulaire comparativement au nombre de ménages en Ontario avec un téléphone cellulaire, l'erreur-type du rapport des estimations est approximativement égale à la racine carrée de la somme des carrés de chaque coefficient de variation considéré séparément multipliée par $\hat{R}$. C'est-à-dire que l'erreur-type d'un rapport $\left(\hat{R} = \hat{X}_1 / \hat{X}_2\right)$ est :

$$\sigma_{\hat{R}} = \hat{R}\sqrt{\alpha_1^{\,2} + \alpha_2^{\,2}}$$

où α_1 et α_2 sont les coefficients de variation de $\hat{X}_1$ et de $\hat{X}_2$ respectivement. Le coefficient de variation de $\hat{R}$ est donné par $\sigma_{\hat{R}} / \hat{R}$. La formule tendra à surestimer l'erreur si $\hat{X}_1$ et $\hat{X}_2$ sont corrélés positivement et à la sous-estimer si $\hat{X}_1$ et $\hat{X}_2$ sont corrélés négativement.

Règle 5 : Estimations de différences entre des rapports

Dans ce cas, les règles 3 et 4 sont combinées. On détermine premièrement les CV pour les deux rapports à l'aide de la règle 4, puis on trouve le CV de leur différence au moyen de la règle 3.

10.1.1 Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques

Les exemples ci-dessous utilisent des données du fichier de l'ESTR de novembre 2002 et sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter. Veuillez noter que les données utilisées dans ces exemples diffèrent des véritables résultats de l'enquête et tiennent seulement lieu de guide.

Exemple 1 : Estimations du nombre de ménages possédant une caractéristique donnée (agrégats)

Supposons qu'un utilisateur estime à 177 859 le nombre de ménages qui n'avaient pas de service téléphonique dans leur résidence au cours de la période de référence (Q01B = 1, Aucun). Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

1) Reportez-vous à la table des coefficients de variation pour le CANADA.

2) L'agrégat estimé (177 859) ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, c'est-à-dire 200 000.

3) On trouve le coefficient de variation pour un agrégat estimé en se reportant à la première entrée autre que des astérisques sur cette rangée, c'est-à-dire 5,6 %.

4) Le coefficient de variation approximatif de l'estimation est donc 5,6 %. Le résultat selon lequel 177 859 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1) ménages n'avaient pas de service téléphonique dans leur résidence au cours de la période de référence, peut être publié sans réserve.

ENQUÊTE SUR LE SERVICE TÉLÉPHONIQUE RÉSIDENTIEL - NOVEMBRE 2002

Tables de variabilité d'échantillonnage approximative — Canada

NUMÉRATEUR DU POURCENTAGE	POURCENTAGE ESTIMÉ												
('000)	**0.1%**	**1.0%**	**2.0%**	**5.0%**	**10.0%**	**15.0%**	**20.0%**	**…**	**35.0%**	**40.0%**	**50.0%**	**70.0%**	**90.0%**
1	79.3	79.0	78.6	77.4	75.3	73.2	71.0		64.0	61.5	56.1	43.5	25.1
2	56.1	55.8	55.6	54.7	53.2	51.7	50.2		45.3	43.5	39.7	30.7	17.7
3	45.8	45.6	45.4	44.7	43.5	42.3	41.0		36.9	35.5	32.4	25.1	14.5
4	39.7	39.5	39.3	38.7	37.7	36.6	35.5		32.0	30.7	28.1	21.7	12.6
5	35.5	35.3	35.1	34.6	33.7	32.7	31.8		28.6	27.5	25.1	19.4	11.2
.													
.													
.													
50	********	11.2	11.1	10.9	10.6	10.3	10.0		9.1	8.7	7.9	6.1	3.5
55	********	10.6	10.6	10.4	10.2	9.9	9.6		8.6	8.3	7.6	5.9	3.4
60	********	10.2	10.1	10.0	9.7	9.4	9.2		8.3	7.9	7.2	5.6	3.2
65	********	9.8	9.7	9.6	9.3	9.1	8.8		7.9	7.6	7.0	5.4	3.1
70	********	9.4	9.4	9.2	9.0	8.7	8.5		7.6	7.3	6.7	5.2	3.0
75	********	9.1	9.1	8.9	8.7	8.5	8.2		7.4	7.1	6.5	5.0	2.9
80	********	8.8	8.8	8.6	8.4	8.2	7.9		7.2	6.9	6.3	4.9	2.8
85	********	8.6	8.5	8.4	8.2	7.9	7.7		6.9	6.7	6.1	4.7	2.7
90	********	8.3	8.3	8.2	7.9	7.7	7.5		6.7	6.5	5.9	4.6	2.6
95	********	8.1	8.1	7.9	7.7	7.5	7.3		6.6	6.3	5.8	4.5	2.6
100	********	7.9	7.9	7.7	7.5	7.3	7.1		6.4	6.1	**5.6**	4.3	2.5
125	********	********	7.0	6.9	6.7	6.5	6.4		5.7	5.5	5.0	3.9	2.2
150	********	********	6.4	6.3	6.1	6.0	5.8		5.2	5.0	4.6	3.5	2.0
200	********	********	**5.6**	5.5	5.3	5.2	5.0		4.5	4.3	4.0	3.1	1.8
250	********	********	********	4.9	4.8	4.6	4.5		4.0	3.9	3.5	2.7	1.6
300	********	********	********	4.5	4.3	4.2	4.1		3.7	3.5	3.2	2.5	1.4
350	********	********	********	4.1	4.0	3.9	3.8		3.4	3.3	3.0	2.3	1.3
400	********	********	********	3.9	3.8	3.7	3.5		3.2	3.1	2.8	2.2	1.3
450	********	********	********	3.6	3.5	3.4	3.3		3.0	2.9	2.6	2.0	1.2
500	********	********	********	3.5	3.4	3.3	3.2		2.9	2.7	2.5	1.9	1.1
.													
.													
.													
8000	********	********	********	********	********	********	********		********	********	********	0.5	0.3
9000	********	********	********	********	********	********	********		********	********	********	********	0.3
10000	********	********	********	********	********	********	********		********	********	********	********	0.3

NOTE : Pour bien utiliser ces tables, veuillez consulter la documentation sur les microdonnées.

Exemple 2 : Estimations de proportions ou de pourcentages de ménages possédant une caractéristique donnée

Supposons qu'un utilisateur estime à 100 791 / 177 859 = 56,7 % la proportion de ménages qui n'avaient pas de service téléphonique dans leur résidence au cours de la période de référence (Q01B = 1, Aucun) ayant déclaré qu'ils n'en avaient pas les moyens (Q02 = 2, Je n'ai pas les moyens ou Q03 = 1, Je n'ai pas les moyens). Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

1) Reportez-vous à la table des coefficients de variation pour le CANADA (voir ci-dessus).

2) Parce que l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c'est-à-dire les ménages qui n'avaient pas de service téléphonique dans leur résidence au cours de la période de référence), il faut utiliser à la fois le pourcentage (56,7 %) et la portion numérateur du pourcentage (100 791) pour déterminer le coefficient de variation.

3) Le numérateur, 100 791, ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, soit 100 000. De même, l'estimation du pourcentage ne figure dans l'en-tête d'aucune colonne; il faut donc utiliser la proportion qui s'en rapproche le plus, soit 50,0 %.

4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 5,6 %, est le coefficient de variation à employer.

5) Le coefficient de variation approximatif de l'estimation est donc 5,6 %. Le résultat selon lequel 56,7 % des ménages n'avaient pas de service téléphonique dans leur résidence au cours de la période de référence parce qu'ils n'en avaient pas les moyens, peut être publié sans réserve.

Exemple 3 : Estimations de différences entre des agrégats ou des pourcentages

Supposons qu'un utilisateur estime à 796 352 / 3 103 651 = 25,7 % la proportion de ménages au Québec (PROV1 = 24, Québec) ayant déclaré un revenu annuel total inférieur au seuil de faible revenu (SFR) (Q10 = 1, Moins élevé que SFR) par rapport à 858 163 / 4 526 715 = 19,0 % des ménages en Ontario (PROV1 = 35, Ontario) ayant déclaré l'avoir fait pour la même raison (Q10 = 1, Moins élevé que SFR). Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

1) L'utilisation des tables des coefficients de variation du QUÉBEC et de l'ONTARIO de la même façon que celle décrite dans l'exemple 1, donne un CV de l'estimation pour les ménages au Québec de 2,8 % et un CV de l'estimation pour les ménages en Ontario de 3,5 %.

ENQUÊTE SUR LE SERVICE TÉLÉPHONIQUE RÉSIDENTIEL - NOVEMBRE 2002

Tables de variabilité d'échantillonnage approximative — Québec

NUMÉRATEUR DU POURCENTAGE	POURCENTAGE ESTIMÉ													
('000)	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	88.9	88.5	88.1	86.7	84.4	82.0	79.6	77.0	74.4	71.7	68.9	62.9	48.7	28.1
2	62.9	62.6	62.3	61.3	59.7	58.0	56.3	54.5	52.6	50.7	48.7	44.5	34.5	19.9
3	51.3	51.1	50.8	50.1	48.7	47.4	45.9	44.5	43.0	41.4	39.8	36.3	28.1	16.2
4	********	44.3	44.0	43.4	42.2	41.0	39.8	38.5	37.2	35.9	34.5	31.5	24.4	14.1
5	********	39.6	39.4	38.8	37.7	36.7	35.6	34.5	33.3	32.1	30.8	28.1	21.8	12.6
.														
.														
.														
200	********	********	********	********	6.0	5.8	5.6	5.4	5.3	5.1	4.9	4.4	3.4	2.0
250	********	********	********	********	5.3	5.2	5.0	4.9	4.7	4.5	4.4	4.0	3.1	1.8
300	********	********	********	********	4.9	4.7	4.6	4.4	4.3	4.1	4.0	3.6	2.8	1.6
350	********	********	********	********	********	4.4	4.3	4.1	4.0	3.8	3.7	3.4	2.6	1.5
400	********	********	********	********	********	4.1	4.0	3.9	3.7	3.6	3.4	3.1	2.4	1.4
450	********	********	********	********	********	3.9	3.8	3.6	3.5	3.4	3.2	3.0	2.3	1.3
500	********	********	********	********	********	********	3.6	3.4	3.3	3.2	3.1	2.8	2.2	1.3
750	********	********	********	********	********	********	********	**2.8**	2.7	2.6	2.5	2.3	1.8	1.0
1000	********	********	********	********	********	********	********	********	********	2.3	2.2	2.0	1.5	0.9
1500	********	********	********	********	********	********	********	********	********	********	********	1.6	1.3	0.7
2000	********	********	********	********	********	********	********	********	********	********	********	********	1.1	0.6

NOTE : Pour bien utiliser ces tables, veuillez consulter la documentation sur les microdonnées.

ENQUÊTE SUR LE SERVICE TÉLÉPHONIQUE RÉSIDENTIEL - NOVEMBRE 2002

Tables de variabilité d'échantillonnage approximative — Ontario

NUMÉRATEUR DU POURCENTAGE	POURCENTAGE ESTIMÉ													
('000)	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	105.6	105.1	104.6	103.0	100.2	97.4	94.5	91.5	88.4	85.2	81.8	74.7	57.9	33.4
2	74.7	74.3	74.0	72.8	70.9	68.9	66.8	64.7	62.5	60.2	57.9	52.8	40.9	23.6
3	61.0	60.7	60.4	59.5	57.9	56.2	54.6	52.8	51.0	49.2	47.2	43.1	33.4	19.3
.														
.														
.														
250	********	********	********	********	6.3	6.2	6.0	5.8	5.6	5.4	5.2	4.7	3.7	2.1
300	********	********	********	********	5.8	5.6	5.5	5.3	5.1	4.9	4.7	4.3	3.3	1.9
350	********	********	********	********	5.4	5.2	5.1	4.9	4.7	4.6	4.4	4.0	3.1	1.8
400	********	********	********	********	5.0	4.9	4.7	4.6	4.4	4.3	4.1	3.7	2.9	1.7
450	********	********	********	********	4.7	4.6	4.5	4.3	4.2	4.0	3.9	3.5	2.7	1.6
500	********	********	********	********	********	4.4	4.2	4.1	4.0	3.8	3.7	3.3	2.6	1.5
750	********	********	********	********	********	********	**3.5**	3.3	3.2	3.1	3.0	2.7	2.1	1.2
1000	********	********	********	********	********	********	********	2.9	2.8	2.7	2.6	2.4	1.8	1.1
1500	********	********	********	********	********	********	********	********	********	2.2	2.1	1.9	1.5	0.9
2000	********	********	********	********	********	********	********	********	********	********	********	1.7	1.3	0.7
3000	********	********	********	********	********	********	********	********	********	********	********	********	1.1	0.6
4000	********	********	********	********	********	********	********	********	********	********	********	********	********	0.5

NOTE : Pour bien utiliser ces tables, veuillez consulter la documentation sur les microdonnées.

2) En utilisant la règle 3, l'erreur-type d'une différence $\left(\hat{d} = \hat{X}_1 - \hat{X}_2\right)$ est :

$$\sigma_{\hat{d}} = \sqrt{\left(\hat{X}_1\alpha_1\right)^2 + \left(\hat{X}_2\alpha_2\right)^2}$$

où $\hat{X}_1$ est l'estimation 1 (Québec), $\hat{X}_2$ est l'estimation 2 (Ontario) et α_1 et α_2 sont les coefficients de variation de $\hat{X}_1$ et de $\hat{X}_2$ respectivement.

C'est-à-dire que l'erreur-type de la différence $\hat{d}$ = 0,257 - 0,190 = 0,067 est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0{,}257)(0{,}028)]^2 + [(0{,}190)(0{,}035)]^2} \\ &= \sqrt{(0{,}0000517) + (0{,}0000442)} \\ &= 0{,}0098\end{aligned}$$

3) Le coefficient de variation de $\hat{d}$ est donné par $\sigma_{\hat{d}} / \hat{d}$ = 0,0098 / 0,067 = 0,146.

4) Le coefficient de variation approximatif de la différence entre les estimations est donc 14,6 %, ce qui peut être diffusé sans réserve.

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime à 796 352 le nombre de ménages au Québec (PROV1 = 24, Québec) ayant déclaré un revenu annuel total inférieur au seuil de faible revenu (SFR) (Q10 = 1, Moins élevé que SFR), par rapport à 858 163 ménages en Ontario (PROV1 = 35, Ontario) ayant déclaré l'avoir fait pour la même raison (Q10 = 1, Moins élevé que SFR). L'utilisateur est intéressé à comparer l'estimation des ménages du Québec à celle des ménages de l'Ontario sous la forme d'un rapport. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

1) Tout d'abord, cette estimation est une estimation d'un rapport, où le numérateur de l'estimation ($\hat{X}_1$) est le nombre de ménages au Québec ayant déclaré un revenu annuel total inférieur au SFR. Le dénominateur de l'estimation ($\hat{X}_2$) est le nombre de ménages en Ontario ayant déclaré un revenu annuel total inférieur au SFR.

2) Reportez-vous aux tables des coefficients de variation pour le QUÉBEC et pour l'ONTARIO (voir ci-dessus).

3) Le numérateur de cette estimation de rapport est 796 352. Le chiffre qui s'en rapproche le plus est 750 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée dans la table de CV pour le Québec, soit 2,8 %.

4) Le dénominateur de cette estimation de rapport est 858 163. Le chiffre qui s'en rapproche le plus est 750 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée dans la table de CV pour l'Ontario, soit 3,5 %.

5) Le coefficient de variation approximatif de l'estimation du rapport est donc donné par la règle 4, qui est :

$$\alpha_{\hat{R}} = \sqrt{{\alpha_1}^2 + {\alpha_2}^2}$$

où α_1 et α_2 sont les coefficients de variation de $\hat{X}_1$ et de $\hat{X}_2$ respectivement. C'est-à-dire que :

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0{,}028)^2 + (0{,}035)^2} \\ &= \sqrt{0{,}000784 + 0{,}001225} \\ &= 0{,}045\end{aligned}$$

6) Le rapport obtenu entre les ménages du Québec et les ménages de l'Ontario qui ont eu un revenu annuel total sous le SFR est 796 352 / 858 163, c'est-à-dire 0,93:1 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1). Le coefficient de variation de cette estimation est 4,5 %, ce qui fait qu'on peut la diffuser sans réserve.

10.2 Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance

Bien que les coefficients de variation soient beaucoup utilisés, l'intervalle de confiance d'une estimation est une mesure plus intuitivement significative de l'erreur d'échantillonnage. Un intervalle de confiance constitue une déclaration du niveau de confiance selon laquelle la valeur vraie pour la population se situe à l'intérieur d'une gamme précisée de valeurs. Par exemple, un intervalle de confiance de 95 % peut être décrit comme suit :

> Si l'échantillonnage de la population est répété indéfiniment, chaque échantillon menant à un nouvel intervalle de confiance pour une estimation, l'intervalle englobera alors dans 95 % des échantillons la valeur vraie de la population.

En utilisant l'erreur-type d'une estimation, des intervalles de confiance pour des estimations peuvent être obtenues en partant de l'hypothèse qu'aux termes d'un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique donnée de la population se répartiront normalement autour de la valeur vraie de la population. Selon cette hypothèse, il y a environ 68 chances sur 100 que l'écart entre une estimation de l'échantillon et la valeur vraie pour la population soit inférieur à une erreur-type, environ 95 chances sur 100 que l'écart soit inférieur à deux erreurs-types et environ 99 chances sur 100 que l'écart soit inférieur à trois erreurs-types. Ces différents degrés de confiance sont désignés sous le nom de niveaux de confiance.

Des intervalles de confiance pour une estimation $\hat{X}$ sont généralement exprimés sous forme de deux chiffres, un inférieur et un supérieur à l'estimation, comme étant $\left(\hat{X} - k,\ \hat{X} + k\right)$, où k est déterminé suivant le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

Des intervalles de confiance pour une estimation peuvent être calculés directement à partir des tables de variabilité d'échantillonnage approximative, en déterminant d'abord à partir de la table appropriée le coefficient de variation de l'estimation $\hat{X}$, puis en utilisant

la formule suivante pour le convertir à un intervalle de confiance $\left(IC_{\hat{x}}\right)$:

$$IC_{\hat{x}} = \left(\hat{X} - t\hat{X}\alpha_{\hat{x}}, \hat{X} + t\hat{X}\alpha_{\hat{x}}\right)$$

où $\alpha_{\hat{x}}$ est le coefficient de variation déterminé de $\hat{X}$, et

$t =$ 1 si l'on désire un intervalle de confiance de 68 %;
$t =$ 1,6 si l'on désire un intervalle de confiance de 90 %;
$t =$ 2 si l'on désire un intervalle de confiance de 95 %;
$t =$ 2,6 si l'on désire un intervalle de confiance de 99 %.

Nota : Les lignes directrices pour la diffusion des estimations s'appliquent également aux intervalles de confiance. S'il est impossible, par exemple, de diffuser une estimation, on ne peut alors pas non plus communiquer un intervalle de confiance.

10.2.1 Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance

Un intervalle de confiance de 95 % pour la proportion estimée de ménages qui n'avaient pas de service téléphonique dans leur résidence au cours de la période de référence ayant déclaré qu'ils n'en avaient pas les moyens (d'après l'exemple 2 à la section 10.1.1) serait calculé comme suit :

$\hat{X}$ = 56,7 % (ou exprimé sous forme de proportion 0,567)

t = 2

$\alpha_{\hat{x}}$ = 5,6 % (0,056 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.

$IC_{\hat{x}}$ = {0,567 - (2) (0,567) (0,056), 0,567 + (2) (0,567) (0,056)}
$IC_{\hat{x}}$ = {0,567 - 0,064, 0,567 + 0,064}
$IC_{\hat{x}}$ = {0,503, 0,631}

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 50,3 et 63,1 % des ménages qui n'avaient pas de service téléphonique dans leur résidence au cours de la période de référence n'avaient pas les moyens d'en avoir.

10.3 Comment utiliser les tables de coefficients de variation pour effectuer un test t

Des erreurs-types peuvent aussi être utilisés pour effectuer des tests d'hypothèses, une procédure destinée à distinguer des paramètres d'une population à l'aide d'estimations d'un échantillon. Ces estimations peuvent être des chiffres, des moyennes, des pourcentages, des rapports, etc. Les tests peuvent être effectués à divers niveaux de signification, où un niveau de signification est la probabilité de conclure que les caractéristiques sont différentes quand, en fait, elles sont identiques.

Supposons que $\hat{X}_1$ et $\hat{X}_2$ sont des estimations d'un échantillon pour deux caractéristiques qui

nous intéressent. Supposons également que l'erreur-type de la différence $\hat{X}_1 - \hat{X}_2$ est $\sigma_{\hat{d}}$.

Si $t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ se situe entre -2 et 2, aucune conclusion à propos de la différence entre les caractéristiques n'est alors justifiée au niveau de signification de 5 %. Si, cependant, ce rapport est inférieur à -2 ou supérieur à +2, la différence observée est significative au niveau de 0,05. C'est-à-dire que la différence entre les estimations est significative.

10.3.1 Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t

Supposons que l'utilisateur désire tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y a pas de différence entre la proportion de ménages au Québec qui ont rapporté un revenu annuel total inférieur au seuil de faible revenu (SFR) et la proportion de ménages en Ontario qui ont aussi rapporté un revenu annuel total inférieur au SFR. D'après l'exemple 3 à la section 10.1.1, il s'est avéré que l'erreur-type de la différence entre ces deux estimations était 0,0098. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0{,}257 - 0{,}190}{0{,}0098} = \frac{0{,}067}{0{,}0098} = 6{,}84$$

Puisque *t* = 6,84 est supérieur à 2, il faut en conclure qu'il existe une différence significative entre les deux estimations au niveau de signification de 0,05.

10.4 Coefficients de variation pour des estimations quantitatives

Il faudrait produire des tables spéciales afin de déterminer l'erreur d'échantillonnage d'estimations quantitatives, ce qui n'a pas été fait, parce que toutes les variables pour l'Enquête sur le service téléphonique résidentiel sont principalement de nature catégorique.

En général cependant, le coefficient de variation d'un total quantitatif sera supérieur au coefficient de variation de l'estimation de la catégorie correspondante (c'est-à-dire l'estimation du nombre de personnes retenues dans l'estimation quantitative). S'il est impossible de diffuser l'estimation de la catégorie correspondante, on ne pourra pas non plus communiquer l'estimation quantitative. Par exemple, le coefficient de variation du nombre total de semaines d'absence du travail serait supérieur au coefficient de variation de la proportion correspondante de travailleurs rémunérés s'en étant absentés. Si, par conséquent, le coefficient de variation de la proportion est inacceptable (rendant la proportion non diffusable), il en sera de même du coefficient de variation de l'estimation quantitative correspondante (rendant cette estimation quantitative non diffusable).

Des coefficients de variation de telles estimations peuvent être calculés, au besoin, pour une estimation précise à l'aide d'une technique appelée pseudo-répétition, ce qui veut dire diviser les enregistrements inclus dans les fichiers de microdonnées en sous-groupes (ou répétitions) et déterminer la variation à l'intérieur de l'estimation de répétition en répétition. Les utilisateurs qui désirent calculer des coefficients de variation pour des estimations quantitatives peuvent communiquer avec Statistique Canada afin d'obtenir des conseils sur l'allocation d'enregistrements à des répétitions appropriées et sur les formules à employer à l'intérieur de ces calculs.

10.5 Tables de coefficients de variation

Consulter le fichier ESTR201012_CVTabsF.pdf pour les tables de coefficients de variation.

10.6 Estimation de variance pour les fichiers maître

Pour juger de la qualité d'une estimation et en calculer le coefficient de variation (CV), on doit d'abord établir l'écart-type. On a aussi besoin de l'écart-type d'estimations pour les intervalles de confiance. Dans l'Enquête sur le service téléphonique résidentiel (ESTR), on utilise des plans d'échantillonnage et d'étalonnage à plusieurs degrés et, par conséquent, il n'y a pas de formule simple d'estimation de variance qu'on puisse appliquer. Il faut donc procéder par approximation, la méthode bootstrap. Par l'utilisation des poids bootstrap et du programme BOOTVAR, dont il est question à la prochaine section, on peut calculer avec exactitude les coefficients de variation et d'autres estimations de la variance.

Méthode bootstrap pour estimer la variance
On prélève indépendamment sur chaque strate un échantillon aléatoire simple de $(n-1)$ des unités n de l'échantillon. À noter que, comme la sélection s'opère avec remise, une unité peut être choisie plusieurs fois. On reprend l'opération R fois pour ainsi obtenir R échantillons bootstrap. On calcule une valeur moyenne initiale de pondération bootstrap à partir des R échantillons pour chaque unité échantillonnée de la strate. Toute l'opération (prélèvement d'échantillons aléatoires simples et repondération de chaque strate) a lieu à B reprises – B étant une valeur élevée –, ce qui donne B valeurs initiales de pondération bootstrap.

On repondère alors comme dans la pondération courante par correction de non-réponse, étalonnage, etc. On obtient B valeurs de pondération bootstrap finale pour chaque unité de l'échantillon. On rapporte la variation des B estimations possibles correspondant aux B valeurs bootstrap à la variance de l'estimateur par pondération courante; la variance peut être estimée par ce moyen.

Progiciels statistiques pour estimer la variance

Bootvar

On signale aux utilisateurs que des poids bootstrap sont fournis et devraient être utilisés pour l'estimation de la variance. BOOTVAR est un macro-programme qui peut être utilisé pour calculer la variance à l'aide des poids bootstrap. Le programme Bootvar est disponible en format SAS ou SPSS. Il est composé de macros qui calculent la variance dans le cas de totaux, de ratios, de différences entre ratios et de régressions linéaires et logistiques.

Le Bootvar peut être téléchargé à partir du site Internet des Centres de données de recherche (CDR) de Statistique Canada. Il faut accepter la licence d'adhésion automatique Bootvar avant de pouvoir lire les fichiers.

SAS: http://www.statcan.gc.ca/rdc-cdr/bootvar_sas-fra.htm
SPSS: http://www.statcan.gc.ca/rdc-cdr/bootvar_spss-fra.htm

Autres progiciels

Autre que Bootvar, il y a des progiciels commerciaux différents qui peuvent effectuer une certaine analyse fondée sur le plan d'enquête pour estimer la variance ; Stata 9 or 10, SUDAAN et WesVar.

Ces méthodes peuvent être adaptées pour l'ESTR d'après un article d'Owen Phillips intitulé *Comment utiliser les poids bootstrap avec Wes Var et SUDAAN* (n° 12-002-X20040027032 au catalogue) paru dans « *Le Bulletin technique et d'information des Centres de données de recherche* (*Index chronologique*) », automne 2004, vol.1 no 2, Statistique Canada, n° 12-002-XIF au catalogue.

11.0 Pondération

Puisque l'Enquête sur le service téléphonique résidentiel (ESTR) faisait appel à un sous-échantillon de l'échantillon de l'Enquête sur la population active (EPA), le calcul des poids pour les enregistrements des données de l'enquête est clairement lié à la procédure de pondération utilisée aux fins de l'EPA. La procédure de pondération employée pour l'EPA est décrite brièvement ci-dessous.

11.1 Procédures de pondération pour l'Enquête sur la population active

Dans le cas de l'EPA, le poids final rattaché à chaque enregistrement est le produit des facteurs suivants : le poids de base, le sous-poids d'une grappe, le poids de stabilisation, le facteur compensatoire pour les non-réponses et le facteur d'ajustement du rapport province-âge-sexe et du niveau infraprovincial. Chacun de ces facteurs est décrit ci-dessous.

Poids de base
Dans le cas d'un échantillon probabiliste, le plan d'échantillonnage lui-même détermine les poids qui doivent être utilisés pour produire des estimations non biaisées de la population. Il faut pondérer chaque enregistrement à l'aide de l'inverse de la probabilité de sélectionner la personne à qui l'enregistrement renvoie. Dans le cas d'un échantillon aléatoire simple de 2 %, comme cette probabilité serait de 0,02 pour chaque personne, il faut pondérer les enregistrements à l'aide de 1 / 0,02 = 50. Étant donné la complexité du plan de l'EPA, les logements situés dans différentes régions auront des poids de base différents. Parce que toutes les personnes admissibles habitant un logement sont interviewées (directement ou par personne interposée), cette probabilité est essentiellement la même que la probabilité de sélection du logement.

Sous-poids d'une grappe
La délimitation des grappes est telle que le nombre de logements inclus dans l'échantillon augmente très légèrement en cas de croissance modérée du parc immobilier. On peut tolérer une croissance importante à l'intérieur d'une grappe isolée avant que l'échantillon additionnel ne pose un problème sur le plan de la collecte des données sur le terrain. S'il y a cependant croissance dans plus d'une grappe incluse à l'intérieur de la tâche d'un intervieweur, l'effet cumulatif de toutes les augmentations peut créer un problème sur le plan de la charge de travail. Dans le cas des grappes dont la croissance est importante, on utilise un sous-échantillon pour que les tâches des intervieweurs demeurent gérables. Le sous-poids d'une grappe représente l'inverse de ce rapport de sous-échantillonnage dans le cas des grappes pour lesquelles il y a eu sous-échantillonnage.

Poids de stabilisation
On a aussi recours à la stabilisation d'un échantillon pour s'attaquer aux problèmes soulevés par la croissance de la taille de ce dernier. Le sous-échantillonnage d'une grappe s'attaquait à une croissance isolée dans des régions relativement petites, tandis que la stabilisation d'un échantillon s'attaque à la croissance lente d'un échantillon au fil du temps qui est le résultat d'un taux fixe d'échantillonnage parallèlement à une augmentation générale de la taille, ou de l'effectif, de la population. La stabilisation d'un échantillon est la suppression aléatoire de logements de l'échantillon, ce qui vise à maintenir la taille de ce dernier à son niveau désiré. On ajuste le poids de base à l'aide du rapport de la taille de l'échantillon, qui repose sur le taux fixe d'échantillonnage, à la taille de l'échantillon désirée. On appelle ce facteur d'ajustement le poids de stabilisation. L'ajustement se fait à l'intérieur de secteurs de stabilisation définis comme étant des logements appartenant à la même région économique de l'assurance-emploi et au même groupe de renouvellement.

Non-réponse
Dans le cas de certains types de non-réponses (comme les ménages temporairement absents ou les refus), les données de l'interview menée le mois précédent auprès d'un ménage, le cas échéant, sont réutilisées pour ce ménage comme données du mois courant.

Dans d'autres cas, on contrebalance les poids des non-réponses en accroissant proportionnellement ceux des ménages ayant répondu au questionnaire. On augmente le poids de chaque enregistrement d'un ménage ayant répondu au questionnaire à l'aide du rapport du nombre de ménages qui auraient dû être interviewés divisé par le nombre de ceux qui l'ont réellement été. On effectue séparément cet ajustement pour les secteurs de non-réponses, qui sont définis par la région économique de l'assurance-emploi, le type de secteur et le groupe de renouvellement. L'ajustement repose sur l'hypothèse voulant que les ménages qui ont été interviewés représentent les caractéristiques de ceux qui auraient dû être interviewés se trouvant à l'intérieur d'un secteur de non-réponse.

Sous-poids de l'Enquête sur la population active
On appelle le sous-poids de l'EPA le produit des facteurs de pondération décrits précédemment. Tous les membres du même ménage échantillonné ont le même sous-poids.

Ajustements au niveau infraprovincial et au rapport province-âge-sexe
On peut utiliser le sous-poids afin de calculer une estimation valable de toute caractéristique pour laquelle on collecte des données au moyen de l'EPA. Cependant, ces estimations seront fondées sur une base contenant des renseignements qui peuvent être périmés depuis plusieurs années et qui ne sont donc pas représentatifs de la population actuelle. Au moyen de renseignements complémentaires plus récents sur la population cible, les poids d'échantillonnage sont ajustés en vue d'améliorer la précision des estimations et la représentativité de l'échantillon de la population actuelle.

On dispose mensuellement d'estimations indépendantes pour divers groupes d'âge-sexe selon la province. Ce sont des projections démographiques fondées sur les données du recensement, les enregistrements des naissances et des décès et les estimations de la migration, les plus récents. À la dernière étape, on utilise ces renseignements auxiliaires pour convertir le sous-poids en poids final, ce qui se fait à l'aide d'une méthode de calibration. Cette méthode assure que les poids finals qu'elle produit équivalent aux projections du recensement pour les variables auxiliaires, c'est-à-dire des totaux pour divers groupes d'âge-sexe, de régions économiques, de régions métropolitaines de recensement, de groupes de renouvellement, de ménages et de la taille de la famille économique. On corrige également les poids de manière à ce que la somme des estimations de la branche d'activité et de la main-d'œuvre du mois précédent, dérivées de l'échantillon du mois en cours, correspondent aux estimations correspondantes de l'échantillon du mois précédent. Il s'agit de l'estimation composite. On applique la méthode de la régression généralisée à l'ensemble de la correction.

Habituellement, on ne tient pas compte de ce poids définitif dans la détermination du poids d'une enquête supplémentaire de l'EPA. On a plutôt recours à la sous-pondération comme l'expliquent les paragraphes qui suivent.

11.2 Procédures de pondération pour l'Enquête sur le service téléphonique résidentiel

Les principes qui sous-tendent le calcul des poids pour l'Enquête sur le service téléphonique résidentiel sont presque identiques à ceux établis aux fins de l'EPA. La pondération pour l'ESTR est par contre faite au niveau du ménage et non au niveau de la personne. Aussi, d'autres ajustements sont apportés aux sous-poids de l'EPA afin de calculer un poids final pour les différents enregistrements inclus dans le fichier de microdonnées de l'ESTR.

1) Un ajustement qui tient compte de l'utilisation d'un sous-échantillon représentant les deux sixièmes pour l'Ontario et le Québec et les trois sixièmes pour tous les autres provinces,

plutôt que de l'échantillon complet de l'EPA.

2) Un ajustement qui tient compte de la non-réponse additionnelle au questionnaire de l'enquête supplémentaire, c'est-à-dire de la non-réponse à l' ESTR provenant des gens qui ont répondu au questionnaire de l'EPA ou pour qui des données de l'EPA du mois précédent ont été reportées. Une analyse a été menée dans le but de déterminer des groupes homogènes de non-réponse. Les groupes ont été définis en tenant compte de diverses caractéristiques telles la taille du ménage, le nombre de tentatives pour contacter le ménage, le type d'interview (procuration ou non) ainsi que d'autres variables démographiques et géographiques.

3) L'ajustement final a fait en sorte que les estimations produites pour un groupe province-âge-sexe coïncident avec les chiffres de population connus pour ce même groupe.

   On tient compte d'ajustement 1) en multipliant le sous-poids (Ws) de l'EPA pour chacun des enregistrements ayant répondu à l'ESTR par :

$$\frac{6}{2} \times \frac{1}{Ws} \text{ ou } \frac{6}{3} \times \frac{1}{Ws}$$

   pour obtenir WEIGHT1.

   On tient compte d'ajustement 2) en tenant compte du facteur d'ajustement de la non-réponse pour chaque groupe de non-réponse. On multiplie le poids WEIGHT1 par le facteur d'ajustement de la non-réponse afin d'obtenir le poids WEIGHT2.

   L'ajustement 3) correspond à un facteur de calage propre à chaque répondant de l'ESTR. Le poids WEIGHT2 est multiplié par ce facteur de calage afin d'obtenir le poids final de l'ESTR. Ce poids, appelé WTHM, est le poids final qui figure dans le fichier maître de l'Enquête sur le service téléphonique résidentiel.