

Guide de l'utilisateur des microdonnées

Enquête sur les travailleurs âgés

2008



Statistique
Canada

Statistics
Canada

Canada

Table des matières

1.0	Introduction	7
2.0	Contexte	9
3.0	Objectifs	11
4.0	Concepts et définitions	13
4.1	Concepts et définitions de l'Enquête sur la population active	13
4.2	Concepts et définitions de l'Enquête sur les travailleurs âgés	14
5.0	Méthodologie de l'enquête	17
5.1	Population visée	17
5.2	Plan de sondage	17
5.2.1	Stratification primaire	17
5.2.2	Types de régions	17
5.2.3	Stratification secondaire	18
5.2.4	Délimitation et sélection des grappes	18
5.2.5	Sélection des logements	19
5.2.6	Sélection des personnes	19
5.3	Taille de l'échantillon	19
5.4	Renouvellement de l'échantillon	20
5.5	Modifications apportées au plan de l'Enquête sur la population active pour l'Enquête sur les travailleurs âgés	20
5.6	Taille de l'échantillon selon la province pour l'Enquête sur les travailleurs âgés	21
6.0	Collecte des données	23
6.1	Réalisation des interviews dans le cadre de l'Enquête sur la population active	23
6.2	Supervision et contrôle de qualité	23
6.3	Non-réponse à l'Enquête sur la population active	24
6.4	Modifications apportées à la collecte des données de l'Enquête sur les travailleurs âgés	24
6.5	Non-réponse à l'Enquête sur les travailleurs âgés	24
7.0	Traitement des données	25
7.1	Saisie des données	25
7.2	Vérification	25
7.3	Codage des questions ouvertes	26
7.4	Imputation	26
7.5	Création de variables dérivées	26
7.6	Pondération	26
7.7	Suppression de renseignements confidentiels	27
8.0	Qualité des données	29
8.1	Taux de réponse	29
8.2	Erreurs relatives à l'enquête	29
8.2.1	Base de sondage	30
8.2.2	Collecte des données	30
8.2.3	Traitement des données	30
8.2.4	Non-réponse	31
8.2.5	Mesure de l'erreur d'échantillonnage	32

9.0	Lignes directrices pour la totalisation, l'analyse et la diffusion de données	33
9.1	Lignes directrices pour l'arrondissement d'estimations	33
9.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation	34
9.3	Définitions de types d'estimations : catégoriques et quantitatives	34
9.3.1	Estimations catégoriques.....	34
9.3.2	Estimations quantitatives	34
9.3.3	Totalisation d'estimations catégoriques.....	35
9.3.4	Totalisation d'estimations quantitatives	35
9.4	Lignes directrices pour l'analyse statistique	36
9.5	Lignes directrices pour la diffusion de coefficients de variation.....	37
9.6	Seuils pour la diffusion des estimations pour l'Enquête sur les travailleurs âgés	39
10.0	Tables de variabilité d'échantillonnage approximative	41
10.1	Comment utiliser les tables de coefficients de variation pour des estimations catégoriques.....	42
10.1.1	Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques	43
10.2	Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance	48
10.2.1	Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance	49
10.3	Comment utiliser les tables de coefficients de variation pour effectuer un test t.....	49
10.3.1	Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t	50
10.4	Coefficients de variation pour des estimations quantitatives.....	50
10.5	Tables des coefficients de variation.....	51
11.0	Pondération	53
11.1	Procédures de pondération pour l'Enquête sur la population active	53
11.2	Procédures de pondération pour l'Enquête sur les travailleurs âgés	54
12.0	Questionnaires	57
12.1	Le questionnaire de l'Enquête sur la population active	57
12.2	Le questionnaire de l'Enquête sur les travailleurs âgés	57
13.0	Cliché d'enregistrement à valeurs univariées	59

1.0 Introduction

Statistique Canada a mené l'Enquête sur les travailleurs âgés (ETA) le 2008 avec l'appui et la collaboration de Ressources humaines et Développement des compétences 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

Services à la clientèle

Division des enquêtes spéciales

Téléphone : 613-951-3321 ou appelez sans frais : 1-800-461-9050

Télécopieur : 613-951-4527

Courriel : des@statcan.ca

Ressources humaines et Développement des compétences Canada

Joni Baran

140 Promenade du Portage

Gatineau, Quebec

Téléphone : 819-997-5322

Télécopieur : 819-953-8479

Courriel : joni.baran@hrsdc-rhdcc.gc.ca

2.0 Contexte

Ces données appuient les initiatives liées à la politique du travail au gouvernement fédéral. Elles orientent l'élaboration des politiques et peuvent aider à la prise de décisions futures.

Les données d'enquête serviront à fournir :

- des estimations nationales et provinciales des ménages comptant des travailleurs âgés de 50 ans ou plus et des estimations à savoir si ces travailleurs ont connu des épisodes de chômage après l'âge de 50 ans;
- des renseignements de niveau national sur les causes d'un épisode de chômage et ce qui peut suivre un épisode de chômage;
- des estimations nationales du nombre d'années de travail auprès du dernier employeur;
- des estimations nationales des types de considérations qui pourraient encourager de nouveaux retraités à réintégrer la population active;
- des estimations nationales de la principale source d'information sur la planification de la retraite;
- des estimations nationales des projets de travail ou de départ à la retraite selon la branche d'activité et la profession;
- des estimations nationales des projets de travail ou de départ à la retraite selon le type de régime de pension;
- des estimations nationales des obstacles recensés à l'épargne retraite.

3.0 Objectifs

Cette enquête sert à recueillir des renseignements qui jetteront de la lumière sur les intentions, les projets et les facteurs connexes qui peuvent influencer sur la décision des travailleurs âgés de partir à la retraite, de continuer à exercer leur « emploi de carrière » ou de chercher du travail dans une autre profession.

Les principaux objectifs de l'enquête consistent à déterminer :

- les facteurs liés à la santé qui influent sur la décision de se retirer d'un « emploi de carrière », de continuer de travailler, ou de commencer un nouvel emploi;
- les considérations financières personnelles qui se répercutent sur la décision de se retirer d'un « emploi de carrière », de continuer de travailler ou de commencer un nouvel emploi;
- le rôle que jouent les régimes de pension dans la décision de départ à la retraite ou de travail;
- le rôle que jouent les personnes à charge dans la décision de départ à la retraite ou de travail;
- le rôle que jouent la profession, la branche d'activité, le sexe, l'état matrimonial et la région géographique dans la décision de départ à la retraite ou de travail.

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 l'intérieur de la section 4.1, tandis que ceux et celles propres à l'Enquête sur les travailleurs âgés (ETA) 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 ou 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 sur les travailleurs âgés

Assurance-emploi (AE)

L'assurance-emploi procure une aide financière temporaire pour les chômeurs canadiens qui sont à la recherche d'un emploi ou qui se recyclent. Les personnes malades, les femmes enceintes, les personnes qui s'occupent d'un nouveau-né ou d'un enfant adopté et les personnes qui s'occupent d'un membre de leur famille gravement malade présentant un risque important de décès ont aussi droit à l'assurance-emploi.

Autres épargnes ou investissements privés

Toute épargne (dans un compte ou sous autre forme) ou placements (p. ex., actions, obligations et fonds commun de placement) que détient un répondant pour lui tenir lieu de source de revenu.

Avoir

Valeur des actifs d'une entreprise ou d'une personne une fois que tous les facteurs ont été pris en compte, soustraits ou réglés.

Catégorie de travailleurs

Indique si une personne est un *employé* (c'est-à-dire si elle travaille contre un traitement ou un salaire, une commission, ou un paiement en nature), un *travailleur autonome* dans une exploitation agricole, une entreprise ou dans l'exercice d'une profession libérale (constituée en société ou non, avec ou sans aide rémunérée).

Compte

Relevé financier des recettes et des paiements, des revenus et des dépenses, des crédits et des débits.

Emploi de carrière

Emploi exercé pendant au moins huit ans à temps plein.

Fonds enregistré de revenu de retraite (FERR)

Une personne crée un FERR à une institution financière. Le FERR est enregistré en vertu de la *Loi de l'impôt sur le revenu* et vise à procurer un revenu de retraite régulier tout en permettant le cumul de l'intérêt sur le capital. Le FERR fonctionne en gros comme une rente. Pour créer le FERR, la personne doit y transférer directement les sommes accumulées dans un Régime enregistré d'épargne-retraite ou les sommes forfaitaires provenant des Régimes de pension agréés. Les sommes retirées du FERR sont imposables. Chaque année, la personne doit retirer une somme minimale à partir de l'année subséquente à la création du FERR.

Hypothèque inversée

Une hypothèque inversée permet à un propriétaire d'obtenir un prêt sur sa résidence pendant qu'il y vit encore.

Indemnité pour accident de travail

Les programmes d'indemnité pour accident de travail protègent les employés contre les difficultés financières liées aux accidents de travail et aux maladies professionnelles.

Ménage

Toute personne ou tout groupe de personnes vivant dans un logement privé. Le ménage peut consister en une personne vivant seule, ou en un groupe de personnes qui ne sont pas apparentées, mais qui partagent le même logement, ou il peut s'agir d'une famille.

Mise à pied

Action de suspendre ou de renvoyer un travailleur sans que la faute lui en revienne.

Pension d'ancien combattant

La *Loi sur les pensions* accorde une pension à toute personne qui souffre d'une incapacité liée au service militaire en temps de paix ou en temps de guerre. La prestation est fonction de la gravité de l'incapacité déterminée par un examen médical et elle est versée selon les taux prévus par la Loi.

Au décès d'un bénéficiaire d'une pension d'invalidité, son conjoint ou conjoint de fait peut recevoir, pendant un an, la même pension. Après cette période d'un an, le survivant reçoit la pension de survivant. Les enfants survivants peuvent être admissibles à une prestation pour orphelin à la suite du décès du bénéficiaire. Le conjoint ou conjoint de fait survivant continue de recevoir les prestations de survivant même s'il se marie de nouveau.

Pension de retraite

Pension ou paiement versé à une personne qui prend sa retraite d'un emploi à temps plein dès qu'elle atteint l'âge de la retraite. Le terme désigne aussi les cotisations accumulées par les employeurs et les employés dans une caisse de retraite.

Régime de pension agréé (RPA)

Un RPA est parrainé par un employeur ou un syndicat et est habituellement financé au moyen des cotisations des employés et de l'employeur. Les RPA doivent remplir certaines conditions et être agréés aux fins de la *Loi de l'impôt sur le revenu* fédérale. Les cotisations à ces régimes sont déductibles dans le calcul du revenu imposable, ce qui fait que l'impôt payable sur le revenu de placements fait l'objet d'un report et que les prestations versées sont imposables.

Régime de pension d'employeur

Régime de pension offert par une entreprise ou un organisme à ses employés. Dans les régimes contributifs, l'employeur et l'employé cotisent à un fonds non imposable durant la période d'épargne. Dans les régimes non contributifs, seul l'employeur contribue au régime. Bien qu'il existe divers régimes de pension d'employeur, ce sont en général des régimes à cotisations déterminées ou à prestations déterminées.

Régime de pensions du Canada (RPC) et Régime de rentes du Québec (RRQ)

Ces régimes sont des programmes d'assurance sociale à gains contributifs qui tiennent lieu de mesures de protection du revenu aux cotisants et à leur famille contre la perte de revenu causée par la retraite, l'invalidité ou le décès. Le RPC a été instauré dans tous les territoires et les provinces, sauf le Québec, qui utilise le RRQ similaire. Une personne peut recevoir des prestations du RPC dès 60 ans.

Régime de retraite professionnel

Régime de retraite instauré par une compagnie ou une organisation au profit de ses employés. Dans un régime contributif, l'employeur et ses employés contribuent ensemble à l'accumulation d'un fonds exempt d'impôt durant la période de cotisation. Dans un régime non contributif, seul l'employeur contribue.

Régime enregistré d'épargne-retraite (REER)

Les REER procurent un revenu de retraite en fonction des cotisations accumulées et de l'intérêt obtenu sur ces cotisations. Ils sont achetés par la personne elle-même ou par le conjoint. Le statut professionnel n'est pas un facteur pertinent lors de l'achat d'un REER. Les cotisations à un REER sont déductibles d'impôt et le revenu de l'investissement est imposable de manière différée. Une personne peut encaisser en tout temps son REER, qui est alors imposé en conséquence. Une personne peut cotiser à un REER jusqu'au jour de son 71^e anniversaire. Elle doit ensuite convertir son REER en FERR (Fonds enregistré de revenu de retraite) avant la fin de l'année de son 71^e anniversaire.

Rente

Option de gestion financière qui procure des paiements réguliers à partir d'un capital (somme forfaitaire) et qui permet le cumul de l'intérêt sur le capital.

Retraité

Dans le contexte de cette enquête, désigne une personne de 50 ans ou plus qui a quitté son emploi ou son employeur principal, et dont au moins 50 % du revenu total provient de sources de revenu de retraite. Il est possible de prendre une retraite, puis de retourner sur le marché du travail dans d'autres fonctions.

Sécurité de la vieillesse (SV) et Supplément de revenu garanti (SRG)

Les prestations de la SV sont des paiements mensuels imposables versés aux personnes âgées de 65 ans ou plus, en fonction de leur nombre d'années de résidence au Canada. Pour sa part, le Supplément de revenu garanti n'est pas imposable et est versé aux prestataires de la SV à faible revenu. Ces deux types de prestations sont fondés sur l'étude du revenu et peuvent faire l'objet d'un mécanisme de récupération si le revenu augmente.

Travailleur à temps plein

Personne qui dit avoir travaillé au moins 30 heures par semaine la plupart des semaines où elle était employée.

Travailleur à temps partiel

Personne qui travaillait moins de 30 heures par semaine. (Dans le cas des personnes qui occupaient plus d'un emploi au cours de la même semaine, les heures consacrées à tous les emplois ont été combinées.)

Travailleur âgé

Aux fins de l'Enquête auprès des travailleurs âgés, un particulier doit être âgé de 50 à 75 ans et participer au marché du travail ou avoir pris sa retraite il y a tout au plus 24 mois.

5.0 Méthodologie de l'enquête

Comme l'Enquête sur les travailleurs âgés (ETA) a été menée en octobre et novembre 2008 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'ETA s'est écartée du plan de base de l'EPA de automne 2008.

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.

5.2.1 Stratification primaire

Les provinces sont divisées en régions économiques (RE) et en régions économiques d'assurance-emploi (REAE). Les RE 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 REAE sont également des régions géographiques et en gros ont la même taille et sont aussi nombreuses que les RE, mais ne sont pas définies de la même façon. On produit des estimations de la population active pour les REAE aux fins de Ressources humaines et Développement des compétences Canada.

Les intersections des régions des deux types susmentionnés forment le premier niveau de stratification pour l'EPA. On traite ces intersections de RE et de REAE 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 REAE.

5.2.2 Types de régions

Les strates primaires (les intersections de RE et de REAE) 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^o 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 grand 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 et 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 de 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 sur les travailleurs âgés

On a utilisé pour l'ETA cinq des six groupes de renouvellement inclus dans l'échantillon de l'EPA. Les groupes de renouvellement 2, 5 et 6 en octobre et 1 et 3 en novembre ont sélectionné. Pour les besoins de l'ETA, on a modifié la couverture de l'EPA afin d'y inclure tous les membres d'un ménage âgés de 50 à 75 ans qui soit travaillaient toujours ou qui ont pris la retraite dans les derniers 24 mois. Cependant, contrairement à l'EPA, pour laquelle des données sont recueillies pour tous les membres admissibles d'un ménage, l'ETA n'a donné lieu à la cueillette d'information qu'auprès d'un seul membre présélectionné d'un ménage. Les réponses par personne interposée n'étaient pas non plus permises pour l'ETA.

5.6 Taille de l'échantillon selon la province pour l'Enquête sur les travailleurs âgés

Le tableau qui suit montre le nombre de répondants inclus dans les groupes de renouvellement échantillonnés de l'EPA qui étaient admissibles pour le supplément de l'ETA.

Province	Taille de l'échantillon
Terre-Neuve-et-Labrador	484
Île-du-Prince-Édouard	347
Nouvelle-Écosse	707
Nouveau-Brunswick	635
Québec	2 263
Ontario	3 641
Manitoba	933
Saskatchewan	927
Alberta	1 192
Colombie-Britannique	1 489
Canada	12 618

6.0 Collecte des données

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

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

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

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

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

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

À la fin des interviews mensuelles menées pour l'EPA, les intervieweurs présentent le questionnaire d'enquête supplémentaire, le cas échéant, qu'il faut faire remplir à certains ou à la totalité des membres 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 sur les travailleurs âgés

On a fait remplir le questionnaire de l'Enquête sur les travailleurs âgés (ETA) par une personne par ménage sélectionnée au hasard. 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 demandait à parler à la personne sélectionnée pour l'ETA. Si la personne sélectionnée n'était pas disponible, l'intervieweur fixait un moment opportun pour rappeler. Comme les « réponses par procuration » n'étaient pas permises à moins que c'était pour aider à la communication avec la personne sélectionnée (c.à.d. traduction, incapacité physique ou mentale, etc.) et que la personne sélectionnée était présente) On a prolongé d'une semaine 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 sur les travailleurs âgés

Pour les ménages ayant répondu au questionnaire de l'EPA, l'étape suivante de la collecte des données consistait à faire remplir le questionnaire de l'ETA. Lorsque cela était possible, l'interview pour l'ETA a été effectuée au même temps que l'interview pour l'EPA. Dans certains cas, un membre du ménage a répondu à l'EPA et un autre, à l'ETA. Si le répondant sélectionné ne pouvait pas faire l'interview de l'ETA immédiatement, l'intervieweur fixait un rendez-vous. La collecte des données pour l'ETA s'est poursuivie pendant quatre jours supplémentaires après la semaine de l'EPA. L'ETA commençait toujours pendant la semaine de l'EPA. Comme toujours dans le cas des enquêtes supplémentaires, l'interview pour l'EPA avait priorité sur celle pour l'ETA.

7.0 Traitement des données

Le principal produit de l'Enquête sur les travailleurs âgés (ETA) 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

Les fichiers-textes électroniques contenant les transmissions quotidiennes de cas terminés sont combinés en un seul fichier d'enquête « brut ». À la fin de la collecte, ce fichier doit comprendre un enregistrement pour chaque personne figurant dans l'échantillon. Avant tout autre traitement, on effectue une vérification pour repérer et supprimer d'éventuels enregistrements en double et exclure les enregistrements classés « non-réponse » ou « hors champ ».

Dans certaines circonstances, des répondants peuvent être considérés comme étant hors du champ de l'ETA. La vaste majorité des cas échantillonnés classés hors champ ont trait à des répondants qui n'ont pas satisfait au critère de l'âge (50 à 75 ans) ou au critère de l'appartenance à la population active (avoir travaillé au cours des 24 mois précédant l'interview).

Par conséquent, la correction se fait en modifiant les données au niveau de la variable individuelle. La première étape de la vérification consiste à déterminer quels sont les éléments des résultats de l'enquête qui doivent être conservés dans le fichier maître de l'enquête. Par la suite, les caractères invalides sont supprimés et les données élémentaires sont dûment formatées. Les champs de texte sont retirés des fichiers principaux et ajoutés à un fichier distinct pour le codage. Le premier type d'erreurs traitées était les erreurs dans l'enchaînement des questions : des questions qui ne s'appliquaient pas au répondant ont été relevées (et auxquelles on n'aurait donc pas dû répondre) et renfermaient des réponses. 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, subséquentes. Dans le cas des sauts de question fondés sur des questions auxquelles on a répondu, toutes les questions sautées sont codées « Saut valide » (6, 96, 996, etc.). Dans le cas des sauts de question fondés sur des réponses « Ne sais pas » ou « Refus », on attribue à toutes les questions sautées le code « Non déclaré » (9, 99, 999, etc.). Aux éléments vides restants, on attribue une valeur numérique (9, 99, 999, etc., selon la longueur de la variable). Ces codes sont réservés à des fins de traitement et signifient que l'élément était « Non déclaré ».

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 11 questions partiellement ou complètement ouvertes. Il s'agit d'éléments relatifs, entre autres, aux aspects suivants : ce que l'on pourrait faire pour inciter les sans-emploi à retourner sur le marché du travail, les facteurs ayant une incidence sur la retraite ou sur le retour au travail après la retraite, les sources de revenus, les raisons de la suppression d'emploi, etc.

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'ETA, il n'y a pas l'imputation par enregistrement donneur a été utilisée pour fournir les éléments d'information manquants aux enregistrements avec non-réponse ponctuelle ou partielle.

7.5 Création de variables dérivées

Au total de 36 données élémentaires incluses dans le fichier de microdonnées ont été calculées en combinant des postes sur le questionnaire pour faciliter l'analyse des données. La plupart sont des variables continues associées à l'âge et aux nombres d'années de service. Ces variables ont été regroupées dans des intervalles pré-déterminés dans le but d'assister le processus analytique.

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 âgées de 65 à 75 ans qui travaillent toujours doit être estimé, cette opération s'effectue en sélectionnant les enregistrements se référant aux individus 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 inclut des identificateurs géographiques plus explicites que le FMGD. Notamment, on y trouve la région économique et la région métropolitaine de recensement en plus de la province. Le FMGD ne renferme pas d'identificateurs géographiques inférieurs au niveau provincial.

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 texte utilisés à l'intérieur des questions ouvertes sont agrégés dans le FMGD. Ils sont regroupés pour convenir aux principaux schémas de codage.

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 les travailleurs âgés (ETA).

Provinces	Échantillon dans le champ de l'enquête	Réponse	Taux de réponse (%)
Terre-Neuve-et-Labrador	484	380	78,5
Île-du-Prince-Édouard	347	274	79,0
Nouvelle-Écosse	707	562	79,5
Nouveau-Brunswick	635	490	77,2
Québec	2 263	1 827	80,7
Ontario	3 641	2 947	80,9
Manitoba	933	776	83,2
Saskatchewan	927	776	83,7
Alberta	1 192	915	76,8
Colombie-Britannique	1 489	1 087	73,0
Canada	12 618	10 034	79,5

Note : Le taux de réponse à l'ETA est le nombre d'individus ayant répondu au questionnaire de l'ETA exprimé sous forme de pourcentage du nombre d'individus dans le champ de l'enquête sélectionnés pour l'ETA.

8.2 Erreurs relatives à l'enquête

Les estimations calculées à partir de cette enquête reposent sur un sous-échantillon d'individus de l'Enquête sur la population active (EPA). 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 du cycle de collecte et de traitement des données afin de contrôler la qualité des données. Ces mesures comprenaient le recours à des intervieweurs

hautement qualifiés, une formation poussée des intervieweurs concernant les procédures d'enquête et du questionnaire, l'observation des intervieweurs en vue de cerner les problèmes liés à la conception du questionnaire ou à une mauvaise compréhension des instructions, des procédures visant à s'assurer que les erreurs de saisie des données étaient réduites au minimum ainsi que des vérifications de la qualité du codage et de contrôle ayant pour but d'attester la logique du traitement.

8.2.1 Base de sondage

Comme l'ETA 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'ETA. La qualité des variables d'échantillon dans la base était très élevée. L'échantillon de l'ETA était composé de 5 groupes de renouvellement 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'ETA 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. La base de sondage de l'ETA exclus aussi les non-réponses complètes à l'EPA et les non-réponses partielles aux variables utilisées dans les critères de sélection.

Les variables de la base de sondage de l'ETA étaient assez à jour puisqu'elles avaient été tirées de l'EPA au plus quatre semaines avant le début de la collecte des données de l'ETA.

8.2.2 Collecte des données

La formation des intervieweurs pour l'ETA consistait à lire le Manuel du surveillant, le Manuel des procédures, le Manuel de l'intervieweur, la pratique avec des cas simulés de l'ETA sur ordinateur pour d'interview assistée par ordinateur seulement et une discussion avec l'intervieweur principal des questions de l'intervieweur avant le début de l'enquête. L'information fournie aux intervieweurs comprenait une description du contexte et des objectifs de l'enquête ainsi qu'un glossaire de terminologie et une série de questions et de réponses. Les intervieweurs ont recueilli l'information pour l'ETA après avoir recueilli l'information de l'EPA. La période de collecte allait de la semaine du 19 octobre au 23 novembre 2008.

Lorsque cela était possible, l'interview pour l'ETA a été effectuée au même temps que l'interview pour l'EPA. Dans certains cas, un membre du ménage a répondu à l'EPA et un autre, à l'ETA. Si le répondant choisi ne pouvait pas répondre à l'interview immédiatement, l'intervieweur fixait un rendez-vous. L'ETA s'est poursuivie pendant quatre jours de plus après la semaine de l'EPA. L'ETA commençait toujours durant la semaine de l'EPA. Comme toujours dans le cas des enquêtes supplémentaires, l'interview pour l'EPA avait priorité sur l'interview pour l'ETA.

Dans les cas où l'interview pour l'EPA était terminée mais non l'interview pour l'ETA, le dossier de l'ETA était versé dans un groupe Z distinct pour être complété.

8.2.3 Traitement des données

Au cours du traitement des données de l'ETA, 641 enregistrements n'ont pu être jumelés à un enregistrement équivalent sur l'EPA. Ceux-ci ont donc été codés comme étant des non-réponses. Parmi ces enregistrements, il y a ceux des répondants qui ont répondu aux questions de l'EPA, mais non aux questions de l'ETA, ou qui n'ont pas répondu à un

nombre suffisant de questions de l'ETA pour que l'enregistrement puisse être considéré comme utilisable.

Le traitement des données de l'ETA a été fait 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.

En raison du nombre important d'opérations de vérification intégrées à l'application d'interview téléphonique assistée par ordinateur (ITAO), l'ensemble de données de l'ETA n'a requis que deux corrections.

- Premièrement, dans le cas de 15 enregistrements, on a attribué à PP_Q05 le code « Non déclaré » après avoir découvert que l'âge auquel les répondants prévoient présenter leur demande de pension du Régime de pensions du Canada ou leur demande de rente du Régime de rentes du Québec était antérieur à l'âge qu'ils avaient au moment de l'enquête.
- Deuxièmement, à la question DW_Q03, on demandait si le répondant était toujours sans emploi après avoir été déplacé. Dans 67 cas, on a attribué le code « Non déclaré » à DW_Q03 parce que la réponse ne concordait pas avec la situation d'emploi actuelle (CS_Q01).

Aucune autre donnée élémentaire n'a été imputée, que ce soit de façon déterministe ou autrement.

Enfin, pour éviter la divulgation de renseignements personnels des répondants, toutes les variables numériques ainsi que certaines variables codées ont été groupées.

8.2.4 Non-réponse

Provinces	Nombre de non-répondants	Nombre de répondants	Taux de réponse (%)	Nombre de répondants connus hors champ	Taille globale de l'échantillon (dans le champ de l'enquête)
Terre-Neuve-et-Labrador	104	380	78,5	0	484
Île-du-Prince-Édouard	73	274	79,0	0	347
Nouvelle-Écosse	145	562	79,5	0	707
Nouveau-Brunswick	145	490	77,2	0	635
Québec	436	1 827	80,7	2	2 263
Ontario	694	2 947	80,9	7	3 641
Manitoba	157	776	83,2	0	933
Saskatchewan	151	776	83,7	1	927
Alberta	277	915	76,8	1	1 193
Colombie-Britannique	402	1,087	73,0	0	1 489
Canada	2 584	10 034	79,5	11	12 618

D'après notre examen des constatations, il est clair que les catégories où les caractéristiques sociodémographiques des répondants choisis ayant eu une incidence sur le taux de réponse ont été très peu nombreuses. Géographiquement, le taux de réponse observé en Colombie-Britannique a été inférieur à tout autre taux de réponse

provincial, mais ce phénomène est une tendance observée actuellement à l'échelle de l'organisme et ne se limite pas à l'ETA. La mise en œuvre de l'ETA comme enquête supplémentaire volontaire de l'EPA a apporté un taux de réponse très satisfaisant de 79,5 pour cent.

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.

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'ETA de 2008, l'on estime que 71,16 % des individus faisant partis de la population de l'ETA à Île-du-Prince-Édouard qui travaillent toujours et n'ont jamais été à la retraite et l'on constate que l'erreur-type de cette estimation est de 0,0311. Le coefficient de variation de l'estimation est donc calculé comme suit :

$$\left(\frac{0,0311}{0,7116} \right) \times 100 \% = 4,37 \%$$

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

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

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

9.1 Lignes directrices pour l'arrondissement d'estimations

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

- a) Les estimations dans le corps principal d'un tableau statistique doivent être arrondies à la centaine près à l'aide de la technique d'arrondissement normale. Selon cette technique, si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, le dernier chiffre à conserver est augmenté de 1. Par exemple, selon la technique d'arrondissement normale à la centaine près, si les deux derniers chiffres se situent entre 00 et 49, ils sont remplacés par 00 et le chiffre précédent (le chiffre des centaines) reste inchangé. Si les derniers chiffres se situent entre 50 et 99, ils sont remplacés par 00 et le chiffre précédent est augmenté de 1.
- b) Les totaux partiels marginaux et les totaux marginaux des tableaux statistiques doivent être calculés à partir de leurs composantes non arrondies correspondantes, puis ensuite être arrondis à leur tour à la centaine près à l'aide de la technique d'arrondissement normale.
- c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir de composantes non arrondies (c'est-à-dire des numérateurs et/ou des dénominateurs), puis être arrondis à leur tour à une décimale à l'aide de la technique d'arrondissement normale. Dans le cas d'un arrondissement normal à un seul chiffre, si le dernier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, le dernier chiffre à conserver est augmenté de 1.
- d) Les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs composantes non arrondies correspondantes, puis être arrondies à leur tour à la centaine près (ou à la décimale près) à l'aide de la technique d'arrondissement normale.
- e) Dans les cas, où, en raison de limitations d'ordre technique ou de toutes autres limites, une technique d'arrondissement autre que la technique normale est utilisée produisant des estimations à être publiées ou autrement diffusées différentes des estimations correspondantes publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces différences dans le ou les documents à publier ou à diffuser.
- f) En aucun cas, les utilisateurs ne doivent publier ou autrement diffuser des estimations non arrondies. Des estimations non arrondies laissent entendre qu'elles sont plus précises qu'elles ne le sont en réalité.

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

Le plan d'échantillonnage utilisé pour l'Enquête sur les travailleurs âgés (ETA) n'était pas autopondéré. Lorsqu'ils produisent des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent appliquer le poids d'enquête 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 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'ETA, 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'ETA.

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 actives et qui sont travailleurs autonomes durant la semaine de référence ou la proportion de personnes qui sont à la retraite constituent des exemples de telles estimations. Une estimation du nombre de personnes possédant une certaine caractéristique peut aussi être désignée comme une estimation d'un agrégat.

Exemples de questions catégoriques :

Q : Est-ce que vous recevez présentement des prestations de retraite du Regime de Pension Canada / de la Regime de Pension Québec?

R : Oui / Non

Q : Qu'est ce qui a contribué à votre décision de retourner sur le marché du travail après votre retraite?

R : Besoin financier / Raison familiale / Toujours prévu retourner pour faire quelque chose de différent / Interaction sociale, quelque chose à faire / Autre - Précisez

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 mois sans travail à la suite d'une perte d'emploi. Le numérateur est une estimation du nombre total de mois sans travail et son dénominateur est le nombre de personnes travailleurs qui ont été déplacés mais qui ont retrouvé un emploi depuis leur déplacement.

Exemples de questions quantitatives :

Q : Combien d'années de service donnant droit à une pension avez-vous accumulées auprès de votre employeur actuel?

R : |_|_| années

Q : Depuis combien de temps êtes-vous sans emploi?

R : |_|_|_| mois

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. Pour obtenir, par exemple, une estimation du nombre total de mois sans travail suivant un déplacement pour les personnes vivant en Ontario, multipliez la valeur déclarée à la question DW_Q04 (le nombre de mois sans emploi (000 to 168)) pour toutes les personnes répondant à cette question par le poids final de l'enregistrement, puis additionnez cette valeur pour tous les enregistrements où la variable PROV = 35 (Ontario) et CS_Q04 = 1 (perdu un emploi suite à une mise à pied, la fermeture d'une usine ou d'une entreprise ou en raison d'une réduction du personnel).

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 mois sans travail suivant un déplacement pour les personnes vivant en Ontario,

- estimez le nombre total de mois sans emploi (\hat{X}) tel qu'il est décrit ci-dessus,
- estimez le nombre de personnes qui travaillent présentement en Ontario qui ont été déplacées (\hat{Y}) incluses dans cette catégorie en additionnant les poids finals

- de tous les enregistrements où la variable PROV = 35 et CS_Q04 = 1, puis
- c) divisez l'estimation a) par l'estimation b) (\hat{X} / \hat{Y}) .

9.4 Lignes directrices pour l'analyse statistique

L'ETA 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 SEXE = 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 SEXE = homme;
- 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. Un excellent moyen d'obtenir une approximation de la variance réelle consiste à recourir aux méthodes de rééchantillonnage, plus particulièrement la méthode bootstrap. Cette méthode, qui se fonde sur une technique de réplication de l'échantillon, produit de bonnes approximations de la valeur réelle de la variance. Un fichier contenant 1 000 poids bootstrap est disponible. Le

calcul de la variance au moyen de 1 000 poids bootstrap comprend le calcul de l'estimation en appliquant chacun de ces 1 000 poids, puis le calcul de la variance de ces 1 000 estimations.

Statistique Canada a développé un produit statistique appelé BOOTVAR qui estime les variances par la méthode bootstrap. BOOTVAR permet d'estimer les variances pour les totaux, les rapports (y compris les proportions), les différences entre rapports (ou proportions), les modèles de régression linéaire et les modèles de régression logistique. La version 3.1 de BOOTVAR (disponible en SAS seulement) permet également d'estimer les centiles et les tests d'indépendance du khi-deux. BOOTVAR est disponible dans les langages de programmation SAS et SPSS. La version la plus récente du programme (SAS version 3.1 et SPSS version 3.0), la documentation à l'intention de l'utilisateur et les renseignements sur des paramètres d'enquête particuliers peuvent être téléchargés gratuitement en suivant les liens ci-après :

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

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

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'ETA, 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 sur les travailleurs âgés

Le tableau ci-dessous fournit une indication de la précision des estimations 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 5 000 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	CV acceptable 0,0 à 16,5 %	CV médiocre 16,6 à 33,3 %	CV inacceptable > 33,3 %
Terre-Neuve-et-Labrador	9 800 et plus	2 600 à < 9 800	moins de 2 600
Île-du-Prince-Édouard	4 000 et plus	1 100 à < 4 000	moins de 1 100
Nouvelle-Écosse	12 400 et plus	3 200 à < 12 400	moins de 3 200
Nouveau-Brunswick	12 000 et plus	3 200 à < 12 000	moins de 3 200
Québec	46 800 et plus	11 800 à < 46 800	moins de 11 800
Ontario	54 300 et plus	13 600 à < 54 300	moins de 13 600
Manitoba	13 300 et plus	3 400 à < 13 300	moins de 3 400
Saskatchewan	11 500 et plus	3 000 à < 11 500	moins de 3 000
Alberta	42 800 et plus	11 100 à < 42 800	moins de 11 100
Colombie-Britannique	42 300 et plus	10 800 à < 42 300	moins de 10 800
Canada	44 600 et plus	11 000 à < 44 600	moins de 11 000

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 à utiliser à l'intérieur des tables de CV qui s'appliqueraient ensuite à l'ensemble entier des caractéristiques.

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

Provinces	Effet du plan	Taille de l'échantillon	Population
Terre-Neuve-et-Labrador	1,39	380	83 283
Île-du-Prince-Édouard	1,45	274	25 089
Nouvelle-Écosse	1,31	562	157 387
Nouveau-Brunswick	1,42	490	123 909
Québec	2,00	1 827	1 212 979
Ontario	2,27	2 947	1 977 841
Manitoba	1,58	776	191 659
Saskatchewan	1,54	776	170 076
Alberta	2,04	915	565 817
Colombie-Britannique	1,77	1 087	750 332
Canada	2,34	10 034	5 258 373

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 que le pourcentage repose sur un sous-groupe de la population. La proportion, par exemple, des personnes retraitées qui bénéficient du Régime de pension du Canada/Régime de pension du Québec est plus fiable que le nombre estimé de personnes retraitées qui bénéficient du Régime de pension du Canada/Régime de pension du Québec. (Remarquez que dans les tables la valeur des coefficients de variation diminue lorsqu'on les lit de gauche à droite.)

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

Lorsque la proportion ou que le pourcentage repose sur un sous-ensemble de la population totale (p. ex., comme ses membres d'un sexe ou d'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 qui travaillent et le numérateur, le nombre de personnes qui travaillent mais qui avaient déjà été à la retraite

Dans le cas où le numérateur n'est pas un sous-ensemble du dénominateur, comme dans l'exemple du rapport du nombre de personnes qui travaillent dans le secteur publique comparativement au nombre de personnes qui travaillent dans le secteur privé, 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 maître de l'ETA de 2008 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 383 042 les individus dans la population de l'ETA appartiennent au groupe d'âge de 65 à 69 ans. 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 (383 042) 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 400 000.
- 3) On trouve le coefficient de variation pour un agrégat estimé en se reportant à la

première entrée autre que des astérisques sur cette rangée, c'est-à-dire 5,2 %.

- 4) Le coefficient de variation approximatif de l'estimation est donc 5,2 %. Le résultat selon lequel il y avait 383 042 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1) personnes âgées de 65 à 69 ans peut être publié sans réserve *).

Enquête sur les travailleurs âgés, 2008

Tables de variabilité d'échantillonnage approximative - Canada

NUMÉRATEUR DU POURCENTAGE (en milliers)	POURCENTAGE ESTIMÉ													
	0,1%	1,0%	2,0%	5,0%	10,0%	15,0%	20,0%	25,0%	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%
1	110,6	110,1	109,5	107,8	105,0	102,0	99,0	95,8	92,6	89,2	85,7	78,2	60,6	35,0
2	78,2	77,8	77,4	76,2	74,2	72,1	70,0	67,7	65,5	63,1	60,6	55,3	42,8	24,7
3	63,8	63,6	63,2	62,3	60,6	58,9	57,1	55,3	53,4	51,5	49,5	45,2	35,0	20,2
4	55,3	55,0	54,8	53,9	52,5	51,0	49,5	47,9	46,3	44,6	42,8	39,1	30,3	17,5
5	49,5	49,2	49,0	48,2	46,9	45,6	44,3	42,8	41,4	39,9	38,3	35,0	27,1	15,6
6	*****	44,9	44,7	44,0	42,8	41,6	40,4	39,1	37,8	36,4	35,0	31,9	24,7	14,3
7	*****	41,6	41,4	40,8	39,7	38,6	37,4	36,2	35,0	33,7	32,4	29,6	22,9	13,2
8	*****	38,9	38,7	38,1	37,1	36,1	35,0	33,9	32,7	31,5	30,3	27,7	21,4	12,4
9	*****	36,7	36,5	35,9	35,0	34,0	33,0	31,9	30,9	29,7	28,6	26,1	20,2	11,7
10	*****	34,8	34,6	34,1	33,2	32,3	31,3	30,3	29,3	28,2	27,1	24,7	19,2	11,1
...
...
...
250	*****	*****	*****	6,8	6,6	6,5	6,3	6,1	5,9	5,6	5,4	4,9	3,8	2,2
300	*****	*****	*****	*****	6,1	5,9	5,7	5,5	5,3	5,1	4,9	4,5	3,5	2,0
350	*****	*****	*****	*****	5,6	5,5	5,3	5,1	4,9	4,8	4,6	4,2	3,2	1,9
400	*****	*****	*****	*****	5,2	5,1	4,9	4,8	4,6	4,5	4,3	3,9	3,0	1,7
450	*****	*****	*****	*****	4,9	4,8	4,7	4,5	4,4	4,2	4,0	3,7	2,9	1,6
500	*****	*****	*****	*****	4,7	4,6	4,4	4,3	4,1	4,0	3,8	3,5	2,7	1,6
750	*****	*****	*****	*****	*****	3,7	3,6	3,5	3,4	3,3	3,1	2,9	2,2	1,3
1 000	*****	*****	*****	*****	*****	*****	3,1	3,0	2,9	2,8	2,7	2,5	1,9	1,1
1 500	*****	*****	*****	*****	*****	*****	*****	*****	2,4	2,3	2,2	2,0	1,6	0,9
2 000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1,9	1,7	1,4	0,8
3 000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1,1	0,6
4 000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	0,6

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

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

Supposons qu'un utilisateur estime à $507\ 770 / 617\ 067 = 82,3\ %$ des personnes âgées de 55 à 59 ans en Ontario travaillent et n'ont jamais été à la retraite. 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 ONTARIO.
- 2) Parce que l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c'est-à-dire les personnes âgées de 55 à 59 ans en Ontario), il faut utiliser à la fois le pourcentage (82,3 %) et la portion numérateur du pourcentage (507 770) pour déterminer le coefficient de variation.
- 3) Le numérateur, 507 770, 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 500 000. 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 90,0 %.
- 4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 1,7 %,

est le coefficient de variation à employer.

- 5) Le coefficient de variation approximatif de l'estimation est donc 1,7 %. Le résultat selon lequel 82,3 % des personnes âgées de 55 à 59 ans en Ontario travaillent et n'ont jamais été à la retraite, peut être publié sans réserve.

Enquête sur les travailleurs âgés, 2008

Tables de variabilité d'échantillonnage approximative - Ontario

NUMÉRATEUR DU POURCENTAGE (en milliers)0,1%	POURCENTAGE ESTIMÉ													
	1,0%	2,0%	5,0%	10,0%	15,0%	20,0%	25,0%	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%	
1	123,3	122,7	122,1	120,2	117,0	113,7	110,3	106,8	103,2	99,4	95,5	87,2	67,6	39,0
2	****	86,8	86,3	85,0	82,7	80,4	78,0	75,5	73,0	70,3	67,6	61,7	47,8	27,6
3	****	70,9	70,5	69,4	67,6	65,7	63,7	61,7	59,6	57,4	55,2	50,4	39,0	22,5
4	****	61,4	61,0	60,1	58,5	56,9	55,2	53,4	51,6	49,7	47,8	43,6	33,8	19,5
5	****	54,9	54,6	53,8	52,3	50,9	49,3	47,8	46,1	44,5	42,7	39,0	30,2	17,4
6	****	50,1	49,8	49,1	47,8	46,4	45,0	43,6	42,1	40,6	39,0	35,6	27,6	15,9
7	****	46,4	46,1	45,4	44,2	43,0	41,7	40,4	39,0	37,6	36,1	33,0	25,5	14,7
8	****	43,4	43,2	42,5	41,4	40,2	39,0	37,8	36,5	35,2	33,8	30,8	23,9	13,8
9	****	40,9	40,7	40,1	39,0	37,9	36,8	35,6	34,4	33,1	31,8	29,1	22,5	13,0
10	****	38,8	38,6	38,0	37,0	36,0	34,9	33,8	32,6	31,4	30,2	27,6	21,4	12,3
11	****	37,0	36,8	36,2	35,3	34,3	33,3	32,2	31,1	30,0	28,8	26,3	20,4	11,8
12	****	35,4	35,2	34,7	33,8	32,8	31,8	30,8	29,8	28,7	27,6	25,2	19,5	11,3
13	****	34,0	33,9	33,3	32,5	31,5	30,6	29,6	28,6	27,6	26,5	24,2	18,7	10,8
...
125	****	****	****	****	10,5	10,2	9,9	9,6	9,2	8,9	8,5	7,8	6,0	3,5
150	****	****	****	****	9,6	9,3	9,0	8,7	8,4	8,1	7,8	7,1	5,5	3,2
200	****	****	****	****	8,0	7,8	7,6	7,3	7,0	6,8	6,6	6,2	4,8	2,8
250	****	****	****	****	7,2	7,0	6,8	6,5	6,3	6,0	6,0	5,5	4,3	2,5
300	****	****	****	****	****	****	6,4	6,2	6,0	5,7	5,5	5,0	3,9	2,3
350	****	****	****	****	****	****	5,9	5,7	5,5	5,3	5,1	4,7	3,6	2,1
400	****	****	****	****	****	****	****	5,3	5,2	5,0	4,8	4,4	3,4	2,0
450	****	****	****	****	****	****	****	5,0	4,9	4,7	4,5	4,1	3,2	1,8
500	****	****	****	****	****	****	****	****	4,6	4,4	4,3	3,9	3,0	1,7
750	****	****	****	****	****	****	****	****	****	****	3,5	3,2	2,5	1,4
1 000	****	****	****	****	****	****	****	****	****	****	****	****	2,1	1,2
1 500	****	****	****	****	****	****	****	****	****	****	****	****	****	1,0

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

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

Supposons qu'un utilisateur estime à 236 869 / 594 791 = 39,8 % des ménages de deux personnes au Québec ayant un revenu du ménage de 50 000 \$ to 100 000 \$, bien que 319 069 / 788 592 = 40,5 % des ménages de deux personnes en Ontario avaient des revenus dans la même intervalle. 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 ONTARIO de la même façon que celle décrite dans l'exemple 2, donne un CV de l'estimation pour Québec de 5,6 % et un CV de l'estimation pour Ontario de 5,5 %.
- 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 $\hat{d} = 0,398 - 0,405 = -0,007$ est :

$$\begin{aligned} \sigma_{\hat{d}} &= \sqrt{[(0,398)(0,056)]^2 + [(0,405)(0,055)]^2} \\ &= \sqrt{(0,0004967) + (0,0004961)} \\ &= 0,032 \end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,032 / 0,007 = 4,571$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 457,1 %. La différence entre les estimations est considérée inacceptable et Statistique Canada recommande de ne pas publier cette estimation. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre F (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les utilisateurs subséquents en garde contre les hauts taux d'erreur associés à l'estimation.

Enquête sur les travailleurs âgés, 2008

Tables de variabilité d'échantillonnage approximative - Québec

NUMÉRATEUR DU POURCENTAGE (en milliers)	POURCENTAGE ESTIMÉ													
	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	115,1	114,6	114,0	112,2	109,2	106,2	103,0	99,7	96,3	92,8	89,2	81,4	63,1	36,4
2	****	81,0	80,6	79,4	77,2	75,1	72,8	70,5	68,1	65,6	63,1	57,6	44,6	25,7
3	****	66,1	65,8	64,8	63,1	61,3	59,5	57,6	55,6	53,6	51,5	47,0	36,4	21,0
4	****	57,3	57,0	56,1	54,6	53,1	51,5	49,9	48,2	46,4	44,6	40,7	31,5	18,2
5	****	51,2	51,0	50,2	48,9	47,5	46,1	44,6	43,1	41,5	39,9	36,4	28,2	16,3
6	****	46,8	46,5	45,8	44,6	43,3	42,0	40,7	39,3	37,9	36,4	33,2	25,7	14,9
7	****	43,3	43,1	42,4	41,3	40,1	38,9	37,7	36,4	35,1	33,7	30,8	23,8	13,8
8	****	40,5	40,3	39,7	38,6	37,5	36,4	35,3	34,1	32,8	31,5	28,8	22,3	12,9
9	****	38,2	38,0	37,4	36,4	35,4	34,3	33,2	32,1	30,9	29,7	27,1	21,0	12,1
10	****	36,2	36,0	35,5	34,5	33,6	32,6	31,5	30,5	29,4	28,2	25,7	19,9	11,5
11	****	34,5	34,4	33,8	32,9	32,0	31,1	30,1	29,0	28,0	26,9	24,5	19,0	11,0
12	****	33,1	32,9	32,4	31,5	30,6	29,7	28,8	27,8	26,8	25,7	23,5	18,2	10,5
13	****	****	31,6	31,1	30,3	29,4	28,6	27,7	26,7	25,7	24,7	22,6	17,5	10,1
...
...
...
125	****	****	****	****	****	9,5	9,2	8,9	8,6	8,3	8,0	7,3	5,6	3,3
150	****	****	****	****	****	8,7	8,4	8,1	7,9	7,6	7,3	6,6	5,1	3,0
200	****	****	****	****	****	****	7,3	7,1	6,8	6,6	6,3	5,8	4,5	2,6
250	****	****	****	****	****	****	****	6,3	6,1	5,9	5,6	5,1	4,0	2,3
300	****	****	****	****	****	****	****	5,8	5,6	5,4	5,1	4,7	3,6	2,1
350	****	****	****	****	****	****	****	****	5,1	5,0	4,8	4,4	3,4	1,9
400	****	****	****	****	****	****	****	****	****	4,6	4,5	4,1	3,2	1,8
450	****	****	****	****	****	****	****	****	****	****	4,2	3,8	3,0	1,7
500	****	****	****	****	****	****	****	****	****	****	****	3,6	2,8	1,6
750	****	****	****	****	****	****	****	****	****	****	****	****	2,3	1,3
1,000	****	****	****	****	****	****	****	****	****	****	****	****	****	1,2

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

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime à 297 735 personnes dans la population de l'ETA au Québec qui travaillent dans le secteur publique, par rapport à 427 088 personnes dans la

population de l'ETA en Ontario qui travaillent dans le secteur public. L'utilisateur est intéressé à comparer l'estimation des Québec à celle des Ontario sous la forme d'un rapport. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation d'un rapport, où le numérateur de l'estimation (\hat{X}_1) est le nombre de personnes au Québec employées dans le secteur public. Le dénominateur de l'estimation (\hat{X}_2) est le nombre de personnes en Ontario employées dans le secteur public.
- 2) Reportez-vous à la table des coefficients de variation pour le QUÉBEC et ONTARIO.
- 3) Le numérateur de cette estimation de rapport est 297 735 - l'estimation pour le Québec. Le chiffre qui s'en rapproche le plus est 300 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 5,8 %.
- 4) Le dénominateur de cette estimation de rapport est 427 088 - l'estimation pour l'Ontario. Le chiffre qui s'en rapproche le plus est 450 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 5,0 %.
- 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 :

$$\begin{aligned} \alpha_{\hat{R}} &= \sqrt{(0,058)^2 + (0,050)^2} \\ &= \sqrt{0,003364 + 0,0025} \\ &= 0,077 \end{aligned}$$

- 6) Le rapport obtenu entre les personnes du Québec et de l'Ontario qui travaillent dans le secteur public 297 735 / 427 088, c'est-à-dire 0,697 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1). Le coefficient de variation de cette estimation est 7,7 %, ce qui fait qu'on peut la diffuser sans réserve.

Exemple 5 : Estimations de différences de rapports

Supposons qu'un utilisateur estime que le rapport entre les personnes sans le Régime enregistré d'épargne-retraite (REER), et les individus avec REERs is 0,423 pour le Québec, tandis qu'il est de 0,319 pour l'Ontario. L'utilisateur est intéressé à comparer les deux rapports pour voir s'il y a une différence statistique entre ceux-ci. Comment l'utilisateur détermine-t-il le coefficient de variation de cette différence?

- 1) Tout d'abord calculez le coefficient de variation approximatif pour le rapport du Québec (\hat{R}_1) et le rapport de l'Ontario (\hat{R}_2) tel qu'il est décrit dans l'exemple 4. Le

CV approximatif pour le rapport du Québec est 5,6 % et 5,1 % pour celui de l'Ontario.

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

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

où α_1 et α_2 sont les coefficients de variation de \hat{R}_1 et \hat{R}_2 respectivement. C'est-à-dire que l'erreur-type de la différence $\hat{d} = 0,423 - 0,319 = 0,104$ est :

$$\begin{aligned} \sigma_{\hat{d}} &= \sqrt{[(0,423)(0,056)]^2 + [(0,319)(0,051)]^2} \\ &= \sqrt{(0,0005611) + (0,0002646)} \\ &= 0,029 \end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,029 / 0,104 = 0,279$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 27,9 %. La différence entre les estimations est considérée médiocre et Statistique Canada recommande de ne pas la publier. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre E (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les utilisateurs subséquents en garde contre les hauts taux d'erreur associés à l'estimation.

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

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

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

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

Des intervalles de confiance pour une estimation \hat{X} sont généralement exprimés sous

forme de deux chiffres, un inférieur et un supérieur à l'estimation, comme étant $(\hat{X} - k, \hat{X} + k)$, 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 personnes âgées de 55 à 59 ans en Ontario qui travaillent et n'ont jamais été retraitées (d'après l'exemple 2 à la section 10.1.1) serait calculé comme suit :

$$\hat{X} = 82,3 \% \text{ (ou exprimé sous forme de proportion } 0,823)$$

$$t = 2$$

$$\alpha_{\hat{x}} = 1,7 \% \text{ (0,017 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.}$$

$$IC_{\hat{x}} = \{0,823 - (2) (0,823) (0,017), 0,823 + (2) (0,823) (0,017)\}$$

$$IC_{\hat{x}} = \{0,823 - 0,028, 0,823 + 0,028\}$$

$$IC_{\hat{x}} = \{0,795, 0,851\}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 79,5 % et 85,1 % personnes âgées de 55 à 59 ans en Ontario travaillent et n'ont jamais été retraitées.

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, elles sont identiques.

Supposons que \hat{X}_1 et \hat{X}_2 sont des estimations d'un échantillon pour deux caractéristiques qui nous intéressent. Supposons également que l'erreur-type de la différence $\hat{X}_1 - \hat{X}_2$ est $\sigma_{\hat{d}}$.

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

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

Supposons que l'utilisateur désire tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y a pas de différence entre la proportion de ménages à deux personnes au Québec avec un revenu de ménage de 50 000 \$ to 100 000 \$ et la proportion de ménages à deux personnes en Ontario avec un revenu de ménage de 50 000 \$ to 100 000 \$. 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 0,032. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0,398 - 0,405}{0,032} = \frac{-0,007}{0,032} = -0,219$$

Puisque $t = -0,219$ est compris entre -2 et +2, donc aucune conclusion peut être faite au niveau de signification de 5 %.

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'ETA 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 mois de travail suivant un déplacement de travail serait supérieur au coefficient de variation de la proportion correspondante de travailleurs ayant vécu un déplacement de travail. 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 ETA2008_CVTabF.pdf pour les tables de coefficients de variation.

11.0 Pondération

Puisque l'Enquête sur les travailleurs âgés (ETA) 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 une croissance dans plus d'une grappe incluse à l'intérieur de la tâche d'un intervieweur, l'effet cumulatif de toutes les augmentations peut créer un problème sur le plan de la charge de travail. Dans le cas des grappes dont la croissance est importante, on utilise un sous-échantillon pour que les tâches des intervieweurs demeurent gérables. Le sous-poids d'une grappe représente l'inverse de ce rapport de sous-échantillonnage dans le cas des grappes pour lesquelles il y a eu sous-échantillonnage.

Poids de stabilisation

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

Non-réponse

Dans le cas de certains types de non-réponses (comme les ménages temporairement absents ou les refus), les données de l'interview menée le mois précédent auprès d'un ménage, le cas

échéant, sont réutilisées pour ce ménage comme données du mois courant.

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

Sous-poids de l'Enquête sur la population active

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

Ajustements au niveau infraprovincial et au rapport province-âge-sexe

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

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

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

11.2 Procédures de pondération pour l'Enquête sur les travailleurs âgés

Les principes qui sous-tendent le calcul des poids pour l'ETA 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'ETA.

1. Ajustement pour l'interview sur place assistée par ordinateur (IPAO) et la non-réponse à l'Enquête sur la population active

On a attribué un poids d'enquête à toutes les unités d'échantillonnage de l'EPA figurant dans le fichier principal de l'EPA (fichier TABS). Toutefois, les échantillons figurant sur le fichier TABS

n'étaient pas tous des répondants pendant le mois où les données ont été recueillies. Dans certains cas, lorsqu'une réponse ne pouvait être obtenue (quelle que soit la raison), des données étaient saisies pour le mois, souvent en reportant les données d'un mois précédent. Étant donné qu'il s'agissait en fait de cas de non-réponse à l'EPA et puisque l'Enquête sur les travailleurs âgés (ETA) devait être sélectionnée auprès des répondants à l'EPA, un ajustement distinct pour la non-réponse a été effectué afin de prendre en compte ces cas avant la pondération de l'ETA. Les groupes d'ajustement pour la non-réponse à l'EPA ont été utilisés pour cette étape de repondération.

De plus, les répondants de l'EPA pour lesquels des données ont été recueillies au moyen d'une interview sur place ont été exclus de l'échantillon de l'ETA. Ces unités d'échantillonnage ont été regroupées avec les non-répondants de l'EPA et incluses dans l'ajustement pour la non-réponse susmentionné.

2. Ajustement pour les groupes de renouvellement

L'échantillon de l'EPA comprend six groupes de renouvellement tandis que l'échantillon de l'ETA n'en utilise que cinq. Un ajustement de 1,2 est donc appliqué au poids de l'ETA afin de compenser le groupe de renouvellement manquant.

3. Ajustement pour le sous-échantillonnage

Lors de la sélection de l'échantillon pour l'ETA, on a communiqué avec tous les ménages de l'EPA dans les cinq groupes de renouvellement de l'ETA comprenant au moins une personne dans la population cible de l'ETA. Si plus d'une personne au sein du ménage appartenait à la population cible, alors une des personnes était sélectionnée au hasard pour l'échantillon de l'ETA. Afin de justifier le sous-échantillonnage, le poids de l'ETA est multiplié par le nombre de membres du ménage appartenant à la population cible.

4. Ajustement pour la non-réponse

Afin de procéder à un ajustement pour la non-réponse à l'ETA, la régression logistique a été utilisée pour estimer la probabilité prévue de réponse pour chaque unité d'échantillonnage. La modélisation a été effectuée au sein de la province. Afin de former des groupes de réponse pour l'ajustement, le fichier de l'échantillon a été classé par probabilité de réponse puis divisé en déciles donnant dix groupes d'ajustement de réponse pour chacune des provinces. Au sein des groupes d'ajustement, le facteur de correction du poids a été calculé comme étant le rapport entre la somme des poids pour toutes les unités d'échantillonnage et la somme des poids pour l'ensemble des répondants.

5. Ajustement pour le calage des chiffres de population prévus

Un calage des poids ajustés pour la non-réponse est alors effectué. Les poids de l'ETA ajustés pour la non-réponse pour les répondants de l'ETA sont combinés aux sous-poids de l'EPA pour les répondants de l'EPA âgés de 50 à 75 ans et ne faisant pas partie de la population cible de l'ETA. Un calage est alors effectué en ajustant les poids afin qu'ils équivalent aux estimations démographiques connues pour la tranche d'âge de 50 à 75 ans. Le calage devait être effectué en incluant l'échantillon de l'EPA non lié à l'ETA, étant donné que les estimations démographiques étaient uniquement disponibles pour l'ensemble de la population âgée de 50 à 75 ans et non pour la population cible de l'ETA.

Le poids qui en résulte WTPM (pour « final weight » en anglais) est le poids final qui figure dans le fichier maître de microdonnées de l'ETA et WTPP est le poids final qui figure dans le fichiers de microdonnées à grande diffusion (FMGD).

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 sur les travailleurs âgés

Le questionnaire de l'Enquête sur les travailleurs âgés (ETA) a servi à recueillir des données pour l'enquête supplémentaire en octobre et novembre 2008. Le fichier ETA2008_QuestF.pdf renferme le questionnaire français.

13.0 Cliché d'enregistrement à valeurs univariées

Consulter le fichier ETA2008_LvCds.pdf pour le cliché d'enregistrement à valeurs univariées.