



Guide de l'utilisateur des microdonnées

Enquête canadienne sur l'utilisation d'Internet 2005



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
4.1	Concepts et définitions de l'Enquête sur la population active	11
4.2	Concepts et définitions de l'Enquête canadienne sur l'utilisation d'Internet	12
5.0	Méthodologie de l'enquête	13
5.1	Population visée	13
5.2	Plan de sondage	13
5.2.1	Stratification primaire	13
5.2.2	Types de régions	13
5.2.3	Stratification secondaire	14
5.2.4	Délimitation et sélection des grappes	14
5.2.5	Sélection des logements	15
5.2.6	Sélection des personnes	15
5.3	Taille de l'échantillon	15
5.4	Renouvellement de l'échantillon	16
5.5	Modifications apportées au plan de l'Enquête sur la population active pour l'Enquête canadienne sur l'utilisation d'Internet	16
5.6	Taille de l'échantillon selon la province pour l'Enquête canadienne sur l'utilisation d'Internet	17
6.0	Collecte des données	19
6.1	Réalisation des interviews dans le cadre de l'Enquête sur la population active	19
6.2	Supervision et contrôle de qualité	19
6.3	Non-réponse à l'Enquête sur la population active	20
6.4	Modifications apportées à la collecte des données de l'Enquête canadienne sur l'utilisation d'Internet	20
6.5	Non-réponse à l'Enquête canadienne sur l'utilisation d'Internet	20
7.0	Traitement des données	21
7.1	Saisie des données	21
7.2	Vérification	21
7.3	Codage des questions ouvertes	21
7.4	Imputation	22
7.5	Création de variables dérivées	22
7.6	Pondération	22
7.7	Suppression de renseignements confidentiels	23
8.0	Qualité des données	25
8.1	Taux de réponse	25
8.2	Erreurs relatives à l'enquête	25
8.2.1	Base de sondage	26
8.2.2	Collecte des données	26
8.2.3	Traitement des données	27
8.2.4	Non-réponse	27
8.2.5	Mesure de l'erreur d'échantillonnage	28

9.0	Lignes directrices pour la totalisation, l'analyse et la diffusion de données	29
9.1	Lignes directrices pour l'arrondissement d'estimations	29
9.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation	30
9.3	Définitions de types d'estimations : catégoriques et quantitatives	30
9.3.1	Estimations catégoriques	30
9.3.2	Estimations quantitatives	30
9.3.3	Totalisation d'estimations catégoriques	31
9.3.4	Totalisation d'estimations quantitatives	31
9.4	Lignes directrices pour l'analyse statistique	32
9.5	Lignes directrices pour la diffusion de coefficients de variation	33
9.6	Seuils pour la diffusion des estimations pour l'Enquête canadienne sur l'utilisation d'Internet	35
10.0	Tables de variabilité d'échantillonnage approximative	37
10.1	Comment utiliser les tables de coefficients de variation pour des estimations catégoriques	38
10.1.1	Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques	39
10.2	Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance	45
10.2.1	Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance	46
10.3	Comment utiliser les tables de coefficients de variation pour effectuer un test t	46
10.3.1	Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t	47
10.4	Coefficients de variation pour des estimations quantitatives	47
10.5	Tables des coefficients de variation	47
11.0	Pondération	49
11.1	Procédures de pondération pour l'Enquête sur la population active	49
11.2	Procédures de pondération pour l'Enquête canadienne sur l'utilisation d'Internet	50
12.0	Questionnaires	53
12.1	Le questionnaire de l'Enquête sur la population active	53
12.2	Le questionnaire de l'Enquête canadienne sur l'utilisation d'Internet	53
13.0	Cliché d'enregistrement à valeurs univariées	55

1.0 Introduction

L'Enquête canadienne sur l'utilisation d'Internet (ECUI) a été menée en 2005 par Statistique Canada pour la Division de la science, de l'innovation et de l'information électronique à Statistique Canada. 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

Larry McKeown

Division de la science, de l'innovation et de l'information électronique

150, promenade du Pré Tunney, Ottawa (Ontario) K1A 0T6

Téléphone : (613) 951-2582

Télécopieur : (613) 951-0009

Courriel : Larry.McKeown@statcan.ca

2.0 Contexte

L'Enquête canadienne sur l'utilisation d'Internet (ECUI), qui remplace l'Enquête sur l'utilisation d'Internet par les ménages (EUIM), a été menée pour la première fois en 2005. L'EUIM avait été réalisée annuellement de 1997 à 2003 afin de mesurer l'utilisation d'Internet par les ménages. Comme la croissance du nombre de ménages qui utilisent Internet se stabilise, on a remanié l'enquête afin de se concentrer sur la façon dont les personnes, plutôt que les ménages, utilisent Internet.

Le fait que l'ECUI est maintenant une enquête auprès des personnes au lieu d'une enquête-ménage permet de mieux l'aligner sur les normes internationales relatives aux statistiques concernant Internet et d'étudier de façon plus poussée l'utilisation d'Internet à large bande et les conséquences éventuelles de l'utilisation d'Internet sur les personnes et la société tant sur le plan social que sur le plan économique.

3.0 Objectifs

Cette enquête a pour objectif de mesurer la demande des Canadiens pour Internet et l'utilisation qu'ils en font. Pour ce faire, on tient compte de l'intensité et de la fréquence d'utilisation, ainsi que de la vitesse de connexion Internet.

D'autres renseignements sont recueillis sur :

- l'accessibilité et l'utilisation d'Internet des Canadiens à la maison, dans les lieux de travail et dans d'autres endroits (par exemple, dans les bibliothèques, les écoles et les cafés);
- le genre de services et d'informations auxquels les gens ont accès à partir de la maison (par exemple, le courrier électronique, les transactions bancaires électroniques, les services éducatifs, les renseignements médicaux et sur la santé);
- la commande et l'achat de biens et services par l'entremise d'Internet pour consommation personnelle ou pour le ménage, de n'importe quel endroit ;
- comment le *lèche-vitrine* sur Internet a exercé une influence sur nos vies; et
- les personnes qui n'utilisent pas Internet et les raisons pour lesquelles elles ne l'utilisent pas. Il est important de comprendre les particularités de ces personnes, car elles ont un impact sur les services gouvernementaux qui seront offerts en ligne et sur le moment où ils le seront.

L'information recueillie servira à mettre à jour et à enrichir les études déjà réalisées par Statistique Canada sur l'utilisation d'Internet.

4.0 Concepts et définitions

Ce chapitre donne un aperçu des concepts et des définitions d'intérêt pour les utilisateurs. Les concepts et les définitions utilisés dans le cadre de l'Enquête sur la population active (EPA) sont décrits à la section 4.1, tandis que ceux propres à l'Enquête canadienne sur l'utilisation d'Internet (ECUI) sont fournis dans la section 4.2. Les utilisateurs sont priés de se reporter au chapitre 12.0 de ce document où figure une copie du (des) questionnaire(s) d'enquête réellement employé(s).

4.1 Concepts et définitions de l'Enquête sur la population active

Situation vis-à-vis de l'activité

Désigne la situation de l'enquêté sur le marché du travail : un membre de la population active âgé de 15 ans et plus (à l'exclusion des pensionnaires d'établissements) est soit occupé, soit en chômage, soit inactif.

Emploi

Les personnes occupées sont celles qui, au cours de la semaine de référence :

- a) ont fait un travail¹ quelconque dans le cadre d'un emploi ou dans une entreprise; ou
- b) avaient un emploi, mais n'étaient pas au travail à cause d'une maladie ou d'une incapacité, pour obligations personnelles ou familiales, pour des vacances, par suite d'un conflit de travail ou du fait de tout autre facteur (cela n'inclut pas les personnes mises à pied, celles qui étaient inactives entre deux emplois occasionnels et celles qui avaient un emploi devant commencer à une date ultérieure).

Chômage

Les chômeurs sont les personnes qui, au cours de la semaine de référence :

- a) avaient été mises à pied temporairement, mais s'attendaient à être rappelées au travail et étaient disponibles pour travailler; ou
- b) étaient sans emploi, avaient activement cherché un emploi au cours des quatre dernières semaines et étaient disponibles pour travailler²; ou
- c) devaient commencer un nouvel emploi dans quatre semaines ou moins à compter de la semaine de référence et étaient disponibles pour travailler.

Inactifs

Les personnes inactives sont celles qui, considérant les conditions qui existaient sur le marché du travail dans leur région, ne désiraient ou ne pouvaient offrir ou fournir leurs services au cours de la semaine de référence. Autrement dit, elles n'étaient ni occupées ni en chômage.

¹ Travail inclut un travail quelconque contre rémunération ou en vue d'un bénéfice, c'est-à-dire tout travail rémunéré accompli pour un employeur ou à son propre compte, cela comprend aussi le travail familial non rémunéré, c'est-à-dire un travail (non rémunéré) qui contribue directement à l'exploitation d'une ferme, d'une entreprise ou d'un cabinet de professionnel appartenant à un membre apparenté du même ménage et exploité par lui. De telles activités peuvent inclure la tenue de livres, la vente de produits, le service aux tables, etc. Des tâches comme le ménage ou l'entretien ménager ne sont pas considérées comme un travail familial non rémunéré.

² Les personnes sont considérées comme étant disponibles pour travailler si elles :

- i) ont déclaré qu'elles auraient pu travailler durant la semaine de référence si on leur avait offert un emploi convenable, ou encore si la raison pour laquelle elles ne pouvaient occuper un emploi était de nature temporaire, notamment du fait d'une maladie ou d'une incapacité, d'obligations personnelles ou familiales, parce qu'elles avaient déjà un emploi devant bientôt débiter, ou qu'elles étaient en vacances (avant 1997, les personnes en vacances n'étaient pas considérées comme étant disponibles).
- ii) étaient des étudiants à temps plein à la recherche d'un travail à temps partiel qui respectaient également la condition i) ci-dessus. Les étudiants à temps plein qui sont actuellement aux études et qui cherchent un travail à temps plein ne sont pas considérés comme étant disponibles au cours de la semaine de référence.

Branche d'activité et de la profession

L'Enquête sur la population active fournit de l'information sur la branche d'activité et la profession des personnes occupées et en chômage et de celles inactives qui ont détenu un emploi au cours des 12 derniers mois. Depuis 1997, ces statistiques reposent sur le Système de classification des industries de l'Amérique du Nord (SCIAN) et la Classification type des professions (CTP de 1991). Avant 1997, la Classification type des industries de 1980 ainsi que la Classification type des professions de 1980 étaient utilisées.

Semaine de référence

Semaine civile complète (du dimanche au samedi) sur laquelle porte, chaque mois, l'Enquête sur la population active. C'est habituellement la semaine où tombe le 15^e jour du mois. Les interviews sont réalisées durant la semaine suivante, appelée la semaine d'enquête, et la situation vis-à-vis de l'activité déterminée est celle de la semaine de référence.

Emploi à temps plein

L'emploi à temps plein tient compte des personnes qui travaillent habituellement 30 heures ou plus par semaine à leur emploi principal ou à leur unique emploi.

Emploi à temps partiel

L'emploi à temps partiel tient compte des personnes qui travaillent habituellement moins de 30 heures par semaine à leur emploi principal ou à leur unique emploi.

4.2 Concepts et définitions de l'Enquête canadienne sur l'utilisation d'Internet

Utilisateurs

Les utilisateurs sont des personnes qui se branchent à Internet au moins une fois au cours des 12 derniers mois pour des motifs personnels (utilisations non commerciales).

Anciens utilisateurs

Les anciens utilisateurs sont des personnes qui ont utilisé Internet par le passé, mais qui ne l'utilisent pas à l'heure actuelle. Un « ancien utilisateur » n'a pas utilisé Internet au cours des 12 derniers mois, mais l'a utilisé à un moment donné dans le passé.

Non-utilisateurs

Les non-utilisateurs sont des personnes qui n'ont jamais utilisé Internet. Pour certaines sections, les non-utilisateurs sont des gens qui n'utilisent pas Internet à partir de la maison.

Utilisations

Les utilisations d'Internet comptent la communication avec le gouvernement, la recherche de renseignements médicaux et sur la santé, l'utilisation éducative, les utilisations précises en matière de communication (réseaux sociaux) et le commerce électronique.

Lèche-vitrine

Les personnes font du lèche-vitrine quand ils comparent les caractéristiques et les prix des produits et/ou services sans nécessairement les acheter.

5.0 Méthodologie de l'enquête

Comme l'Enquête canadienne sur l'utilisation d'Internet (ECUI) a été menée du 13 novembre au 6 décembre 2005 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. Le plan de l'EPA est décrit brièvement à l'intérieur des sections 5.1 à 5.4³. Les sections 5.5 et 5.6 décrivent comment l'ECUI s'est écartée du plan de base de l'EPA de 2005.

5.1 Population visée

L'EPA est une enquête mensuelle réalisée auprès des ménages. Son échantillon est représentatif de la population canadienne civile non institutionnalisée de 15 ans et plus des 10 provinces du Canada. Sont spécifiquement exclus du champ de l'enquête les résidents du Yukon, des Territoires du Nord-Ouest et du Nunavut, les personnes qui vivent dans les réserves indiennes, les membres à temps plein des Forces armées canadiennes et les pensionnaires d'établissements. Réunies, ces personnes exclues de l'enquête représentent environ 2 % de la population de 15 ans et plus.

5.2 Plan de sondage

L'EPA a fait l'objet d'un remaniement poussé, dont le point culminant a été l'introduction du nouveau plan à la fin de 1994. L'échantillon de l'EPA repose sur un échantillonnage probabiliste stratifié faisant appel à un plan à plusieurs degrés à tous les stades de ce dernier. Les principes du plan sont les mêmes pour chaque province. Un diagramme résumant les stades du plan figure dans le document EPA_AnnexeA.pdf.

5.2.1 Stratification primaire

Les provinces sont divisées en régions économiques (RÉ) et en régions économiques d'assurance-emploi (RÉAE). Les RÉ sont des régions géographiques d'une structure économique plus ou moins homogène formées à la suite d'ententes fédérales-provinciales. Elles sont relativement stables au fil du temps. Les RÉAE sont également des régions géographiques et en gros ont la même taille et sont aussi nombreuses que les RÉ, mais ne sont pas définies de la même façon. On produit des estimations de la population active pour les RÉAE aux fins de Service Canada.

Les intersections des régions des deux types susmentionnés forment le premier niveau de stratification pour l'EPA. On traite ces intersections de RÉ et de RÉAE comme des strates primaires et on effectue une stratification plus poussée à l'intérieur d'elles (voir la section 5.2.3). À noter que la stratification incluse dans le plan actuel de l'EPA respecte aussi un troisième ensemble de régions, les régions métropolitaines de recensement (RMR), puisque chaque RMR est également une RÉAE.

5.2.2 Types de régions

Les strates primaires (les intersections de RÉ et de RÉAE) sont en outre désagrégées en trois types de régions : rurales, urbaines et éloignées. Les régions rurales et urbaines sont généralement fondées sur les définitions du recensement de régions urbaines et de régions rurales, à quelques exceptions près destinées à permettre la formation de strates dans certaines régions. Les régions urbaines vont des RMR les plus grandes jusqu'aux

³ 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.

villages les plus petits classés suivant le Recensement de 1991 dans la catégorie des régions urbaines (de 1 000 habitants ou plus), tandis que les régions rurales se composent des régions non désignées régions urbaines ni régions éloignées.

Toutes les régions urbaines sont, en outre, subdivisées en deux types : celles pour lesquelles on utilise une liste d'appartements et une base aréolaire et celles pour lesquelles on emploie uniquement une telle base.

Environ 1 % de la population visée par l'EPA se trouve dans des régions éloignées des provinces qui sont moins accessibles que d'autres régions pour les intervieweurs affectés à l'EPA. À des fins administratives, cette portion de la population est échantillonnée séparément à l'aide de la liste des régions éloignées. Certaines populations, non rassemblées dans des endroits comptant 25 habitants ou plus, sont exclues de la base de sondage.

5.2.3 Stratification secondaire

Dans les régions urbaines où le nombre d'immeubles d'appartements est suffisamment grand, les strates sont subdivisées en listes d'appartements et en bases aréolaires. Une liste d'appartements est un registre tenu à jour dans les 18 plus grands centres du Canada. Cela vise à assurer une meilleure représentation des résidents d'appartements à l'intérieur de l'échantillon et à réduire l'effet de croissance dans les grappes, attribuable à la construction de nouveaux immeubles d'appartements. Dans les principaux centres, les strates d'appartements sont, en outre, subdivisées en strates de faibles revenus et en strates ordinaires.

Lorsque cela est possible et/ou nécessaire, la liste des régions urbaines est, en plus, subdivisée en strates ordinaires, en strates de revenus élevés et en strates de faible densité de population. La plupart des régions urbaines font partie des strates urbaines ordinaires, qui, en fait, englobent la majorité de la population canadienne. Les strates de revenus élevés se trouvent dans les principales régions urbaines, tandis que les strates urbaines de faible densité se composent de petites villes géographiquement dispersées.

Dans les régions rurales, la densité de population peut varier grandement, c'est-à-dire aller d'une densité de population relativement élevée à une faible densité de population, ce qui entraîne la formation de strates reflétant ces variations. Les différentes stratégies de stratification pour les régions rurales ont été fondées non seulement sur la concentration de la population, mais également sur le coût-efficacité et les contraintes auxquelles les intervieweurs sont confrontés.

Dans chaque province, le nombre de peuplements éloignés échantillonnés est proportionnel au nombre de logements, sans autre stratification. On sélectionne les logements à l'aide d'une méthode d'échantillonnage systématique dans chacun des endroits échantillonnés.

5.2.4 Délimitation et sélection des grappes

On ne sélectionne pas directement les ménages à l'intérieur des strates finales. On divise plutôt chaque strate en grappes, puis on sélectionne un échantillon de grappes à l'intérieur de la strate. On échantillonne ensuite les logements à partir des grappes sélectionnées. On utilise différentes méthodes pour définir les grappes, suivant le type de strate.

À l'intérieur de chaque strate urbaine incluse sur la liste des régions urbaines, on forme un certain nombre de groupes géographiquement continus de logements, ou grappes, à

partir des chiffres du Recensement de 1991. Ces grappes sont généralement un ensemble d'un ou de plusieurs îlots, ou côtés d'îlot. La sélection d'un échantillon de grappes (toujours six ou un multiple de six grappes) à partir de chacune de ces strates secondaires représente le premier degré d'échantillonnage dans la plupart des régions urbaines. Dans certaines autres régions urbaines, on utilise comme grappes des secteurs de dénombrement (SD) du recensement. Dans les strates urbaines de faible densité de population, on suit un plan à trois degrés ou stades. Aux termes de ce plan, on échantillonne deux villes à l'intérieur d'une strate, puis 6 ou 24 grappes à l'intérieur de chaque ville.

Pour les strates d'appartements des régions urbaines, plutôt que de définir des grappes, on utilise l'immeuble d'appartements comme unité primaire d'échantillonnage. On échantillonne les immeubles d'appartements à partir de la liste, la probabilité d'échantillonnage étant proportionnelle au nombre d'unités que renferme chaque immeuble.

À l'intérieur de chacune des strates secondaires des régions rurales, on effectue, si nécessaire, une autre stratification afin de refléter les différences entre un certain nombre de caractéristiques socio-économiques de chaque strate. À l'intérieur de chaque strate d'une région rurale, on échantillonne comme grappes six SD ou deux ou trois groupes de SD.

5.2.5 Sélection des logements

Dans les trois types de régions que sont les régions urbaines, rurales et éloignées, des enquêteurs itinérants visitent premièrement les grappes sélectionnées, puis on dresse une liste de tous les logements privés faisant partie des grappes. On sélectionne ensuite à partir de cette liste un échantillon de logements. Le rendement de l'échantillon dépend du type de strate. Sur la liste de régions urbaines, par exemple, les rendements de l'échantillon sont de six ou de huit logements, selon la taille du centre. Le rendement de chaque grappe inscrite sur la liste d'appartements d'une région urbaine, est de cinq logements, tandis que dans les régions rurales et dans les parties de centres formées par des SD le rendement de chaque grappe est de dix logements. Dans toutes les grappes, on échantillonne systématiquement les logements, ce qui représente le dernier degré d'échantillonnage.

5.2.6 Sélection des personnes

On recueille des renseignements démographiques sur tous les membres du ménage pour lesquels le logement sélectionné constitue le lieu de résidence habituel. Les renseignements obtenus dans le cadre de l'EPA concernent tous les membres civils du ménage âgés de 15 ans ou plus. Pour les personnes âgées, c'est-à-dire celles de 70 ans et plus, on réduit le fardeau de la personne en réutilisant leurs réponses à l'interview initiale au cours des cinq mois suivants l'enquête.

5.3 Taille de l'échantillon

On détermine la taille de l'échantillon des personnes admissibles dans le cadre de l'EPA de façon à respecter les exigences en matière de précision statistique établies pour diverses caractéristiques de la population active aux niveaux provincial et infraprovincial et à répondre aux besoins des administrations fédérale, provinciales et municipales et à ceux d'une foule d'autres utilisateurs de données.

L'échantillon mensuel de l'EPA se compose d'environ 60 000 logements. Après en avoir exclu

les logements trouvés vacants, les logements démolis ou ceux convertis à des fins non résidentielles, ceux n'abritant que des personnes inadmissibles, les logements en construction et les logements saisonniers, il reste à peu près 54 000 logements occupés par une ou par plusieurs personnes admissibles. On recueille de l'information aux fins de l'EPA à partir de ces logements sur environ 102 000 civils âgés de 15 ans et plus.

5.4 Renouvellement de l'échantillon

L'EPA utilise un plan de sondage avec renouvellement de panel, suivant lequel les ménages sélectionnés restent dans l'échantillon pendant six mois consécutifs. L'échantillon complet est formé de six sous-échantillons ou panels représentatifs. Chaque mois, on remplace le panel qui fait partie de l'échantillon depuis six mois. Les ménages qui cessent de faire partie de l'échantillon sont remplacés par d'autres ménages du même secteur ou d'un secteur comparable. Il en résulte un chevauchement des cinq sixièmes de l'échantillon d'un mois à l'autre, ce qui procure un plan de sondage efficace pour estimer les variations d'un mois à l'autre. La rotation après six mois évite d'imposer un fardeau trop lourd aux enquêtés des ménages sélectionnés aux fins de l'enquête.

Grâce à la caractéristique que constituent les groupes de renouvellement, on peut mener facilement des enquêtes supplémentaires à l'aide du plan de l'EPA, mais en utilisant un échantillon moins complet.

5.5 Modifications apportées au plan de l'Enquête sur la population active pour l'Enquête canadienne sur l'utilisation d'Internet

On a utilisé pour l'ECUI cinq des six groupes de renouvellement inclus dans l'échantillon de l'EPA de novembre 2005. Pour les besoins de l'ECUI, on a modifié la couverture de l'EPA afin d'y inclure tous les membres d'un ménage âgés de 18 ans et plus. Contrairement à l'EPA, pour laquelle des données sont recueillies pour tous les membres admissibles d'un ménage, l'ECUI n'a donné lieu à la cueillette d'information qu'auprès d'un seul membre sélectionné au hasard d'un ménage. Les réponses par personne interposée n'étaient pas non plus permises pour l'ECUI.

5.6 Taille de l'échantillon selon la province pour l'Enquête canadienne sur l'utilisation d'Internet

Le tableau qui suit montre le nombre de membres (âgés de 18 ans et plus) des ménages inclus dans les groupes de renouvellement échantillonnés de l'EPA qui étaient admissibles pour le supplément de l'ECUI. Ce tableau comprend les membres des ménages qui étaient non-répondants à l'EPA.

Provinces	Taille de l'échantillon
Terre-Neuve-et-Labrador	1 537
Île-du-Prince-Édouard	1 127
Nouvelle-Écosse	2 282
Nouveau-Brunswick	2 236
Québec	7 996
Ontario	12 545
Manitoba	2 926
Saskatchewan	3 092
Alberta	4 257
Colombie-Britannique	5 082
Canada	43 080

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 personne interposée », 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 du 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 canadienne sur l'utilisation d'Internet

Une personne par ménage sélectionnée au hasard répond au questionnaire de l'Enquête canadienne sur l'utilisation d'Internet (ECUI).. La sélection au hasard a été effectuée au moment de l'interview.

À la fin de l'interview menée pour l'Enquête sur la population active, l'intervieweur a demandé à parler à la personne sélectionnée pour l'ECUI. Si la personne sélectionnée n'était pas disponible, l'intervieweur fixait un moment opportun pour rappeler. Comme les « réponses par personne interposée » n'étaient pas permises, on a prolongé de plus de deux semaines la période de collecte des données afin de laisser aux intervieweurs le temps de communiquer avec les gens présélectionnés.

6.5 Non-réponse à l'Enquête canadienne sur l'utilisation d'Internet

Pour les ménages ayant répondu au questionnaire de l'EPA, l'étape suivante de la collecte des données consistait à poser les questions de l'ECUI. Au total, 43 080 personnes étaient admissibles pour l'enquête supplémentaire; le nombre final d'interview aux fins de l'ECUI s'est élevé à 30 466 de ces personnes, ce qui a donné un taux de réponse de 70,7 %. 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 canadienne sur l'utilisation d'Internet (ECUI) 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.

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 l'enquêté 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 parfois à des questions subséquentes.

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 à ces questions.

7.3 Codage des questions ouvertes

Quelques données élémentaires ont été consignées sur le questionnaire par les intervieweurs sous forme de questions ouvertes. On a inclus dans l'enquête au total 18 questions partiellement ou complètement ouvertes. C'étaient des questions portant sur l'utilisation antérieure, le lieu d'utilisation, l'utilisation individuelle, le gouvernement en direct, l'utilisation à des fins éducatives, le commerce électronique, pas d'utilisation au domicile, l'origine ethnique et le revenu.

7.4 Imputation

L'imputation est le processus qui fournit des valeurs valides concernant les variables qui ont été retenues pour être modifiées, soit en raison de renseignements invalides, soit en raison de renseignements manquants. Les nouvelles valeurs sont établies de façon à préserver la structure sous-jacente des données et à garantir que les enregistrements qui en résultent passeront tous les contrôles requis. En d'autres mots, l'objectif n'est pas de reproduire les véritables valeurs des microdonnées mais plutôt d'établir des enregistrements intrinsèquement cohérents qui permettront de produire de bonnes estimations agrégées.

Nous pouvons faire la distinction entre trois types de non-réponse. La non-réponse complète est lorsque le répondant ne fournit pas le nombre minimal de réponses. Ces enregistrements sont supprimés et seront pris en compte lors du processus de pondération (voir le chapitre 11.0). La non-réponse ponctuelle est lorsque le répondant ne fournit pas une réponse à une question mais qu'il passe à la question suivante. Ce type de non-réponse est habituellement traité en utilisant le code « non déclaré » ou en ayant recours à l'imputation. Enfin, la non-réponse partielle est lorsque le répondant fournit le nombre minimal de réponses mais ne termine pas l'interview. Ces enregistrements peuvent être traités comme des cas de non-réponse complète ou ponctuelle.

Dans le cas de l'ECUI, l'imputation par enregistrement donneur a été utilisée pour fournir les éléments d'information manquants aux enregistrements avec non-réponse ponctuelle ou partielle. De plus amples renseignements sur le processus d'imputation sont donnés au chapitre 8.0 (Qualité des données).

7.5 Création de variables dérivées

Un certain nombre de données élémentaires incluses dans le fichier de microdonnées ont été calculées en combinant des réponses sur le questionnaire pour faciliter l'analyse des données. Un exemple est la variable de quintile de revenu du ménage HINCQUIN (G_HQUINT dans le Fichier de microdonnées à grande diffusion), qui est créée à partir d'information sur le revenu recueillie au cours de l'interview. On a eu recours à la technique de l'imputation dans les cas pour lesquels la variable sur le revenu manquait (voir la section 8.2.4 pour obtenir plus de détails sur la méthode utilisée pour faire l'imputation du revenu).

7.6 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.

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 personnes utilisant habituellement Internet à partir de la maison doit être estimé, cette opération s'effectue en sélectionnant les enregistrements se référant aux personnes incluses 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.

7.7 Suppression de renseignements confidentiels

Il convient de souligner que les fichiers de microdonnées « à grande diffusion » (FMGD) peuvent différer des fichiers « maîtres » de l'enquête que conserve Statistique Canada. Ces différences sont habituellement le résultat de mesures prises pour protéger l'anonymat des répondants à une enquête. Les mesures les plus courantes sont la suppression de variables du fichier, le regroupement de valeurs en des catégories plus étendues et le codage de valeurs spécifiques à la catégorie « non déclaré ». 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.

Le fichier maître de données de l'enquête comprend l'âge exact des répondants, tandis que le FMGD ne comprend que des regroupements d'âges.

Lorsque cela est nécessaire, certains des codes de réponse utilisés à l'intérieur des questions ouvertes sont agrégés dans le FMGD. Ils sont regroupés pour convenir aux principaux schémas de codage.

Pour certaines variables sensibles au risque d'identification d'individus, le FMGD peut avoir subi de la suppression locale, c'est-à-dire que certaines valeurs présentes dans le fichier maître peuvent avoir été codées comme « non déclarées » sur le FMGD.

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 canadienne sur l'utilisation d'Internet (ECUI).

Provinces	Ménages sélectionnés de l'EPA	Taux de réponse à l'EPA*	Ménages de l'EPA admissibles à l'ECUI	Personnes répondantes à l'ECUI	Taux de réponse à l'ECUI**	Taux de réponse global à l'ECUI***
		%			%	
Terre-Neuve-et-Labrador	1 592	90,5	1 537	1 021	66,4	64,1
Île-du-Prince-Édouard	1 177	91,9	1 127	805	71,4	68,4
Nouvelle-Écosse	2 391	91,9	2 282	1 518	66,5	63,5
Nouveau-Brunswick	2 321	92,0	2 236	1 517	67,8	65,4
Québec	8 266	91,3	7 996	5 923	74,1	71,7
Ontario	12 968	91,7	12 545	8 653	69,0	66,7
Manitoba	3 020	92,1	2 926	2 094	71,6	69,3
Saskatchewan	3 181	92,6	3 092	2 252	72,8	70,8
Alberta	4 469	90,4	4 257	3 122	73,3	69,9
Colombie-Britannique	5 321	90,8	5 082	3 561	70,1	66,9
Canada	44 706	91,5	43 080	30 466	70,7	68,1

Note : Les chiffres de l'EPA s'appliquent aux ménages alors que les chiffres de l'ECUI s'appliquent aux personnes sélectionnées à l'intérieur des ménages (seulement une personne est sélectionnée par ménage). Le taux de réponse global de l'ECUI est fondé sur les enregistrements des ménages dans tous les sous-échantillons de l'EPA utilisés par l'ECUI. Les ménages de l'EPA admissibles à l'ECUI comprennent les répondants à l'EPA du mois courant ainsi que ceux reportés du mois précédent.

- * Le taux de réponse à l'EPA est le nombre de ménages ayant répondu au questionnaire de l'EPA exprimé sous forme de pourcentage du nombre de ménages sélectionnés pour l'EPA.
- ** Le taux de réponse à l'ECUI est le nombre de personnes ayant répondu au questionnaire de l'ECUI exprimé sous forme de pourcentage du nombre de ménages de l'EPA admissibles à l'ECUI.
- *** Le taux de réponse global à l'ECUI est le nombre de personnes ayant répondu au questionnaire de l'ECUI exprimé sous forme de pourcentage du nombre de ménages ayant été sélectionnés pour l'EPA.

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 enquêtés 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 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'ECUI 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'ECUI. La qualité des variables d'échantillon dans la base était très élevée. L'échantillon de l'ECUI était composé de cinq groupes de renouvellement de l'EPA. Les critères de sélection utilisés pour l'ECUI (tel que le groupe de renouvellement) n'étaient manquants pour aucun des enregistrements de l'EPA.

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'ECUI 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.

Certaines variables incluses dans la base de sondage peuvent jouer un rôle crucial relativement à l'application logicielle utilisée dans le cadre de l'enquête. Dans le cas d'une application d'interview téléphonique assistée par ordinateur (ITAO), par exemple, chaque enregistrement doit comporter un code de province exact. Il exige, en outre, un codage exact du champ du fuseau horaire correspondant à la province et de chaque champ de numéro de téléphone. Une telle analyse de la base de sondage fournit des réactions importantes sur la qualité de la base utilisée dans le cadre de l'enquête.

8.2.2 Collecte des données

La formation des intervieweurs pour l'ECUI consistait à lire le Manuel de l'intervieweur, à mettre en pratique des cas simulés de l'ECUI sur ordinateur et à poser des questions à l'intervieweur principal 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'ECUI après avoir recueilli l'information de l'EPA. La période de collecte allait du 13 novembre au 6 décembre 2005.

8.2.3 Traitement des données

Au cours du traitement des données de l'ECUI, 32 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.

Le traitement des données de l'ECUI s'est déroulé par étapes, y compris la vérification, le codage, le contrôle, l'imputation, l'estimation, la confidentialité, etc. À chaque étape, une photo des fichiers de sortie est prise et il est facile de faire une vérification en comparant les fichiers de l'étape en cours avec ceux de l'étape antérieure. Cette façon de procéder a beaucoup amélioré l'étape de traitement des données.

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 personnes qui ont répondu au questionnaire d'enquête de façon à le contrebalancer pour ceux qui n'y ont 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. Afin de produire des données complètes concernant la répartition du revenu personnel et des ménages et permettant le calcul des totaux des commandes et des achats par l'entremise d'Internet, des valeurs ont été imputées lorsqu'elles étaient manquantes.

Les imputations ont été faites à l'aide de donneurs qui avaient été choisis au moyen d'une fonction de caractérisation. Pour tous les enregistrements contenant de la non-réponse ponctuelle ou de la non-réponse partielle (également appelés enregistrements receveurs), nous avons comparé certaines de leurs caractéristiques avec celles de tous les donneurs. Lorsque les caractéristiques étaient les mêmes entre un donneur et le receveur, une valeur était ajoutée au résultat de ce donneur. Le donneur ayant obtenu le résultat le plus élevé était réputé être le donneur « le plus semblable », et il était choisi pour fournir les éléments d'information manquants des non-répondants. S'il y avait plus d'un donneur avec le résultat le plus élevé, on procédait à une sélection aléatoire. Le bassin de donneurs était constitué de telle façon que la valeur imputée assignée au receveur, parallèlement à d'autres éléments d'information non imputés du receveur, serait encore acceptée à la vérification.

L'imputation a été faite en deux étapes. Tout d'abord, on a imputé le revenu personnel et celui du ménage (ensemble si nécessaire, et toujours en se servant du même donneur) et en deuxième lieu, les variables du commerce électronique, ces deux types d'imputation n'utilisant pas les mêmes données auxiliaires. Le tableau qui suit montre le taux d'imputation pour chacune des variables.

	Étape 1		Étape 2		
	Revenu personnel	Revenu du ménage	Commerce électronique	Nombre de commandes	Valeur des commandes
Imputé	8 670	10 816	159	216	297
Total	30 466	30 466	30 466	30 466	30 466
Taux (%)	28,5	35,5	0,5	0,7	1,0

Le processus d'imputation de l'ECUI a donné de bons résultats et il a contribué à compléter les réponses incomplètes avec l'expérience d'autres répondants ayant des caractéristiques semblables ou identiques. Il augmentera le nombre d'unités utilisées dans toute analyse effectuée par des chercheurs.

À noter que le fichier de microdonnées à grande diffusion ne renferme aucun indicateur d'imputation ni de variables de revenu. Ce fait aurait comme répercussion, une étape de confidentialité additionnelle.

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, l'on estime que 28,6 % des Canadiennes et des Canadiens n'avaient jamais utilisé Internet à la maison, au travail, à l'école ou à tout autre endroit (EV_Q01 = 2, NON), et l'on constate que l'erreur-type de cette estimation est de 0,017. Le coefficient de variation de l'estimation est donc calculé comme suit :

$$\left(\frac{0,017}{0,286} \right) \times 100 \% = 5,9 \%$$

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 canadienne sur l'utilisation d'Internet (ECUI) 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'ECUI, 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'ECUI.

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 personnes qui n'ont jamais utilisé Internet ou la proportion des personnes qui, au cours des 12 derniers mois, ont utilisé Internet à partir de la maison pour le courrier électronique constituent des exemples de telles estimations. Une estimation du nombre de personnes possédant une certaine caractéristique peut aussi être désignée une estimation d'un agrégat.

Exemples de questions catégoriques :

Q : Avez-vous déjà utilisé Internet (courrier électronique ou Web) à partir de la maison, du travail, de l'école ou de quelque autre endroit à des fins personnelles non commerciales?

R : Oui / Non

Q : Durant un mois type, à quelle fréquence utilisez-vous Internet à votre domicile?

R : Au moins une fois par jour / Au moins une fois par semaine (mais pas tous les jours) / Au moins une fois par mois (mais pas chaque semaine) / Moins d'une fois par mois

9.3.2 Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes et 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 personnes dans la population visée par l'enquête qui contribuent à cette quantité totale.

Un exemple d'estimation quantitative est le nombre moyen de commandes de produits ou de services qui ont été effectuées par les Canadiens en 2005 sur Internet. Le numérateur est une estimation du nombre total de commandes effectuées et son dénominateur est le nombre de personnes ayant fait au moins une commande.

Exemples de questions quantitatives :

Q : Au cours des 12 derniers mois, combien de commandes différentes de produits ou services avez-vous passées sur Internet?

R : |_|_|_| Nombre de transactions

Q : Au cours des 12 derniers mois, à combien estimez-vous, en dollars canadiens, la valeur totale des produits et services que vous avez commandés sur Internet?

R : |_|_|_|_|_| Montant total, arrondissez au dollar près

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 :

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

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. Par exemple, pour obtenir une estimation du nombre total de commandes de produits ou de services effectuées par les Canadiens en 2005 sur Internet et payées directement sur Internet en utilisant une carte de crédit ou de débit, multipliez la valeur déclarée à la question EC_Q03 (nombre de commandes différentes de produits ou services) par le poids final de l'enregistrement, puis additionnez cette valeur pour tous les enregistrements où la variable EC_Q07 = 1 (payé directement sur Internet (utilisant carte de crédit ou de débit)).

Pour obtenir une moyenne pondérée de la forme \hat{X} / \hat{Y} , le numérateur (\hat{X}) est calculé comme une estimation quantitative et le dénominateur (\hat{Y}) est calculé comme une estimation catégorique. Pour estimer, par exemple, le nombre moyen de commandes de produits ou de services effectuées par les Canadiens en 2005 sur Internet et payées directement sur Internet en utilisant une carte de crédit ou de débit,

- a) estimez le nombre total de commandes (\hat{X}) tel qu'il est décrit ci-dessus,
- b) estimez le nombre de personnes (\hat{Y}) inclus dans cette catégorie en additionnant les poids finals de tous les enregistrements où la variable EC_Q07 = 1, puis
- c) divisez l'estimation a) par l'estimation b) (\hat{X} / \hat{Y}).

9.4 Lignes directrices pour l'analyse statistique

L'ECUI 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 répondants de sexe masculin. Les étapes à suivre pour remettre à l'échelle les poids sont les suivantes :

- 1) sélectionner tous les répondants du fichier qui ont déclaré CSEX = 1, homme;
- 2) calculer le poids MOYEN pour ces enregistrements en additionnant les poids originaux des personnes établis à partir du fichier de microdonnées pour ces enregistrements puis diviser cette somme par le nombre de répondants ayant déclaré CSEX = 1;
- 3) pour chacun de ces répondants, calculer un poids REMIS À L'ÉCHELLE égal au poids original de la personne divisé par le poids MOYEN;
- 4) effectuer l'analyse portant sur ces répondants 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'ECUI, 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 de répondants 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	<p>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 %.</p> <p>Aucune mise en garde n'est requise.</p>
2) Médiocre	<p>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 %.</p> <p>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.</p>
3) Inacceptable	<p>Les estimations proviennent d'une taille d'échantillon inférieure à 30, ou présentent des coefficients de variation très élevés, supérieurs à 33,3 %.</p> <p>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 :</p> <p>« 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. »</p>

9.6 Seuils pour la diffusion des estimations pour l'Enquête canadienne sur l'utilisation d'Internet

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 7 000 personnes possédant une caractéristique donnée à Terre-Neuve-et-Labrador est médiocre.

Veillez 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	24 900 et plus	6 400 à < 24 900	moins de 6 400
Île-du-Prince-Édouard	7 500 et plus	1 900 à < 7 500	moins de 1 900
Nouvelle-Écosse	30 500 et plus	7 700 à < 30 500	moins de 7 700
Nouveau-Brunswick	27 200 et plus	6 900 à < 27 200	moins de 6 900
Québec	103 600 et plus	25 800 à < 103 600	moins de 25 800
Ontario	135 400 et plus	33 600 à < 135 400	moins de 33 600
Manitoba	29 900 et plus	7 500 à < 29 900	moins de 7 500
Saskatchewan	21 200 et plus	5 300 à < 21 200	moins de 5 300
Alberta	62 100 et plus	15 500 à < 62 100	moins de 15 500
Colombie-Britannique	82 200 et plus	20 600 à < 82 200	moins de 20 600
Provinces de l'Atlantique	28 400 et plus	7 100 à < 28 400	moins de 7 100
Manitoba et Saskatchewan	26 500 et plus	6 600 à < 26 500	moins de 6 600
Provinces des Prairies	50 900 et plus	12 600 à < 50 900	moins de 12 600
Canada	101 000 et plus	24 900 à < 101 000	moins de 24 900

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 canadienne sur l'utilisation d'Internet (ECUI).

Provinces et régions	Effet du plan	Taille de l'échantillon	Population
Terre-Neuve-et-Labrador	1,81	1 021	408 815
Île-du-Prince-Édouard	1,68	805	105 735
Nouvelle-Écosse	1,82	1 518	724 591
Nouveau-Brunswick	2,03	1 517	583 011
Québec	2,87	5 923	5 930 348
Ontario	3,36	8 653	9 640 686
Manitoba	2,10	2 094	843 239
Saskatchewan	1,90	2 252	707 607
Alberta	2,22	3 122	2 443 168
Colombie-Britannique	2,47	3 561	3 312 097
Provinces de l'Atlantique	2,10	4 861	1 822 152
Manitoba et Saskatchewan	2,06	4 346	1 550 846
Provinces des Prairies	2,63	7 468	3 994 014
Canada	3,41	30 466	24 699 297

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 personnes 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 personnes 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 le pourcentage repose sur un sous-groupe de la population. Par exemple, la proportion de personnes qui n'ont jamais utilisé Internet est plus fiable que le nombre estimé de personnes qui n'ont jamais utilisé Internet. Vous remarquerez que dans les tables, la valeur des coefficients de variation diminue lorsqu'on les lit de gauche à droite.

Lorsque la proportion ou 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 le pourcentage repose sur un sous-ensemble de la population totale (p. ex., le sexe ou un groupe d'âge particulier), 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 ($\hat{d} = \hat{X}_1 - \hat{X}_2$) est :

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

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 personnes ayant accès à un ordinateur ou un autre appareil permettant d'accéder à Internet à partir du domicile et le numérateur, le nombre de personnes ayant accès à un ordinateur ou un autre appareil permettant d'accéder à Internet à partir du domicile mais n'ayant pas actuellement une connexion à Internet à la maison.

Dans le cas où le numérateur n'est pas un sous-ensemble du dénominateur, comme dans l'exemple du rapport du nombre de personnes au Québec qui, au cours des 12 derniers mois, ont utilisé Internet à partir du domicile pour effectuer des opérations bancaires électroniques comparativement au nombre de personnes en Ontario qui, au cours des 12 derniers mois, ont utilisé Internet à partir du domicile pour effectuer des opérations bancaires électroniques, 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 ($\hat{R} = \hat{X}_1 / \hat{X}_2$) 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'Enquête canadienne sur l'utilisation d'Internet de 2005 et sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter.

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

Supposons qu'un utilisateur estime à 7 064 905 le nombre de Canadiens qui n'ont jamais utilisé Internet (EV_Q01 = 2, NON). 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 (7 064 905) 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 7 000 000.
- 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 1,7 %.

ENQUÊTE CANADIENNE SUR L'UTILISATION D'INTERNET, 2005													
Tables de variabilité d'échantillonnage approximative - Canada													
POURCENTAGE ESTIMÉ													
NUMÉRATEUR DU POURCENTAGE (en milliers)	0,1%	1,0%	2,0%	5,0%	10,0%	15,0%	...	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%
1	166,1	165,3	164,5	162,0	157,6	153,2		139,0	134,0	128,7	117,5	91,0	52,5
2	117,4	116,9	116,3	114,5	111,5	108,3		98,3	94,7	91,0	83,1	64,4	37,2
3	95,9	95,5	95,0	93,5	91,0	88,4		80,3	77,3	74,3	67,8	52,5	30,3
4	83,0	82,7	82,2	81,0	78,8	76,6		69,5	67,0	64,4	58,7	45,5	26,3
5	74,3	73,9	73,6	72,4	70,5	68,5		62,2	59,9	57,6	52,5	40,7	23,5
.													
.													
.													
60	****	21,3	21,2	20,9	20,4	19,8		17,9	17,3	16,6	15,2	11,7	6,8
65	****	20,5	20,4	20,1	19,6	19,0		17,2	16,6	16,0	14,6	11,3	6,5
70	****	19,8	19,7	19,4	18,8	18,3		16,6	16,0	15,4	14,0	10,9	6,3
75	****	19,1	19,0	18,7	18,2	17,7		16,1	15,5	14,9	13,6	10,5	6,1
80	****	18,5	18,4	18,1	17,6	17,1		15,5	15,0	14,4	13,1	10,2	5,9
85	****	17,9	17,8	17,6	17,1	16,6		15,1	14,5	14,0	12,7	9,9	5,7
90	****	17,4	17,3	17,1	16,6	16,1		14,7	14,1	13,6	12,4	9,6	5,5
95	****	17,0	16,9	16,6	16,2	15,7		14,3	13,7	13,2	12,1	9,3	5,4
100	****	16,5	16,4	16,2	15,8	15,3		13,9	13,4	12,9	11,7	9,1	5,3
125	****	14,8	14,7	14,5	14,1	13,7		12,4	12,0	11,5	10,5	8,1	4,7
150	****	13,5	13,4	13,2	12,9	12,5		11,4	10,9	10,5	9,6	7,4	4,3
200	****	11,7	11,6	11,5	11,1	10,8		9,8	9,5	9,1	8,3	6,4	3,7
250	****	****	10,4	10,2	10,0	9,7		8,8	8,5	8,1	7,4	5,8	3,3
300	****	****	9,5	9,4	9,1	8,8		8,0	7,7	7,4	6,8	5,3	3,0
350	****	****	8,8	8,7	8,4	8,2		7,4	7,2	6,9	6,3	4,9	2,8
400	****	****	8,2	8,1	7,9	7,7		7,0	6,7	6,4	5,9	4,6	2,6
450	****	****	7,8	7,6	7,4	7,2		6,6	6,3	6,1	5,5	4,3	2,5
500	****	****	****	7,2	7,0	6,9		6,2	6,0	5,8	5,3	4,1	2,3
750	****	****	****	5,9	5,8	5,6		5,1	4,9	4,7	4,3	3,3	1,9
1 000	****	****	****	5,1	5,0	4,8		4,4	4,2	4,1	3,7	2,9	1,7
1 500	****	****	****	****	4,1	4,0		3,6	3,5	3,3	3,0	2,3	1,4
2 000	****	****	****	****	3,5	3,4		3,1	3,0	2,9	2,6	2,0	1,2
3 000	****	****	****	****	****	****		2,6	2,5	2,4	2,3	2,1	1,7
4 000	****	****	****	****	****	****		2,2	2,1	2,0	1,9	1,4	0,8
5 000	****	****	****	****	****	****		2,0	1,9	1,8	1,7	1,3	0,7
6 000	****	****	****	****	****	****		1,8	1,7	1,7	1,5	1,2	0,7
7 000	****	****	****	****	****	****		1,7	1,6	1,5	1,4	1,1	0,6
8 000	****	****	****	****	****	****		****	1,5	1,4	1,3	1,0	0,6
9 000	****	****	****	****	****	****		****	****	1,4	1,2	1,0	0,6
10 000	****	****	****	****	****	****		****	****	****	1,2	0,9	0,5
12 500	****	****	****	****	****	****		****	****	****	****	0,8	0,5
15 000	****	****	****	****	****	****		****	****	****	****	0,7	0,4
20 000	****	****	****	****	****	****		****	****	****	****	****	0,4

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

- 4) Le coefficient de variation approximatif de l'estimation est donc 1,7 %. Le résultat selon lequel il y avait 7 064 905 Canadiens (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1) qui n'ont jamais utilisé Internet, peut être publié sans réserve.

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

Supposons qu'un utilisateur estime à $2\,224\,386 / 7\,064\,905 = 31,5 \%$ la proportion de personnes qui ont un ordinateur ou un autre appareil permettant d'accéder à Internet à partir du domicile (NU_Q03 = 1, OUI) parmi ceux qui n'ont jamais utilisé Internet (EV_Q01 = 2, NON). 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 personnes qui n'ont jamais utilisé Internet), il faut utiliser à la fois le pourcentage (31,5 %) et la portion numérateur du pourcentage (2 224 386) pour déterminer le coefficient de variation.
- 3) Le numérateur, 2 224 386, 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 2 000 000. 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 30,0 %.
- 4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 3,1 %, est le coefficient de variation à employer.
- 5) Le coefficient de variation approximatif de l'estimation est donc 3,1 %. Le résultat selon lequel 31,5 % des personnes qui n'ont jamais utilisé Internet ont un ordinateur ou un autre appareil permettant d'accéder à Internet à partir du domicile, peut être publié sans réserve.

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

Supposons qu'un utilisateur estime à $2\,722\,655 / 5\,930\,348 = 45,9 \%$ la proportion de personnes au Québec (PROVINCE = 24) qui ont utilisé Internet à partir du domicile au cours des 12 derniers mois pour le courrier électronique (SU_Q01 = 1, OUI) par rapport à $5\,891\,906 / 9\,640\,686 = 61,1 \%$ la proportion de personnes en Ontario (PROVINCE = 35) qui ont utilisé Internet à partir du domicile au cours des 12 derniers mois pour le courrier électronique (SU_Q01 = 1, OUI). Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

- 1) L'utilisation de la table 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 personnes au Québec de 1,7 % et un CV de l'estimation pour les personnes en Ontario de 1,4 %.

ENQUÊTE CANADIENNE SUR L'UTILISATION D'INTERNET, 2005

Tables de variabilité d'échantillonnage approximative - Québec
POURCENTAGE ESTIMÉNUMÉRATEUR DU
POURCENTAGE

(en milliers)	0,1%	1,0%	2,0%	5,0%	10,0%	15,0%	...	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%
1	169,3	168,6	167,7	165,1	160,7	156,2		141,8	136,6	131,2	119,8	92,8	53,6
2	119,7	119,2	118,6	116,8	113,7	110,5		100,2	96,6	92,8	84,7	65,6	37,9
3	97,8	97,3	96,8	95,3	92,8	90,2		81,8	78,9	75,8	69,2	53,6	30,9
4	84,7	84,3	83,9	82,6	80,4	78,1		70,9	68,3	65,6	59,9	46,4	26,8
5	75,7	75,4	75,0	73,9	71,9	69,9		63,4	61,1	58,7	53,6	41,5	24,0
.													
.													
.													
60	****	****	21,7	21,3	20,8	20,2		18,3	17,6	16,9	15,5	12,0	6,9
65	****	****	20,8	20,5	19,9	19,4		17,6	16,9	16,3	14,9	11,5	6,6
70	****	****	20,0	19,7	19,2	18,7		16,9	16,3	15,7	14,3	11,1	6,4
75	****	****	19,4	19,1	18,6	18,0		16,4	15,8	15,2	13,8	10,7	6,2
80	****	****	18,8	18,5	18,0	17,5		15,8	15,3	14,7	13,4	10,4	6,0
85	****	****	18,2	17,9	17,4	16,9		15,4	14,8	14,2	13,0	10,1	5,8
90	****	****	17,7	17,4	16,9	16,5		14,9	14,4	13,8	12,6	9,8	5,6
95	****	****	17,2	16,9	16,5	16,0		14,5	14,0	13,5	12,3	9,5	5,5
100	****	****	16,8	16,5	16,1	15,6		14,2	13,7	13,1	12,0	9,3	5,4
125	****	****	****	14,8	14,4	14,0		12,7	12,2	11,7	10,7	8,3	4,8
150	****	****	****	13,5	13,1	12,8		11,6	11,2	10,7	9,8	7,6	4,4
200	****	****	****	11,7	11,4	11,0		10,0	9,7	9,3	8,5	6,6	3,8
250	****	****	****	10,4	10,2	9,9		9,0	8,6	8,3	7,6	5,9	3,4
300	****	****	****	****	9,3	9,0		8,2	7,9	7,6	6,9	5,4	3,1
350	****	****	****	****	8,6	8,3		7,6	7,3	7,0	6,4	5,0	2,9
400	****	****	****	****	8,0	7,8		7,1	6,8	6,6	6,0	4,6	2,7
450	****	****	****	****	7,6	7,4		6,7	6,4	6,2	5,6	4,4	2,5
500	****	****	****	****	7,2	7,0		6,3	6,1	5,9	5,4	4,2	2,4
750	****	****	****	****	****	5,7		5,2	5,0	4,8	4,4	3,4	2,0
1 000	****	****	****	****	****	****		4,5	4,3	4,2	3,8	2,9	1,7
1 500	****	****	****	****	****	****		3,7	3,5	3,4	3,1	2,4	1,4
2 000	****	****	****	****	****	****		****	3,1	2,9	2,7	2,1	1,2
3 000	****	****	****	****	****	****		****	****	****	****	1,7	1,0
4 000	****	****	****	****	****	****		****	****	****	****	1,5	0,8
5 000	****	****	****	****	****	****		****	****	****	****	****	0,8

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

ENQUÊTE CANADIENNE SUR L'UTILISATION D'INTERNET, 2005

Tables de variabilité d'échantillonnage approximative - Ontario
POURCENTAGE ESTIMÉNUMÉRATEUR DU
POURCENTAGE

(en milliers)	0,1%	1,0%	2,0%	5,0%	10,0%	15,0%	...	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%
1	193,3	192,4	191,5	188,5	183,5	178,3		161,8	155,9	149,8	136,8	105,9	61,2
2	136,7	136,1	135,4	133,3	129,7	126,1		114,4	110,3	105,9	96,7	74,9	43,2
3	111,6	111,1	110,5	108,8	105,9	102,9		93,4	90,0	86,5	79,0	61,2	35,3
4	96,6	96,2	95,7	94,2	91,7	89,2		80,9	78,0	74,9	68,4	53,0	30,6
5	86,4	86,1	85,6	84,3	82,1	79,7		72,4	69,7	67,0	61,2	47,4	27,4
.													
.													
.													
60	****	24,8	24,7	24,3	23,7	23,0		20,9	20,1	19,3	17,7	13,7	7,9
65	****	23,9	23,7	23,4	22,8	22,1		20,1	19,3	18,6	17,0	13,1	7,6
70	****	23,0	22,9	22,5	21,9	21,3		19,3	18,6	17,9	16,3	12,7	7,3
75	****	22,2	22,1	21,8	21,2	20,6		18,7	18,0	17,3	15,8	12,2	7,1
80	****	21,5	21,4	21,1	20,5	19,9		18,1	17,4	16,7	15,3	11,8	6,8
85	****	20,9	20,8	20,4	19,9	19,3		17,6	16,9	16,2	14,8	11,5	6,6
90	****	20,3	20,2	19,9	19,3	18,8		17,1	16,4	15,8	14,4	11,2	6,4
95	****	19,7	19,6	19,3	18,8	18,3		16,6	16,0	15,4	14,0	10,9	6,3
100	****	****	19,1	18,8	18,3	17,8		16,2	15,6	15,0	13,7	10,6	6,1
125	****	****	17,1	16,9	16,4	15,9		14,5	13,9	13,4	12,2	9,5	5,5
150	****	****	15,6	15,4	15,0	14,6		13,2	12,7	12,2	11,2	8,6	5,0
200	****	****	****	13,3	13,0	12,6		11,4	11,0	10,6	9,7	7,5	4,3
250	****	****	****	11,9	11,6	11,3		10,2	9,9	9,5	8,6	6,7	3,9
300	****	****	****	10,9	10,6	10,3		9,3	9,0	8,6	7,9	6,1	3,5
350	****	****	****	10,1	9,8	9,5		8,6	8,3	8,0	7,3	5,7	3,3
400	****	****	****	9,4	9,2	8,9		8,1	7,8	7,5	6,8	5,3	3,1
450	****	****	****	8,9	8,6	8,4		7,6	7,4	7,1	6,4	5,0	2,9
500	****	****	****	****	8,2	8,0		7,2	7,0	6,7	6,1	4,7	2,7
750	****	****	****	****	6,7	6,5		5,9	5,7	5,5	5,0	3,9	2,2
1 000	****	****	****	****	****	5,6		5,1	4,9	4,7	4,3	3,3	1,9
1 500	****	****	****	****	****	****		4,2	4,0	3,9	3,5	2,7	1,6
2 000	****	****	****	****	****	****		3,6	3,5	3,3	3,1	2,4	1,4
3 000	****	****	****	****	****	****		****	2,8	2,7	2,5	1,9	1,1
4 000	****	****	****	****	****	****		****	****	****	2,2	1,7	1,0
5 000	****	****	****	****	****	****		****	****	****	****	1,5	0,9
6 000	****	****	****	****	****	****		****	****	****	****	1,4	0,8
7 000	****	****	****	****	****	****		****	****	****	****	****	0,7
8 000	****	****	****	****	****	****		****	****	****	****	****	0,7

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 ($\hat{d} = \hat{X}_1 - \hat{X}_2$) est :

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

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 $\hat{d} = 0,459 - 0,611 = -0,152$ est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0,459)(0,017)]^2 + [(0,611)(0,014)]^2} \\ &= \sqrt{(0,000061) + (0,000073)} \\ &= 0,012\end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,012 / 0,152 = 0,079$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 7,9 % ce qui peut être diffusé sans réserve.

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime à 2 722 655 le nombre personnes au Québec qui ont utilisé Internet à partir du domicile au cours des 12 derniers mois pour le courrier électronique (SU_Q01 = 1, OUI) et à 5 891 906 le nombre personnes en Ontario dans la même situation (SU_Q01 = 1, OUI). L'utilisateur est intéressé à comparer l'estimation des personnes au Québec à celle des personnes en 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 personnes au Québec qui ont utilisé Internet à partir du domicile au cours des 12 derniers mois pour le courrier électronique. Le dénominateur de l'estimation (\hat{X}_2) est le nombre personnes en Ontario qui ont utilisé Internet à partir du domicile au cours des 12 derniers mois pour le courrier électronique.
- 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 2 722 655. Le chiffre qui s'en rapproche le plus est 3 000 000. 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 QUÉBEC, soit 1,7 %.
- 4) Le dénominateur de cette estimation de rapport est 5 891 906. Le chiffre qui s'en rapproche le plus est 6 000 000. 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 ONTARIO, soit 1,4 %.
- 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 \hat{X}_2 respectivement.

C'est-à-dire que :

$$\begin{aligned}
 \alpha_{\hat{R}} &= \sqrt{(0,017)^2 + (0,014)^2} \\
 &= \sqrt{0,000289 + 0,000196} \\
 &= 0,022
 \end{aligned}$$

- 6) Le rapport obtenu entre les personnes au Québec et les personnes en l'Ontario qui ont utilisé Internet à partir du domicile au cours des 12 derniers mois pour le courrier électronique est 2 722 655 / 5 891 906, c'est-à-dire 0,46:1 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1). Le coefficient de variation de cette estimation est 2,2 %, 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 déterminée 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 $(\hat{X} - k, \hat{X} + k)$, 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 ($IC_{\hat{X}}$) :

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

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 Canadiens qui n'ont jamais utilisé Internet et qui ont un ordinateur ou un autre appareil permettant d'accéder à Internet à partir du domicile (d'après l'exemple 2 à la section 10.1.1) serait calculé comme suit :

$$\hat{X} = 31,5 \% \text{ (ou exprimé sous forme de proportion } 0,315)$$

$$t = 2$$

$$\alpha_{\hat{x}} = 3,1 \% \text{ (0,031 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.}$$

$$IC_{\hat{x}} = \{0,315 - (2) (0,315) (0,031), 0,315 + (2) (0,315) (0,031)\}$$

$$IC_{\hat{x}} = \{0,315 - 0,020, 0,315 + 0,020\}$$

$$IC_{\hat{x}} = \{0,295, 0,335\}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 29,5 % et 33,5 % des Canadiens qui n'ont jamais utilisé Internet ont un ordinateur ou un autre appareil permettant d'accéder à Internet à partir du domicile.

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

Des erreurs-types peuvent aussi être utilisées 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, elle 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 personnes au Québec qui ont utilisé Internet à partir du domicile, au cours des 12 derniers mois, pour le courrier électronique et la proportion de personnes en Ontario qui ont utilisé Internet à partir du domicile, au cours des 12 derniers mois, pour le courrier électronique. 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 0,012. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0,459 - 0,611}{0,012} = \frac{-0,152}{0,012} = -12,7$$

Puisque $t = -12,7$ est inférieur 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 la plupart des variables pour l'ECUI 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 commandes de produits ou de services effectuées par les Canadiens en 2005 sur Internet serait supérieur au coefficient de variation de la proportion correspondante de Canadiens qui ont effectué une commande de produits ou de services. 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 des coefficients de variation

Consulter le fichier ECUI2005_CVTabF.pdf pour les tables de coefficient de variation.

11.0 Pondération

Puisque l'Enquête canadienne sur l'utilisation d'Internet (ECUI) 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 canadienne sur l'utilisation d'Internet

Les principes qui sous-tendent le calcul des poids pour l'ECUI sont identiques à ceux établis aux fins de l'EPA. Toutefois, 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'ECUI.

- 1) Un ajustement qui tient compte de l'utilisation d'un sous-échantillon représentant les cinq sixièmes, plutôt que de l'échantillon complet de l'EPA.
- 2) Un ajustement qui tient compte de la sélection aléatoire d'un répondant à partir du ménage sélectionné.
- 3) Un ajustement qui tient compte des non-réponses additionnelles au questionnaire de l'enquête supplémentaire, c'est-à-dire des non-réponses au questionnaire de l'ECUI attribuables à des personnes qui ont répondu au questionnaire de l'EPA ou pour qui des données de l'EPA du mois précédent ont été reportées. La procédure est similaire à l'ajustement des poids des non-réponses au questionnaire de l'EPA, mais les regroupements reposent sur des variables différentes.
- 4) Un dernier ajustement est fait afin d'apparier les projections du recensement pour des groupes d'âge-sexe-province indépendants et des chiffres de régions métropolitaines de recensement (à l'intérieur d'une opération d'étalonnage).

Le poids qui en résulte, WTPM, est le poids final qui figure dans le fichier de microdonnées de l'ECUI.

12.0 Questionnaires

12.1 Le questionnaire de l'Enquête sur la population active

Le questionnaire de l'Enquête sur la population active (EPA_QuestF.pdf) sert à recueillir des données au sujet de l'activité actuelle et de la dernière activité sur le marché du travail de tous les membres d'un ménage âgés de 15 ans ou plus. Il comprend des questions sur les heures de travail, l'ancienneté dans l'emploi, le type de travail, la raison des heures perdues ou d'absence, la recherche d'emploi réalisée, la disponibilité pour le travail et la fréquentation scolaire.

12.2 Le questionnaire de l'Enquête canadienne sur l'utilisation d'Internet

Le questionnaire de l'Enquête canadienne sur l'utilisation d'Internet (ECUI) a servi à recueillir des données pour l'enquête supplémentaire en 2005. Le fichier ECUI2005_QuestF.pdf renferme le questionnaire français.

13.0 Cliché d'enregistrement à valeurs univariées

Consulter le fichier ECUI2005_LvCds.pdf pour le cliché d'enregistrement à chiffres univariés.