



Guide de l'utilisateur des microdonnées

Enquête sur la couverture de l'assurance-emploi 2004 et 2005



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 la couverture de l'assurance-emploi	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	18
5.2.3	Stratification secondaire	18
5.2.4	Délimitation et sélection des grappes	19
5.2.5	Sélection des logements	19
5.2.6	Sélection des personnes	20
5.3	Taille de l'échantillon	20
5.4	Renouvellement de l'échantillon	20
5.5	Modifications apportées au plan de l'Enquête sur la population active pour l'Enquête sur la couverture de l'assurance-emploi	20
5.5.1	Population cible	21
5.5.2	Type 4 : Un cas particulier	21
5.5.3	Sous-échantillonnage	22
5.5.4	Autres exclusions	22
5.6	Taille de l'échantillon selon la province pour l'Enquête sur la couverture de l'assurance-emploi	23
6.0	Collecte des données	25
6.1	Réalisation des interviews dans le cadre de l'Enquête sur la population active	25
6.2	Supervision et contrôle de qualité	25
6.3	Non-réponse à l'Enquête sur la population active	26
6.4	Modifications apportées à la collecte des données de l'Enquête sur la couverture de l'assurance-emploi	26
6.5	Non-réponse à l'Enquête sur la couverture de l'assurance-emploi	26
7.0	Traitement des données	27
7.1	Saisie des données	27
7.2	Vérification et contrôle	27
7.3	Codage des questions ouvertes	28
7.4	Imputation	28
7.5	Création de variables dérivées	29
7.5.1	Regroupement des éléments de données continus	29
7.5.2	Regroupement de questions identiques	29
7.5.3	Regroupement de données de l'Enquête sur la population active et de	

	l'Enquête sur la couverture de l'assurance-emploi	30
7.5.4	Regroupement de deux questions différentes ou plus	30
7.5.5	Classification de la protection par le régime d'assurance-emploi : la variable COV	31
7.6	Pondération.....	33
7.7	Suppression de renseignements confidentiels	34
8.0	Qualité des données	35
8.1	Taux de réponse	35
8.2	Erreurs relatives à l'enquête	35
8.2.1	Base de sondage	36
8.2.2	Collecte des données	36
8.2.3	Traitement des données	36
8.2.4	Non-réponse	37
8.2.5	Mesure de l'erreur d'échantillonnage.....	38
9.0	Lignes directrices pour la totalisation, l'analyse et la diffusion de données	39
9.1	Lignes directrices pour l'arrondissement d'estimations	39
9.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation	40
9.3	Définitions de types d'estimations : catégoriques et quantitatives	40
9.3.1	Estimations catégoriques.....	40
9.3.2	Estimations quantitatives	41
9.3.3	Totalisation d'estimations catégoriques.....	41
9.3.4	Totalisation d'estimations quantitatives	41
9.4	Lignes directrices pour l'analyse statistique	42
9.5	Lignes directrices pour la diffusion de coefficients de variation.....	43
9.6	Seuils pour la diffusion des estimations pour l'Enquête sur la couverture de l'assurance-emploi	45
10.0	Tables de variabilité d'échantillonnage approximative	47
10.1	Comment utiliser les tables de coefficients de variation pour des estimations catégoriques.....	48
10.1.1	Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques	50
10.2	Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance	56
10.2.1	Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance	57
10.3	Comment utiliser les tables de coefficients de variation pour effectuer un test t.....	57
10.3.1	Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t	58
10.4	Coefficients de variation pour des estimations quantitatives.....	58
10.5	Tables des coefficients de variation.....	59
11.0	Pondération	61
11.1	Procédures de pondération pour l'Enquête sur la population active	61
11.2	Procédures de pondération pour l'Enquête sur la couverture de l'assurance-emploi	63
12.0	Questionnaires	65
12.1	Le questionnaire de l'Enquête sur la population active	65

12.2	Les questionnaires de l'Enquête sur la couverture de l'assurance-emploi.....	65
13.0	Cliché d'enregistrement à valeurs univariées	67

1.0 Introduction

L'Enquête sur la couverture de l'assurance-emploi (ECAE) est menée par Statistique Canada avec la collaboration et l'appui de Ressources humaines et Développement social Canada (auparavant 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

2.0 Contexte

Lancée en 1997, l'Enquête sur la couverture de l'assurance-emploi (ECAE) avait comme principal objectif de nous permettre de mieux comprendre le rapport entre le nombre de personnes touchant des prestations d'assurance-emploi ou admissibles à ces prestations et le nombre de chômeurs selon l'Enquête sur la population active (EPA).

Les données administratives de l'assurance-emploi sont limitées quant à la population couverte et les variables accessibles : on dispose de renseignements sur les demandes acceptées mais on ne dispose d'aucun renseignement sur les demandes rejetées ou sur les non-demandeurs. Dans les données administratives, il manque aussi des renseignements démographiques et sur les ménages, lesquels sont nécessaires pour l'analyse sociale.

Afin de brosser un tableau plus complet de la protection offerte par le régime d'assurance-emploi, l'enquête cible également les autres personnes qui peuvent avoir été admissibles aux prestations d'assurance-emploi en raison d'une récente interruption d'emploi ou d'une situation de sous-emploi.

La portée de l'enquête a été élargie en 2000 de façon à couvrir l'accès aux prestations de maternité et aux prestations parentales. Ces changements ont été apportés un an avant l'élargissement du programme de prestations parentales, en janvier 2001.

3.0 Objectifs

L'enquête sur la couverture de l'assurance-emploi (ECAE) vise principalement à évaluer le rendement du régime d'assurance-emploi, en déterminant le nombre de personnes protégées par le régime et la proportion de gens recevant des prestations de même qu'en cernant les groupes n'ayant pas accès à l'assurance-emploi.

Les données servent à mesurer la couverture de la population canadienne par l'assurance-emploi et le rôle que jouent les prestations d'assurance-emploi dans la contribution au revenu personnel et au revenu du ménage en période de chômage ou de sous-emploi. Les chômeurs, les personnes qui travaillent (p. ex. les prestataires qui ont un revenu) et les personnes classées inactives par l'EPA font l'objet d'une analyse sur le sujet. Un nombre important de personnes des deux derniers groupes reçoivent aussi des prestations d'assurance-emploi.

Les facteurs cités le plus souvent pour expliquer les écarts de couverture d'assurance-emploi sont les suivants : inadmissibilité à l'assurance-emploi, prestations épuisées, en période d'attente après une cessation d'emploi, ou aucune demande de prestation d'assurance-emploi. L'ampleur de ces facteurs et d'autres facteurs, ainsi que leurs liens avec les caractéristiques personnelles, les cycles saisonniers et économiques, et les régions du Canada peuvent être examinés à l'aide de la présente enquête pour améliorer notre compréhension des raisons pour lesquelles certains chômeurs ne reçoivent pas de prestations d'assurance-emploi.

Les analystes seront en mesure d'observer les caractéristiques et la situation des gens non protégés par le régime d'assurance-emploi et des personnes ayant épuisé leurs prestations, l'intensité avec laquelle les chômeurs font des démarches pour trouver un emploi, les attentes à l'égard d'un rappel au travail ainsi que le recours à des sources de revenu de remplacement.

Les données de l'enquête concernant les prestations de maternité et les prestations parentales permettent de répondre aux questions sur le nombre de mères d'un enfant en bas âge qui ont reçu ces prestations, les raisons pour lesquelles elles n'en reçoivent pas et le partage des prestations parentales avec leur conjoint. L'enquête permet aussi d'examiner le choix du moment du retour au travail et les circonstances s'y rattachant, ainsi que le caractère adéquat du revenu des ménages ayant de jeunes enfants, et plus encore.

Enquête sur la couverture de l'assurance-emploi

L'enquête est conçue pour produire une série de mesures précises sur la population sans emploi afin d'identifier les groupes qui sont le moins susceptibles de recevoir des prestations. Parmi ces groupes, on compte les suivants :

- les personnes sans emploi depuis longtemps
- les nouveaux venus sur le marché du travail et les étudiants
- les personnes qui deviennent chômeurs après un emploi non assuré
- les personnes qui ont quitté leur travail volontairement
- les personnes qui sont admissibles de par leurs antécédents d'emploi, mais qui ne demandent pas de prestations ni n'en reçoivent d'aucune façon

Couverture de l'assurance-emploi des personnes sans emploi

On a utilisé les données de l'enquête pour classer les personnes dans les catégories « potentiellement admissibles » à l'assurance-emploi ou « potentiellement non admissibles » à l'assurance-emploi, selon l'information communiquée par les répondants au sujet des prestations demandées et des prestations touchées, de leurs raisons pour ne pas recevoir ou demander de prestations et de leur expérience récente du marché du travail. On utilise l'expression « potentiellement admissibles à l'assurance-emploi »

pour décrire des personnes sans emploi qui, au cours de la semaine de référence, recevaient des prestations d'assurance-emploi ou pouvaient en recevoir en raison de leur emploi assurable récent et de leur perte de travail subséquente. L'expression « potentiellement non admissibles » décrit la situation des personnes qui n'ont pas reçu de prestations et qui n'auraient pu en recevoir même si elles en avaient fait la demande, selon l'information déclarée.

L'Enquête sur la couverture de l'assurance-emploi donne une idée de la composition de la population des personnes sans emploi, et particulièrement de celles qui n'ont pas reçu de prestations d'assurance-emploi durant la semaine de référence. L'enquête est un meilleur indicateur de qui a ou n'a pas accès à des prestations d'assurance-emploi que les indicateurs du ratio prestataires-chômeurs (P-C). Un ratio prestataires-chômeurs (P-C) est calculé pour une semaine donnée en divisant le nombre de prestataires ordinaires de l'assurance-emploi par le nombre total de chômeurs. En 2005, 55 % des personnes sans emploi étaient potentiellement admissibles à l'assurance-emploi. Sur le nombre de personnes potentiellement admissibles, 83 % pouvaient satisfaire aux exigences d'admissibilité du programme et étaient très susceptibles de recevoir des prestations au cours de leur période de chômage si elles le demandaient. Le dernier 17 % ne répondaient pas aux exigences d'admissibilité.

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 la couverture de l'assurance-emploi (ECAE) sont fournis dans la section 4.2. Les utilisateurs sont priés de se reporter au chapitre 12.0 de ce document où figure une copie du (des) questionnaire(s) d'enquête réellement employé(s).

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

Situation vis-à-vis de l'activité

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

Emploi

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

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

Chômage

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

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

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.

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.

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 la couverture de l'assurance-emploi

Type

L'échantillon de l'ECAE comprend cinq types de sous-populations distincts auxquels s'intéressent les chercheurs :

- 1) les chômeurs au cours de la semaine de référence;
- 2) les travailleurs à temps partiel au cours de la semaine de référence;
- 3) les personnes inactives au cours de la semaine de référence;
- 4) les travailleurs à temps plein au cours de la semaine de référence ayant débuté leur emploi actuel au cours des trois derniers mois;
- 5) les mères d'un enfant de moins d'un an ayant travaillé au cours de la semaine de référence.

Le type de sous-population détermine souvent les questions posées.

Mères

Dans l'ECAE, le terme « mère » renvoie à la mère (par naissance ou adoption) d'un enfant qui avait moins de un an pendant la semaine de référence de l'EPA. De nombreuses mères ne faisaient pas partie de l'échantillon avant 2000. Les mères travaillant à temps plein et les mères ne faisant pas partie de la population active et n'ayant pas travaillé dans les deux dernières années (ou n'ayant jamais travaillé) n'étaient notamment pas visées par l'enquête avant 2000.

Population « ordinaire »

Personne qui n'est pas la mère d'un enfant en bas âge durant la semaine de référence de l'enquête (voir la définition des « mères » ci-dessus).

Échantillon initial

Population ciblée par l'ECAE avant qu'elle ne soit élargie de façon à englober toutes les mères d'un enfant en bas âge.

L'enquête initiale visait les personnes appartenant au :

- type 1 (même définition qu'à l'heure actuelle),
- type 2 (y compris les mères travaillant à temps partiel),
- type 3 (à l'exception des mères qui n'ont pas travaillé en deux ans) et
- type 4 (y compris les mères qui ont connu une récente interruption de travail).

Il importe de signaler que seule la définition des personnes du type 1 (les chômeurs) n'a pas changé depuis 1997.

Semaine de référence

L'échantillon utilisé pour l'enquête est composé de personnes qui ont terminé leur participation à l'EPA. Même si les interviews sont réalisées de trois à sept semaines après celles de l'EPA, la semaine de référence est la même.

Mois de référence

Le mois de référence est le mois qui inclut la semaine de référence. C'est la période de référence pour les questions sur le revenu.

Année de référence

Pour les « mères », l'année de référence est la période de 12 mois précédant la naissance ou l'adoption de l'enfant.

Pour les membres de la population « ordinaire » visée par l'ECAE, l'année de référence est la période de 12 mois qui se termine par le mois de référence.

Travail durant la semaine de référence

Le « travail durant la semaine de référence » renvoie à tout travail d'une heure ou plus exécuté contre rémunération ou afin de réaliser un profit.

Emploi à temps plein/temps partiel

Dans le contexte de l'enquête, on entend par « employé à temps plein » les personnes qui travaillent habituellement 30 heures ou plus par semaine à un ou plusieurs emplois. L'emploi à temps partiel tient compte de toutes les autres personnes travaillant habituellement moins de 30 heures par semaine.

Dans le cadre de l'EPA, le travail à temps partiel est défini différemment pour les personnes qui ont plusieurs emplois : le critère des 30 heures s'applique uniquement à l'emploi principal.

Emploi assurable

Signifie un emploi assuré par le régime d'assurance-emploi contre un arrêt de rémunération. Le travail autonome et certains autres emplois sont exclus. L'enquête permet de déterminer s'il s'agit d'un emploi assurable si des cotisations d'assurance-emploi sont déduites de la rémunération et selon la catégorie de travail.

Prestataire de l'assurance-emploi

Un prestataire est une personne qui a présenté une demande de prestations d'assurance-emploi au cours d'une période donnée.

Bénéficiaire de l'assurance-emploi

Un bénéficiaire est une personne qui, sur présentation d'une demande de prestations, est admissible et reçoit des prestations pendant une période donnée (la semaine de référence, par exemple, ou le mois de référence ou depuis la dernière interruption de travail).

Potentiellement admissible au régime d'assurance-emploi

Expression visant les chômeurs qui, pendant la semaine de référence, ont reçu des prestations ou étaient à même d'en recevoir puisqu'ils avaient récemment occupé un emploi assurable qu'ils avaient perdu. Cela inclut toutes les personnes au chômage qui ont occupé un emploi assurable dans les 12 derniers mois et qui n'ont pas abandonné leur emploi sans motif valable ou en raison d'un retour aux études.

Admissible au régime d'assurance-emploi

Il s'agit d'un sous-ensemble de la population potentiellement admissible au régime d'assurance-emploi. Il englobe les personnes qui reçoivent ou qui comptent recevoir des prestations d'assurance-emploi pendant leur période de chômage actuelle ainsi que celles qui ont occupé un emploi rémunéré dans l'année précédant la perte ou le départ de leur dernier emploi et qui ont vraisemblablement accumulé un nombre suffisant d'heures pour être admissibles au régime d'assurance-emploi.

Potentiellement non admissible au régime d'assurance-emploi

Ce groupe comprend les chômeurs qui n'ont pas eu d'emploi assurable dans les 12 derniers mois de même que les personnes qui ont quitté leur emploi sans motif valable ou pour retourner aux études.

5.0 Méthodologie de l'enquête

Comme l'Enquête sur la couverture de l'assurance-emploi (ECAE) est menée depuis 1997 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'ECAE s'est écartée du plan de base de l'EPA.

5.1 Population visée

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

5.2 Plan de sondage

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

5.2.1 Stratification primaire

Les provinces sont divisées en régions économiques (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 social 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.

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.

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

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

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

5.2.3 Stratification secondaire

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

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

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

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

5.2.4 Délimitation et sélection des grappes

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

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

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

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

5.2.5 Sélection des logements

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

5.2.6 Sélection des personnes

On recueille des renseignements démographiques sur tous les membres du ménage pour lesquels le logement sélectionné constitue le lieu de résidence habituel. Les renseignements obtenus dans le cadre de l'EPA concernent tous les membres civils du ménage âgés de 15 ans 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 la couverture de l'assurance-emploi

L'ECAE est menée chaque année sur quatre cycles. Elle a recours pour chaque cycle au groupe de renouvellement qui vient de terminer une période de six mois dans l'EPA. La collecte de données pour l'ECAE suit celle de l'EPA pour les mois de mars, juin, octobre et décembre. Un second groupe de renouvellement est ajouté à l'échantillon à chaque cycle pour les mères d'un

enfant en bas âge.

Les estimations de l'enquête sont produites pour l'année de référence en faisant la moyenne des quatre cycles de l'enquête.

5.5.1 Population cible

La population cible de l'enquête est formée d'une sous-population de l'EPA et regroupe cinq groupes (ou types) de personnes susceptibles de toucher des prestations d'assurance-emploi :

- 1) les chômeurs au cours de la semaine de référence;
- 2) les travailleurs à temps partiel au cours de la semaine de référence;
- 3) les personnes inactives au cours de la semaine de référence;
- 4) les travailleurs à temps plein au cours de la semaine de référence ayant débuté leur emploi actuel au cours des trois derniers mois;
- 5) les mères d'un enfant de moins d'un an ayant travaillé au cours de la semaine de référence.

Le principal groupe cible est formé des chômeurs et des sans-emploi, mais les travailleurs à temps partiel peuvent également toucher des prestations (par exemple, s'ils ont récemment connu un arrêt de rémunération et s'ils sont admissibles à conserver leurs prestations d'assurance-emploi tout en travaillant en raison de la faiblesse de leurs gains provenant d'un emploi).

Un groupe de renouvellement de l'EPA est normalement formé d'environ 5 500 personnes faisant partie de l'un des cinq groupes cibles (sur un échantillon total d'environ 22 000 personnes de 15 ans et plus). Les travailleurs à temps plein ainsi que les personnes n'ayant pas travaillé pendant deux ans et ne faisant pas partie de la population active pendant la semaine de référence étaient les principaux exclus.

5.5.2 Type 4 : Un cas particulier

Les répondants appartenant au type 4 ne sont pas tous ciblés par l'enquête. Seuls ceux qui ont connu une interruption de travail dans les deux mois précédant la semaine de référence de l'enquête doivent être interviewés. Comme ce renseignement n'était pas disponible dans la base de sondage de l'EPA, tous les travailleurs à temps plein occupant leur emploi actuel depuis peu de temps ont été sélectionnés. La question sur l'interruption de travail est posée aux répondants, et aucune autre question n'est posée à ceux qui disent avoir travaillé continuellement pendant les deux mois précédant la semaine de référence. Ils sont hors du champ de l'enquête et sont retirés à l'étape du traitement (voir la section 7.2). Environ 40 % des personnes sélectionnées et appartenant au type 4 sont retirées de l'enquête pour cette raison dans une année.

5.5.3 Sous-échantillonnage

À l'étape initiale, le sous-échantillonnage a été effectué de façon à obtenir un échantillon cible de 3 600 personnes et d'équilibrer la représentation des groupes en fonction de la pertinence du régime d'assurance-emploi pour eux. Les critères de sous-échantillonnage sont résumés comme suit, selon le type.

Type 1

Toutes les personnes ont été sélectionnées.

Type 2

- a) les étudiants à temps plein ont été sous-échantillonnés à 70 %;
- b) les personnes qui ont travaillé entre 20 et 24 heures pendant la semaine de référence;
- c) les personnes qui ont travaillé entre 25 et 29 heures pendant la semaine de référence ont été sous-échantillonnées à un taux approximatif de 50 %;
- d) le reste de l'échantillon a été sous-échantillonné à 100 %.

Type 3

- a) les étudiants à temps plein qui ont quitté leur dernier emploi pour poursuivre leurs études et les étudiants à temps plein qui n'ont pas quitté leur emploi il y a plus d'une année ont été sous-échantillonnés à 50 %;
- b) le reste de l'échantillon a été sous-échantillonné à 100 %.

Type 4

Toutes les personnes ont été sélectionnées.

Type 5

Toutes les personnes ont été sélectionnées.

5.5.4 Autres exclusions

À la seconde étape du sous-échantillonnage, lorsque trois personnes ou plus ciblées par l'ECAE vivaient dans le même ménage, deux d'entre elles seulement étaient sélectionnées, sauf si elles étaient toutes au chômage. Trois personnes au maximum étaient alors sélectionnées. Cette façon de procéder permettait de réduire le fardeau de réponse dans le ménage.

Certaines personnes qui n'ont pas répondu à l'interview de l'EPA (données imputées) ou qui n'ont pas autorisé le personnel de l'EPA à mener des interviews téléphoniques auprès d'elles ont aussi été exclues de l'ECAE.

5.6 Taille de l'échantillon selon la province pour l'Enquête sur la couverture de l'assurance-emploi

Le tableau suivant montre le nombre de personnes des groupes de renouvellement de l'EPA qui ont été sélectionnées pour l'échantillon de l'ECAE.

Province	Taille de l'échantillon selon l'année d'enquête	
	2004	2005
Terre-Neuve-et-Labrador	777	632
Île-du-Prince-Édouard	459	407
Nouvelle-Écosse	914	791
Nouveau-Brunswick	857	802
Québec	2 595	2 697
Ontario	4 284	4 247
Manitoba	1 002	938
Saskatchewan	1 022	1 057
Alberta	1 530	1 458
Colombie-Britannique	1 282	1 502
Canada	14 722	14 531

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 la couverture de l'assurance-emploi

Les membres des ménages sélectionnés pour l'Enquête sur la couverture de l'assurance-emploi (ECAE) sont contactés de trois à sept semaines après la dernière interview dans le cadre de l'EPA. Toutes les interviews sont réalisées au téléphone, et les réponses par personne interposée ne sont pas permises pour l'ECAE. Il peut y avoir plus d'une personne sélectionnée dans chaque ménage, mais il n'y en a jamais plus de trois.

6.5 Non-réponse à l'Enquête sur la couverture de l'assurance-emploi

Comme c'est le cas pour l'EPA, les intervieweurs sont priés de faire tous les efforts raisonnables pour réaliser une interview dans le cadre de l'ECAE. Si un refus est essuyé à la première tentative, un intervieweur principal effectue un suivi. Toutefois, contrairement à ce qui se passe avec l'EPA, aucune lettre n'est envoyée pour obtenir la collaboration des répondants.

7.0 Traitement des données

Le principal produit de l'Enquête sur la couverture de l'assurance-emploi (ECAE) 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 et contrôle

Les fichiers textes électroniques contenant les transmissions quotidiennes des cas terminés sont regroupés pour former le fichier « brut » de l'enquête. À la fin de la période de collecte, ce fichier devrait contenir un enregistrement pour chaque personne échantillonnée. Avant de poursuivre le traitement, une vérification est effectuée pour repérer et éliminer les enregistrements en double ainsi que pour retirer les enregistrements des non-répondants et ceux qui sont hors du champ de l'enquête.

Les répondants peuvent être hors du champ de l'enquête pour un certain nombre de raisons. La très vaste majorité des enregistrements inadmissibles appartiennent au type 4 (voir la section 5.5.2). Un petit nombre d'autres enregistrements sont retirés après qu'ait été vérifiée l'exactitude de l'information utilisée dans l'échantillonnage. Enfin, un très petit pourcentage de l'échantillon est inadmissible à l'ECAE au moment de l'interview en raison d'un décès, du déménagement en établissement ou d'un déménagement à l'extérieur du pays.

Les enregistrements contenant des réponses partielles sont conservés si les répondants ont donné des réponses aux questions permettant d'établir la variable COV de la couverture par le régime d'assurance-emploi (voir la section 7.5.5).

Le contrôle consiste à modifier les données au niveau des variables individuelles. Il faut déterminer dans un premier temps quels éléments de l'enquête doivent être conservés dans le fichier principal. Les caractères invalides sont ensuite supprimés et les données élémentaires sont formatées de façon appropriée. Les zones de texte sont retirées des fichiers principaux et le texte est versé dans un fichier distinct à des fins de codage.

Les erreurs dans le déroulement du questionnaire, où l'on a relevé des questions qui ne s'appliquaient pas au répondant (et auxquelles on n'aurait donc pas dû répondre) et qui renfermaient des réponses, constituaient le premier type d'erreurs traitées. Dans ces cas, une vérification par ordinateur a éliminé automatiquement les données superflues en suivant l'ordre du questionnaire dicté par les réponses à des questions antérieures et subséquentes, parfois. Pour ce qui est des sauts fondés sur des questions auxquelles une réponse a été donnée, le code « Enchaînement valide » est attribué à toutes les questions sautées (6, 96, 996 ...). Le code « Non déclaré » est par ailleurs attribué à toutes les questions sautées parce que le répondant ne savait pas la réponse (Ne sait pas) ou a refusé d'y répondre (Refus) (9, 99, 999...). La valeur « Non déclaré » (9, 99, 999, etc., selon la longueur de la variable) est attribuée aux autres questions auxquelles aucune réponse n'a été donnée.

Il n'y a eu aucun autre type de contrôle des réponses. En conséquence, des erreurs de cohérence internes peuvent ressortir au moment de l'analyse. Il convient de souligner l'exemple de l'élément sur les gains horaires (HRLYEARN), qui comprend un petit pourcentage de valeurs aberrantes et une erreur de cohérence interne (des personnes qui travaillent et ne déclarent aucun revenu).

7.3 Codage des questions ouvertes

Quelques données élémentaires ont été consignées sur le questionnaire par les intervieweurs sous forme de questions ouvertes. Dans l'ECAE, des codes standards sont attribués aux descriptions de l'industrie et de la profession fournies par les répondants (Système de classification des industries de l'Amérique du Nord [SCIAN] et Classification type des professions [CTP de 1991] ainsi qu'au pays de naissance. De plus, les champs « Autre, précisez » contenant un nombre significatif de réponses en toutes lettres ont été examinés et codés selon les catégories existantes. De nouvelles catégories ont parfois été créées pour faciliter les analyses de l'information textuelle. Il s'agissait des éléments concernant la raison de ne pas demander ou recevoir de prestations, de l'industrie, de la profession, de la raison de l'interruption d'emploi, de la méthode de recherche d'emploi utilisée, de la raison pour laquelle le conjoint ne demande pas de prestations ou de la raison pour laquelle les deux parents le font.

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.

La technique d'imputation a été utilisée pour qu'il n'y ait plus de renseignements manquants dus aux problèmes d'applications en 2000 et 2001 ou pour en réduire le nombre. Cette technique n'a pas été utilisée les années suivantes. Les utilisateurs trouveront des renseignements précis concernant les éléments dans les notes des livres des codes du fichier principal de l'enquête.

Il n'y a eu aucune imputation pour l'Enquête sur la couverture de l'assurance-emploi pour 2004 et 2005.

7.5 Création de variables dérivées

Un grand nombre de données élémentaires incluses dans le fichier de microdonnées ont été calculées en combinant des postes sur le questionnaire pour faciliter l'analyse des données. Un court nom rappelant la description de la variable (en anglais) a été attribué à tous les éléments du fichier de microdonnées à grande diffusion (FMGD).

Il y a plusieurs types de variables dérivées dans le fichier de données. La présente section contient des renseignements généraux sur chaque type de variables dérivées. Le livre des codes disponible pour le FMGD (voir le chapitre 13.0) comprend une note désignant tous les éléments du questionnaire utilisés pour créer chaque variable dérivée.

7.5.1 Regroupement des éléments de données continus

La plupart des éléments de données recueillis comme variables continues sont compris dans le FMGD comme variables regroupées uniquement. Voici des exemples : âge du répondant (AGECAT), durée d'occupation de l'emploi (TENURE_G), nombre de semaines travaillées dans l'année de référence (WEEKSCAT) et avis avant la perte de l'emploi (NOTICE_W).

Ailleurs, les réponses nominales ont été regroupées pour créer des catégories significatives ou pour réduire le risque d'identification des individus avec des ensembles de réponses uniques. C'est le cas pour le plus haut niveau de scolarité atteint (EDUC), l'industrie et la profession (NAICS6 et OCC6), les méthodes de recherche d'emploi (JOBSRCH), l'aide requise pour trouver un emploi (HELPFIND), les dispositions relatives à la garde des enfants (CHLDCARE), l'intervalle de revenu du ménage (M_HHINC – pour les mères seulement), le type de famille économique (EFAMILY) et quelques autres variables.

7.5.2 Regroupement de questions identiques

Avant 2004, le questionnaire de l'ECAE (voir le chapitre 12.0) contenait un certain nombre de questions reprises deux ou trois fois avec un libellé légèrement différent. Cette pratique était fréquente pour les questions concernant les prestations d'assurance-emploi demandées et reçues. Un libellé différent a été utilisé pour les mères afin de tenir compte d'une période de référence différente, la naissance ou l'adoption d'un enfant, ainsi que pour les répondants qui travaillent et ceux qui ne travaillent pas afin que les questions s'appliquent davantage à la situation du répondant. Dans la plupart des cas, la variable dérivée a été créée en regroupant tout simplement les réponses à deux questions. C'est le cas pour les répondants travaillant au moment de l'interview (WORKNOW), le type de prestations reçues pendant la semaine ou le mois de référence

(BENTYP), la durée des prestations (BENWEEKS), le montant des prestations (BENAMNT), la réception d'un préavis en cas d'interruption de travail (NOTICE), le fait d'arrêter de travailler après la naissance ou l'adoption (BREAKWRK) et les prestations parentales demandées par le conjoint (SPCLAIM).

Dans le même ordre d'idées, les questions concernant l'emploi après la naissance ou l'adoption (EMPAGREE, SAMEMP et WORKCOND) et la garde des enfants (CHLDCARE) étaient formulées différemment selon qu'elles étaient posées aux mères en congé ou aux mères qui étaient déjà retournées au travail.

7.5.3 Regroupement de données de l'Enquête sur la population active et de l'Enquête sur la couverture de l'assurance-emploi

Les questions sur l'employeur et les conditions d'emploi n'ont été posées dans l'ECAE que si l'information n'était pas déjà disponible dans l'Enquête sur la population active (EPA). Dans l'EPA, ces questions portent sur l'emploi actuel ou, pour certains éléments, sur l'emploi précédent s'il était occupé l'année précédente. Les responsables de l'ECAE désirent recueillir ces données concernant tous les répondants qui ont travaillé dans les deux années précédentes. En règle générale, le nom de la variable qui se trouve dans le fichier de microdonnées de l'EPA était utilisé (FTPT, HRLYEARN). Un bon nombre de ces variables liées à l'emploi ont été regroupées pour le fichier de microdonnées à grande diffusion de l'ECAE.

7.5.4 Regroupement de deux questions différentes ou plus

Les variables comme la syndicalisation (UNIONCA), les conditions de travail (WRKTYP), la raison pour laquelle il y a eu arrêt de travail (RWORK), la demande de prestations (CLAIM), le fait de toucher des prestations (BENEFIT), la raison pour laquelle le répondant ne touche ni ne demande de prestations (RNBENRW), le fait de toucher des paiements supplémentaires (ADDPAYM) et la recherche de travail à l'extérieur de la collectivité (LOOKOUT) sont dérivées à l'aide de plus d'une question.

Dans ces cas, l'algorithme utilisé pour créer la nouvelle variable est habituellement assez intuitif. Par exemple, la variable sur les conditions de travail est créée en combinant le statut d'employé à temps plein ou à temps partiel, la nature permanente ou temporaire de l'emploi et la raison pour laquelle l'emploi est temporaire ainsi que la catégorie de travailleur.

Statut d'emploi à temps plein ou à temps partiel (FTPT)

Univers : Employés rémunérés lors du dernier emploi ou de l'emploi actuel

- 1 Temps plein
- 2 Temps partiel

Statut d'emploi permanent ou temporaire (PERMTEMP) (disponible seulement dans le fichier maître de l'enquête)

Univers : Employés rémunérés lors du dernier emploi ou de l'emploi actuel

- 1 Permanent
- 2 Non permanent, emploi saisonnier
- 3 Non permanent, emploi temporaire, d'une durée déterminée ou à contrat
- 4 Non permanent, travail occasionnel
- 5 Non permanent, emploi obtenu par l'intermédiaire d'une agence de placement temporaire
- 6 Non permanent, autre

Catégorie de travailleur à l'emploi principal (COW)

Univers : Répondants qui ont déjà travaillé

- 1 Employé du secteur public ou privé
- 2 Travailleur indépendant qui a une entreprise constituée en société ou non, avec ou sans employés
- 3 Secteur privé, travaille dans une entreprise familiale, sans rémunération

Type de conditions de travail (WRKTYP) (variable dérivée)

Univers : Répondants qui ont déjà travaillé

- 01 Employé permanent, à temps plein (FTPT = 1 et PERMTEMP = 1)
- 02 Employé permanent, à temps partiel (FTPT = 2 et PERMTEMP = 1)
- 03 Employé permanent, travaillant un nombre d'heures indéterminées (FTPT = 9 et PERMTEMP = 1)
- 04 Employé non permanent, emploi saisonnier (PERMTEMP = 2)
- 05 Employé non permanent, autre (PERMTEMP = 3, 4 ou 5)
- 06 Travailleur indépendant (COW = 2)

D'autres variables dérivées sont établies en utilisant des règles plus complexes. C'est le cas de la variable dérivée COV, créée pour représenter la protection par le régime d'assurance-emploi.

7.5.5 Classification de la protection par le régime d'assurance-emploi : la variable COV

L'ECAE fournit des renseignements sur la situation des personnes sans emploi relativement aux prestations d'assurance-emploi. Il s'agit d'une enquête et non pas d'une source de données administratives. Les données administratives de l'assurance-emploi représentent les décisions concrètes des agents au sujet des demandes de prestations reçues par Ressources humaines et Développement social Canada (RHDSC). D'autre part, dans cette même enquête, les estimations du degré de protection de la population canadienne par le régime d'assurance-emploi sont préparées en fonction des comportements, des événements et des perceptions signalés par les répondants au cours d'une enquête téléphonique auprès des ménages.

Voici la classification utilisée par RHDSC concernant la protection par le régime d'assurance-emploi. Les catégories ont été établies selon l'ordre hiérarchique exposé ci-dessous.

Les quatre premières catégories sont incompatibles et regroupent tous les répondants qui ont reçu des prestations depuis leur dernier emploi ou qui s'attendaient à recevoir des

prestations pour la semaine de référence au moment de l'interview. Certains répondants dans ces quatre groupes ont quitté leur emploi, sont retournés aux études, étaient des travailleurs autonomes dans leur dernier emploi ou ont été sans travail pendant plus d'une année. Malgré cela, le fait qu'ils ont touché des prestations d'assurance-emploi dans la dernière année établit clairement leur admissibilité.

COV = 1	Le répondant a reçu des prestations ordinaires d'assurance-emploi pendant la semaine de référence (BENEFIT et BENTYP).
COV = 2	Le répondant a touché des prestations spéciales d'assurance-emploi pendant la semaine de référence (BENEFIT et BENTYP).
COV = 3	Le répondant n'a pas touché de prestations pendant la semaine de référence, mais il s'attendait à en toucher (BENEFIT et RNBENRW). On estime que les personnes peuvent recevoir des prestations lorsqu'elles disent avoir présenté une demande en ce sens et n'avoir pas reçu de prestations pendant la semaine de référence, mais qu'elles attendent toujours les prestations pour la semaine de référence, qu'elles sont dans une période d'attente ou que les prestations sont retenues en raison du versement d'une indemnité de départ ou d'autres paiements ou pour d'autres raisons.
COV = 4	Le répondant n'a pas touché de prestations pour la semaine de référence, mais il a reçu certaines prestations d'assurance-emploi depuis son dernier emploi (dans les 12 derniers mois).

Viennent ensuite les répondants qui n'ont pas cotisé au régime d'assurance-emploi et qui sont donc potentiellement non admissibles aux prestations.

COV = 12	Le répondant n'a jamais travaillé.
COV = 11	Le dernier travail du répondant remonte à plus de 12 mois.
COV = 10	Le répondant n'était pas un employé rémunéré à son dernier emploi ou il a déclaré ne pas avoir cotisé au régime d'assurance-emploi à son dernier emploi (WRKTYP et RNBENRW).

Viennent ensuite les autres répondants qui ont cotisé au régime d'assurance-emploi mais qui sont potentiellement non admissibles à cause de la raison pour laquelle ils ont quitté leur dernier emploi.

COV = 9	Le répondant a déclaré ne pas avoir demandé ou reçu de prestations parce qu'il a poursuivi des études ou qu'il a donné la poursuite de ses études comme raison pour quitter son dernier emploi (RNBENRW ou RSWORK).
COV = 8	Le répondant a déclaré ne pas avoir demandé ou reçu de prestations parce qu'il a quitté volontairement son dernier emploi et que d'autres répondants ont indiqué qu'ils avaient quitté leur dernier emploi.

Pour les autres répondants (environ un chômeur sur sept), l'admissibilité à l'assurance-emploi est déterminée en fonction du nombre d'heures travaillées dans l'année précédant l'interruption de travail.

Les trois dernières catégories de la classification de la variable COV reposent largement (mais non exclusivement) sur une estimation (établie à partir des données de l'enquête)

du nombre d'heures assurables travaillées l'année précédente. L'estimation prend en ligne de compte le nombre de semaines travaillées pendant l'année en question, le nombre d'heures travaillées par semaine en moyenne à un emploi à temps plein et le nombre d'heures travaillées en moyenne à un emploi à temps partiel. Le nombre habituel d'heures travaillées au plus récent emploi ou le nombre moyen d'heures travaillées pour tous les employés à temps partiel et à temps plein sont utilisés en cas de non-réponse. Le seuil est fixé à 700 heures pour tous, le plus haut critère au pays.

COV = 7	<p>Le répondant a déclaré ne pas avoir demandé ou reçu de prestations parce qu'il n'avait pas accumulé un nombre suffisant d'heures de travail assurables ou parce qu'il n'avait pas travaillé récemment (RNBENRW).</p> <p>Le répondant dont la durée du dernier emploi était inférieure ou égale à trois mois, étant donné qu'il est impossible de se renseigner sur la nature assurable des heures travaillées aux emplois précédents occupés pendant l'année (il peut s'agir d'un travail autonome ou d'un autre emploi non assuré) (TENURE_G).</p> <p>L'estimation du nombre d'heures assurables est inférieure à 700.</p>
COV = 5	<p>L'estimation du nombre d'heures assurables est de 700 ou plus, mais le répondant n'a pas présenté de demande de prestations.</p>
COV = 6	<p>L'estimation du nombre d'heures assurables est de 700 ou plus et le répondant a présenté une demande de prestations (mais il ne les a pas reçues).</p>

Cela termine la définition de la variable COV. La variable dérivée ELIGIBLE résume la variable COV de la façon suivante :

- 1 potentiellement admissible, admissible (COV = 1 à 6)
- 2 potentiellement admissible, non admissible (COV = 7)
- 3 potentiellement non admissible (COV = 8 à 12).

La principale mesure de la protection du régime d'assurance-emploi publiée à partir des données de l'enquête exprime l'estimation de l'admissibilité (ELIGIBLE = 1) en pourcentage des personnes potentiellement admissibles (ELIGIBLE = 1 ou 2).

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 admissibles à l'assurance-emploi, doit être estimé, cette opération s'effectue en sélectionnant les enregistrements se référant aux personnes incluses à l'intérieur de l'échantillon qui présentent cette caractéristique et en additionnant les poids inscrits dans ces enregistrements.

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

7.7 Suppression de renseignements confidentiels

Il convient de souligner que les fichiers de microdonnées « à grande diffusion » (FMGD) peuvent différer des fichiers « maîtres » de l'enquête que conserve Statistique Canada. Ces différences sont habituellement le résultat de mesures prises pour protéger l'anonymat des répondants à une enquête. Les mesures les plus courantes sont la suppression de données élémentaires et le regroupement de valeurs en des catégories plus étendues. Dans le cas de certaines variables susceptibles d'identifier des individus, certains éléments du FMGD peuvent avoir été supprimés, c'est-à-dire que le code « non déclaré » peut avoir été attribué dans le FMGD à certaines valeurs présentes dans le fichier principal.

Le fichier principal de l'enquête comprend des identificateurs géographiques pour les dix provinces et pour les régions économiques de l'assurance-emploi. Pour sa part, le FMGD ne contient pas d'identificateur géographique inférieur au niveau provincial, et certaines provinces ont été regroupées (la région de l'Atlantique ainsi que le Manitoba et la Saskatchewan). Le regroupement des provinces a été effectué pour éviter une suppression excessive de variables utiles.

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. Dans le même ordre d'idées, des précisions sur l'industrie et la profession, la durée de l'emploi, le nombre de mois depuis le dernier travail et l'âge du bébé en mois (mères seulement) ainsi que plusieurs autres variables détaillées ne sont disponibles que dans le fichier principal de l'enquête.

Les utilisateurs ayant besoin d'avoir accès à de l'information exclue des fichiers de microdonnées peuvent acheter des totalisations spéciales. Les estimations produites seront communiquées à l'utilisateur, sous réserve du respect des lignes directrices pour l'analyse et la diffusion dont le chapitre 9.0 de ce document fournit un aperçu.

8.0 Qualité des données

8.1 Taux de réponse

Les tableaux ci-dessous présentent la taille des ménages, le nombre de répondants et le taux de réponse à l'Enquête sur la couverture de l'assurance-emploi (ECAE).

Province	2004			2005		
	Échantillon admissible	Réponse	Taux de réponse (%)	Échantillon admissible	Réponse	Taux de réponse (%)
Terre-Neuve-et-Labrador	733	617	84	603	529	88
Île-du-Prince-Édouard	434	358	82	377	340	90
Nouvelle-Écosse	850	731	86	737	654	89
Nouveau-Brunswick	800	695	87	757	631	83
Québec	2 427	2 127	88	2 542	2 140	84
Ontario	4 044	3 392	84	4 004	3 436	86
Manitoba	948	834	88	873	756	87
Saskatchewan	967	849	88	998	891	89
Alberta	1 423	1 200	84	1 334	1 123	84
Colombie-Britannique	1 216	1 050	86	1 419	1 185	84
Canada	13 842	11 853	86	13 644	11 685	86

Nota Le taux de réponse à l'ECAE correspond au nombre d'individus ayant répondu au questionnaire de l'ECAE exprimé sous forme de pourcentage du nombre d'individus admissibles sélectionnés pour l'ECAE (voir les sections 5.5.2 et 7.2).

8.2 Erreurs relatives à l'enquête

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

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

Sur un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations calculées à partir de l'enquête. Toutefois, les erreurs systématiques contribuent à biaiser les

estimations de l'enquête. Énormément de temps et d'efforts ont été consacrés à réduire les erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été prises à chacune des étapes 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'ECAE était une enquête supplémentaire à l'Enquête sur la population active (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'ECAE. La qualité des variables d'échantillon dans la base était très élevée. L'échantillon de l'ECAE était composé d'un groupe de renouvellement de pour la population « ordinaire » de l'ECAE et de deux groupes de renouvellement pour les « mères ».

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'ECAE 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. De plus, la base de sondage de l'ECAE ne tient pas compte de la non-réponse complète à l'EPA et de la non-réponse partielle concernant les variables utilisées dans les critères de sélection

Les variables dans la base de sondage de l'ECAE étaient passablement à jour puisqu'elles ont été recueillies dans le cadre de l'EPA trois semaines tout au plus avant le début de la période de collecte de l'ECAE.

8.2.2 Collecte des données

La formation des intervieweurs pour l'ECAE consistait à lire le Manuel de l'intervieweur, la pratique avec des cas simulés de l'ECAE sur ordinateur et une discussion avec l'intervieweur principal des questions de l'intervieweur avant le début de l'enquête. L'information fournie aux intervieweurs comprenait une description du contexte et des objectifs de l'enquête ainsi qu'un glossaire de terminologie et une série de questions et de réponses. Les intervieweurs ont commencé la collecte des données de l'ECAE deux semaines après la collecte des données de l'EPA de janvier, avril, juillet et novembre. La collecte des données a duré cinq semaines pour chaque cycle de l'ECAE.

8.2.3 Traitement des données

Le traitement des données de l'ECAE a été fait par étapes, y compris la vérification, le codage, le contrôle, l'estimation, la confidentialité, etc. Une photo des fichiers de sortie est prise à chaque étape et un rapport montrant les changements apportés à chaque variable d'une étape à l'autre est établi. La vérification de ces rapports de traitement réduit grandement le risque d'erreur.

Vérification

Les fichiers textes électroniques contenant les transmissions quotidiennes des dossiers terminés sont regroupés pour créer le fichier « brut » de l'enquête. Tous les enregistrements de l'ECAE pourraient être jumelés aux enregistrements correspondants de l'EPA. Aucun enregistrement n'est perdu ou supprimé.

Des enregistrements en double sont parfois créés en raison de problèmes de transmission. L'un des deux enregistrements identiques est alors supprimé ou, si les enregistrements en double ne sont pas parfaitement identiques, l'enregistrement le plus récent est conservé. On trouve rarement des enregistrements en double dans l'ECAE.

Contrôle

Le contrôle consiste à modifier les données au niveau des variables individuelles. La principale forme de contrôle relativement aux données de l'ECAE porte sur l'enchaînement des questions (voir la section 7.2). Les rapports produits par le système de contrôle de l'enchaînement des questions ont été minutieusement examinés pour repérer les erreurs possibles survenues au traitement. Cet examen portait particulièrement sur les éléments comprenant une incidence élevée de réponses « Non déclaré » et sur les éléments où une réponse valide a été remplacée par un « Enchaînement valide » ou un code « Non déclaré ». Très peu de cas restent inexplicables. Le processus de vérification a toutefois mis au jour un certain nombre d'erreurs de réponse (voir la section 8.2.4).

Codage

L'industrie et la profession ont été codées par un groupe de personnes spécialement formées, ce qui a permis de réduire le risque d'erreur de codage. Les éléments propres à l'enquête sont probablement plus vulnérables aux erreurs de codage ou à un code non uniforme d'une année à l'autre. Aucune mesure précise des erreurs de codage n'est disponible.

Variables dérivées

Un grand nombre de variables dérivées ont été créées à partir des données recueillies dans le cadre de l'ECAE. Toutes les variables dérivées sont énoncées dans les tables de décision. Pour chaque variable, le processus génère un tableau récapitulatif sur les règles appliquées et sur la fréquence des règles. La répartition de chaque variable dérivée a été comparée à celles des éléments du questionnaire utilisés pour la créer. Les variables dérivées ont aussi été recoupées avec des variables connexes pour garantir la cohérence interne et limiter le risque d'erreur dans les règles de dérivation. Une comparaison de la répartition sur la période allant de 2004 à 2005 a également été effectuée pour assurer la comparabilité chronologique de l'information comprise dans les FMGD. Les incohérences ne pouvant être résolues ont été signalées dans le livre des codes (note).

8.2.4 Non-réponse

L'une des principales sources d'erreurs non dues à l'échantillonnage observées dans le cadre des enquêtes est l'effet de la non-réponse sur les résultats de l'enquête. L'étendue d'une non-réponse varie d'une non-réponse partielle (le fait de ne pas répondre simplement à une ou à des questions) à une non-réponse totale. S'il y a eu non-réponse totale, c'est parce que l'intervieweur a été incapable de communiquer avec le répondant 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. Elle était invariablement plus importante chez les employés à temps plein (types 4 et 5) au fil des ans et légèrement plus importante chez les hommes, mais il n'y avait aucune différence marquée entre les grands groupes d'âge.

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.

Il n'y a eu aucune imputation de données pour compenser pour la non-réponse totale ou partielle dans l'ECAE.

8.2.5 Mesure de l'erreur d'échantillonnage

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

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

En raison, cependant, de la diversité des estimations pouvant être produites à partir d'une enquête, l'erreur-type d'une estimation est habituellement exprimée en fonction de l'estimation à laquelle elle se rapporte. La mesure résultante, appelée coefficient de variation (CV) d'une estimation, s'obtient en divisant l'erreur-type de l'estimation par l'estimation elle-même et s'exprime en pourcentage de l'estimation.

Par exemple, supposons que, d'après les résultats de l'enquête de 2003, l'on estime que 81% des personnes sont admissibles à l'assurance-emploi parmi les personnes potentiellement admissibles et l'on constate que l'erreur-type de cette estimation est de 0,03. Le coefficient de variation de l'estimation est donc calculé comme suit :

$$\left(\frac{0,03}{0,81} \right) \times 100 \% = 3,7 \%$$

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 la couverture de l'assurance-emploi (ECAE) 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'ECAE, 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'ECAE.

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 chômeurs qui ont reçu des prestations d'assurance-emploi pendant la semaine de référence ou la proportion de chômeurs admissibles à l'assurance-emploi 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 : Des cotisations d'assurance-emploi étaient-elles déduites de votre traitement ou salaire à ce travail chez (nom de l'employeur)?

R : Oui / Non

Q : Quel genre de prestations avez-vous reçu pendant cette semaine?

R : Formation / Ordinaires / Maternité (si femme seulement) / Parentales / Maladie / Pêche / Autre

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 d'absence du travail à la suite de la naissance ou de l'adoption d'un enfant. Le numérateur correspond à une estimation du nombre total de mois d'absence de toutes les mères pour lesquelles cette information est disponible (mères déjà retournées au travail ou dont les intentions sont connues) et son dénominateur est le nombre de mères s'absentant pendant une période connue.

Exemples de questions quantitatives :

Q : Combien de mois ce congé a-t-il duré?

R : |_|_| mois

Q : Au cours des semaines pendant lesquelles vous avez travaillé à temps plein, combien d'heures par semaine est-ce que vous avez travaillé en moyenne?

R : |_|_|_| heures

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 semaines visées par des prestations d'assurance-emploi touchées par les mères d'un enfant qui sont déjà retournées au travail, multipliez la valeur déclarée à la variable

dérivée BENWEEKS (semaines pour lesquelles elles ont reçu des prestations d'assurance-emploi) par le poids final de l'enregistrement, puis additionnez cette valeur pour tous les enregistrements où les variables MOTHER = 1 et WORKNOW = 1 (mère d'un enfant de moins d'un an qui travaille actuellement)..

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, les mères ont reçu des prestations d'assurance-emploi,

- a) estimez le nombre total de semaines (\hat{X}) tel qu'il est décrit ci-dessus,
- b) estimez le nombre de mères travaillant actuellement (\hat{Y}) incluses dans cette catégorie en additionnant les poids finals de tous les enregistrements où les variables MOTHER = 1 et WORKNOW = 1, puis
- c) divisez l'estimation a) par l'estimation b) (\hat{X} / \hat{Y}).

9.4 Lignes directrices pour l'analyse statistique

L'ECAE 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

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'ECAE, 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 la couverture de l'assurance-emploi

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 de Type = 1 possédant une caractéristique donnée dans les provinces de l'Atlantique est médiocre.

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

Provinces et régions pour TYPE = 1	CV acceptable 0,0 à 16,5 %	CV médiocre 16,6 à 33,3 %	CV inacceptable > 33,3 %
Provinces de l'Atlantique	8 300 et plus	2 100 à < 8 300	moins de 2 100
Québec	38 600 et plus	10 400 à < 38 600	moins de 10 400
Ontario	35 600 et plus	9 300 à < 35 600	moins de 9 300
Manitoba et Saskatchewan	15 800 et plus	5 000 à < 15 800	moins de 5 000
Alberta	27 300 et plus	8 900 à < 27 300	moins de 8 900
Colombie-Britannique	34 500 et plus	10 200 à < 34 500	moins de 10 200
Provinces de l'Ouest	28 800 et plus	7 700 à < 28 800	moins de 7 700
Canada	35 400 et plus	8 900 à < 35 400	moins de 8 900

Provinces et régions pour MOTHER = 1	CV acceptable 0,0 à 16,5 %	CV médiocre 16,6 à 33,3 %	CV inacceptable > 33,3 %
Provinces de l'Atlantique	11 900 et plus	5 200 à < 11 900	moins de 5 200
Québec	30 400 et plus	10 500 à < 30 400	moins de 10 500
Ontario	44 300 et plus	14 000 à < 44 300	moins de 14 000
Manitoba et Saskatchewan	7 000 et plus	2 100 à < 7 000	moins de 2 100
Alberta	14 500 et plus	4 900 à < 14 500	moins de 4 900
Colombie-Britannique	23 100 et plus	10 400 à < 23 100	moins de 10 400
Provinces de l'Ouest	25 100 et plus	7 500 à < 25 100	moins de 7 500
Canada	43 900 et plus	11 900 à < 43 900	moins de 11 900

Provinces et régions pour MOTHER = 0 et TYPE = 3	CV acceptable 0,0 à 16,5 %	CV médiocre 16,6 à 33,3 %	CV inacceptable > 33,3 %
Provinces de l'Atlantique	23 600 et plus	6 500 à < 23 600	moins de 6 500
Québec	59 600 et plus	16 500 à < 59 600	moins de 16 500
Ontario	63 800 et plus	16 900 à < 63 800	moins de 16 900
Manitoba et Saskatchewan	16 200 et plus	4 400 à < 16 200	moins de 4 400
Alberta	35 000 et plus	10 100 à < 35 000	moins de 10 100
Colombie-Britannique	47 500 et plus	13 900 à < 47 500	moins de 13 900
Provinces de l'Ouest	39 200 et plus	10 200 à < 39 200	moins de 10 200
Canada	54 200 et plus	13 600 à < 54 200	moins de 13 600

Provinces et régions pour MOTHER = 0 et TYPE = 2 ou 4	CV acceptable 0,0 à 16,5 %	CV médiocre 16,6 à 33,3 %	CV inacceptable > 33,3 %
Provinces de l'Atlantique	21 000 et plus	5 600 à < 21 000	moins de 5 600
Québec	47 200 et plus	12 200 à < 47 200	moins de 12 200
Ontario	42 300 et plus	10 700 à < 42 300	moins de 10 700
Manitoba et Saskatchewan	18 600 et plus	4 900 à < 18 600	moins de 4 900
Alberta	23 200 et plus	6 000 à < 23 200	moins de 6 000
Colombie-Britannique	49 900 et plus	13 400 à < 49 900	moins de 13 400
Provinces de l'Ouest	37 000 et plus	9 300 à < 37 000	moins de 9 300
Canada	36 700 et plus	9 100 à < 36 700	moins de 9 100

10.0 Tables de variabilité d'échantillonnage approximative

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

Les coefficients de variation sont calculés à l'aide de la formule de la variance pour un échantillonnage aléatoire simple et en y incorporant un facteur qui reflète la nature du plan d'échantillonnage, qui est à plusieurs degrés et qui prévoit la formation de grappes. Ce facteur, appelé l'effet du plan, a été déterminé en calculant premièrement les effets du plan pour une gamme étendue de caractéristiques, puis en choisissant parmi ceux-ci une valeur modérée 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 le type d'enquête, le statut de la mère et la province, qui ont été utilisés pour produire les tables de variabilité d'échantillonnage approximative de l'Enquête sur la couverture de l'assurance-emploi (ECAE).

Province et région pour TYPE = 1	Effet du plan	Taille de l'échantillon	Population
Provinces de l'Atlantique	1,37	693	122 900
Québec	1,76	493	333 700
Ontario	1,44	606	444 407
Manitoba et Saskatchewan	2,37	212	54 283
Alberta	1,92	142	82 458
Colombie-Britannique	1,58	194	149 883
Provinces de l'Ouest	1,67	548	286 623
Canada	1,96	2 340	1 187 630

Province et région pour MOTHER = 1	Effet du plan	Taille de l'échantillon	Population
Provinces de l'Atlantique	6,46	169	20 409
Québec	3,61	209	78 329
Ontario	5,03	430	147 582
Manitoba et Saskatchewan	1,91	189	25 905
Alberta	2,44	155	39 630
Colombie-Britannique	3,96	95	38 300
Provinces de l'Ouest	3,82	439	103 835
Canada	4,89	1 247	350 156

Province et région pour MOTHER = 0 et TYPE = 3	Effet du plan	Taille de l'échantillon	Population
Provinces de l'Atlantique	3,20	700	164 742
Québec	2,49	513	394 228
Ontario	2,28	785	662 952
Manitoba et Saskatchewan	2,05	446	112 219
Alberta	2,11	313	176 632
Colombie-Britannique	1,81	249	225 519
Provinces de l'Ouest	2,27	1 008	514 370
Canada	2,64	3 006	1 736 292

Province et région pour MOTHER = 0 et TYPE = 2 ou 4	Effet du plan	Taille de l'échantillon	Population
Provinces de l'Atlantique	2,53	790	200 085
Québec	1,71	858	693 418
Ontario	1,54	1 500	1 166 654
Manitoba et Saskatchewan	2,24	818	204 163
Alberta	1,14	566	337 558
Colombie-Britannique	1,78	497	429 435
Provinces de l'Ouest	2,03	1 881	971 156
Canada	1,68	5 029	3 031 314

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, de chômeurs qui reçoivent des prestations régulières d'assurance-emploi pendant la semaine de référence est plus fiable que le nombre estimé de chômeurs qui reçoivent des prestations régulières d'assurance-emploi pendant la semaine de référence. (Remarquez que dans les tables la valeur des coefficients de variation diminue lorsqu'on les lit de gauche à droite.)

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

Lorsque la proportion ou que le pourcentage repose sur un sous-ensemble de la population totale (p. ex., comme ses membres d'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 chômeurs potentiellement admissibles à l'assurance-emploi et le numérateur, le nombre de chômeurs admissibles à l'assurance-emploi.

Dans le cas où le numérateur n'est pas un sous-ensemble du dénominateur, comme dans l'exemple du rapport du nombre de chômeurs qui reçoivent des prestations régulières d'assurance-emploi comparativement au nombre de chômeurs qui reçoivent tout autre type de prestations, l'erreur-type du rapport des estimations est approximativement égale à la racine carrée de la somme des carrés de chaque coefficient de variation considéré séparément multipliée par \hat{R} . C'est-à-dire que l'erreur-type d'un rapport ($\hat{R} = \hat{X}_1 / \hat{X}_2$) est :

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

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

Règle 5 : Estimations de différences entre des rapports

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

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

Les exemples ci-dessous utilisent des données du fichier de l'ECAE de 2003 et sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter. Veuillez noter que les données utilisées dans ces exemples diffèrent des véritables résultats de l'enquête et tiennent seulement lieu de guide.

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

Supposons qu'un utilisateur estime que 400 393 chômeurs ont reçu des prestations régulières d'assurance-emploi pendant la semaine de référence. 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 TYPE = 1 CANADA.

Enquête sur la couverture de l'assurance-emploi, 2000 à 2003														
Tables de variabilité d'échantillonnage approximative - TYPE = 1 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	98,9	98,4	97,9	96,4	93,9	91,2	88,5	85,7	82,8	79,8	76,6	70,0	54,2	31,3
2	*****	69,6	69,3	68,2	66,4	64,5	62,6	60,6	58,5	56,4	54,2	49,5	38,3	22,1
3	*****	56,8	56,6	55,7	54,2	52,7	51,1	49,5	47,8	46,1	44,2	40,4	31,3	18,1
4	*****	49,2	49,0	48,2	46,9	45,6	44,2	42,8	41,4	39,9	38,3	35,0	27,1	15,6
5	*****	44,0	43,8	43,1	42,0	40,8	39,6	38,3	37,0	35,7	34,3	31,3	24,2	14,0
6	*****	40,2	40,0	39,4	38,3	37,2	36,1	35,0	33,8	32,6	31,3	28,6	22,1	12,8
7	*****	37,2	37,0	36,5	35,5	34,5	33,4	32,4	31,3	30,2	29,0	26,4	20,5	11,8
.
.
.
250	*****	*****	*****	*****	*****	*****	*****	5,4	5,2	5,0	4,8	4,4	3,4	2,0
300	*****	*****	*****	*****	*****	*****	*****	4,9	4,8	4,6	4,4	4,0	3,1	1,8
350	*****	*****	*****	*****	*****	*****	*****	*****	4,4	4,3	4,1	3,7	2,9	1,7
400	*****	*****	*****	*****	*****	*****	*****	*****	*****	4,0	3,8	3,5	2,7	1,6
450	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	3,6	3,3	2,6	1,5
500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	3,1	2,4	1,4
750	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,0	1,1
1 000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1,0

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

- L'agrégat estimé (400 393) 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.
- 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 4,0 %.
- Le coefficient de variation approximatif de l'estimation est donc 4,0 %. Le résultat selon lequel il y avait 400 393 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1) chômeurs qui ont reçu des prestations régulières d'assurance-emploi pendant la semaine de référence, peut être publié sans réserve.

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

Supposons qu'un utilisateur estime que $605\,777 / 740\,586 = 81,8\%$ des chômeurs potentiellement admissibles à l'assurance-emploi étaient admissibles à l'assurance-emploi. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

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

- 2) Parce que l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c'est-à-dire les chômeurs potentiellement admissibles à l'assurance-emploi, il faut utiliser à la fois le pourcentage (81,8 %) et la portion numérateur du pourcentage (605 777) pour déterminer le coefficient de variation.
- 3) Le numérateur, 605 777, 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 1,4 %, est le coefficient de variation à employer.
- 5) Le coefficient de variation approximatif de l'estimation est donc 1,4 %. Le résultat selon lequel 81,8 % des chômeurs potentiellement admissibles à l'assurance-emploi étaient admissibles à l'assurance-emploi, peut être publié sans réserve.

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

Supposons qu'un utilisateur estime que $543\,846 / 718\,300 = 75,7\%$ de la population ordinaire employée au Québec a cotisé à l'assurance-emploi, tandis que $775\,530 / 1\,172\,069 = 66,2\%$ de la population ordinaire en Ontario a cotisé à l'assurance-emploi. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

Enquête sur la couverture de l'assurance-emploi, 2000 à 2003														
Tables de variabilité d'échantillonnage approximative - MOTHER = 0 et TYPE = 2 ou 4 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	*****	103,2	102,7	101,1	98,4	95,7	92,8	89,9	86,8	83,6	80,4	73,4	56,8	32,8
2	*****	73,0	72,6	71,5	69,6	67,6	65,6	63,5	61,4	59,1	56,8	51,9	40,2	23,2
3	*****	59,6	59,3	58,4	56,8	55,2	53,6	51,9	50,1	48,3	46,4	42,4	32,8	18,9
4	*****	51,6	51,4	50,6	49,2	47,8	46,4	44,9	43,4	41,8	40,2	36,7	28,4	16,4
5	*****	46,2	45,9	45,2	44,0	42,8	41,5	40,2	38,8	37,4	35,9	32,8	25,4	14,7
6	*****	42,1	41,9	41,3	40,2	39,1	37,9	36,7	35,4	34,1	32,8	30,0	23,2	13,4
7	*****	39,0	38,8	38,2	37,2	36,2	35,1	34,0	32,8	31,6	30,4	27,7	21,5	12,4
.
.
150	*****	*****	*****	*****	*****	*****	*****	7,3	7,1	6,8	6,6	6,0	4,6	2,7
200	*****	*****	*****	*****	*****	*****	*****	*****	1,0	5,9	5,7	5,2	4,0	2,3
250	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	5,1	4,6	3,6	2,1
300	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	4,2	3,3	1,9
350	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	3,9	3,0	1,8
400	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,8	1,6
450	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,7	1,5
500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,5	1,5

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

Enquête sur la couverture de l'assurance-emploi, 2000 à 2003

Tables de variabilité d'échantillonnage approximative - MOTHER = 0 et TYPE = 2 ou 4 Ontario

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	122,9	122,4	121,8	119,9	116,7	113,4	110,0	106,5	102,9	99,2	95,3	87,0	67,4	38,9
2	*****	86,5	86,1	84,8	82,5	80,2	77,8	75,3	72,8	70,1	67,4	61,5	47,6	27,5
3	*****	70,7	70,3	69,2	67,4	65,5	63,5	61,5	59,4	57,2	55,0	50,2	38,9	22,5
4	*****	61,2	60,9	59,9	58,3	56,7	55,0	53,3	51,4	49,6	47,6	43,5	33,7	19,4
5	*****	54,7	54,4	53,6	52,2	50,7	49,2	47,6	46,0	44,3	42,6	38,9	30,1	17,4
6	*****	50,0	49,7	48,9	47,6	46,3	44,9	43,5	42,0	40,5	38,9	35,5	27,5	15,9
7	*****	46,3	46,0	45,3	44,1	42,9	41,6	40,3	38,9	37,5	36,0	32,9	25,5	14,7
.
.
.
350	*****	*****	*****	*****	*****	*****	*****	*****	5,5	5,3	5,1	4,6	3,6	2,1
400	*****	*****	*****	*****	*****	*****	*****	*****	*****	5,0	4,8	4,3	3,4	1,9
450	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	4,5	4,1	3,2	1,8
500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	3,9	3,0	1,7
750	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,5	1,4
1 000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1,2

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

1) L'utilisation de la table des coefficients de variation pour MOTHER = 0 et TYPE = 2 ou 4 QUÉBEC et la table des coefficients de variation pour MOTHER = 0 et TYPE = 2 ou 4 ONTARIO de la même façon que celle décrite dans l'exemple 2, donne un CV de l'estimation pour le Québec de 2,5 % et un CV de l'estimation pour l'Ontario de 2,5 %.

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 $\hat{d} = 0,757 - 0,662 = 0,095$ est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0,757)(0,025)]^2 + [(0,662)(0,025)]^2} \\ &= \sqrt{(0,0003581) + (0,0002739)} \\ &= 0,025\end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,025 / 0,095 = 0,263$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 26,3 %. La différence entre les estimations est considérée médiocre et Statistique Canada recommande que cette estimation soit signalée par la lettre E (ou un autre identificateur similaire). Elle devrait être accompagnée d'une mise en garde avertissant les utilisateurs subséquents du niveau élevé d'erreur associé à cette estimation.

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime que 543 846 personnes de la population ordinaire employée au Québec ont cotisé à l'assurance-emploi, par rapport à 775 530 personnes de la population ordinaire employée en Ontario ont cotisé à l'assurance-emploi. L'utilisateur est intéressé à comparer l'estimation du Québec en comparaison de l'estimation de l'Ontario sous la forme d'un rapport. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation d'un rapport, où le numérateur de l'estimation (\hat{X}_1) est le nombre de personnes employées au Québec qui ont cotisé à l'assurance-emploi. Le dénominateur de l'estimation (\hat{X}_2) est le nombre de personnes employées en Ontario qui ont cotisé à l'assurance-emploi.
- 2) Reportez-vous à la table des coefficients de variation pour MOTHER = 0 et TYPE = 2 ou 4 QUÉBEC et MOTHER = 0 et TYPE = 2 ou 4 ONTARIO.
- 3) Le numérateur de cette estimation de rapport est 543 846. Le chiffre qui s'en rapproche le plus est 500 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 2,5 %.
- 4) Le dénominateur de cette estimation de rapport est 775 530. Le chiffre qui s'en rapproche le plus est 750 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 2,5 %.
- 5) Le coefficient de variation approximatif de l'estimation du rapport est donc donné par la règle 4, qui est :

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

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement.

C'est-à-dire que :

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0,025)^2 + (0,025)^2} \\ &= \sqrt{0,000625 + 0,000625} \\ &= 0,035\end{aligned}$$

- 6) Le rapport obtenu entre les personnes de la population ordinaire employée au Québec en comparaison de celles de l'Ontario qui cotisent à l'assurance-emploi est 543,846 / 775,530, c'est-à-dire 0,70 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1). Le coefficient de variation de cette estimation est 3,5 %, 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 âgées de 15 à 24 ans de la population ordinaire employée qui ont cotisé à l'assurance-emploi et les personnes âgées de 25 à 44 ans de la population ordinaire employée qui ont cotisé à l'assurance-emploi équivaut à 1,24 au Manitoba et en Saskatchewan, alors qu'il équivaut à 1,21 en Alberta. 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 Manitoba et de la Saskatchewan (\hat{R}_1) et le rapport de l'Alberta (\hat{R}_2) tel qu'il est décrit dans l'exemple 4. Le CV approximatif pour le rapport du Manitoba et de la Saskatchewan est 12,8 % et 11,3 % pour celui de l'Alberta.
- 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 $\hat{d} = 1,24 - 1,21 = 0,03$ est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(1,24)(0,128)]^2 + [(1,21)(0,113)]^2} \\ &= \sqrt{(0,025192)^2 + (0,018695)^2} \\ &= 0,209\end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,209 / 0,03 = 6,967$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 696,7 %. La différence entre les estimations est considérée inacceptable 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 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.

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 chômeurs qui étaient potentiellement admissibles à l'assurance-emploi étaient admissibles à l'assurance-emploi (d'après l'exemple 2 à la section 10.1.1) serait calculé comme suit :

$$\hat{X} = 81,8 \% \text{ (ou exprimé sous forme de proportion } 0,818)$$

$$t = 2$$

$$\alpha_{\hat{x}} = 1,4 \% \text{ (0,014 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.}$$

$$IC_{\hat{x}} = \{0,818 - (2) (0,818) (0,014), 0,818 + (2) (0,818) (0,014)\}$$

$$IC_{\hat{x}} = \{0,818 - 0,023, 0,818 + 0,023\}$$

$$IC_{\hat{x}} = \{0,795, 0,841\}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 79,5 % et 84,1 % des chômeurs qui étaient potentiellement admissibles à l'assurance-emploi étaient admissibles à l'assurance-emploi.

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

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

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

Si $t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ se situe entre -2 et 2, aucune conclusion à propos de la différence entre les caractéristiques n'est alors justifiée au niveau de signification de 5 %. Si, cependant, ce rapport

est inférieur à -2 ou supérieur à +2, la différence observée est significative au niveau de 0,05. C'est-à-dire que la différence entre les estimations est significative.

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

Supposons que l'utilisateur désire tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y a pas de différence entre la proportion de la population ordinaire employée au Québec qui a cotisé à l'assurance-emploi et la proportion de la population ordinaire employée en Ontario qui a cotisé à l'assurance-emploi. 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 0,025. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_d} = \frac{0,757 - 0,662}{0,025} = \frac{0,095}{0,025} = 3,80$$

Puisque $t = 3,80$ est supérieur à 2, il faut en conclure qu'il existe une différence significative entre les deux estimations au niveau de signification de 0,05.

10.4 Coefficients de variation pour des estimations quantitatives

Il faudrait produire des tables spéciales afin de déterminer l'erreur d'échantillonnage d'estimations quantitatives, ce qui n'a pas été fait, parce que la plupart des variables pour l'ECAE 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 chômeurs qui reçoivent des prestations régulières de l'assurance-emploi serait supérieur au coefficient de variation de la proportion correspondante de chômeurs qui reçoivent des prestations régulières d'assurance-emploi. 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 ECAE2004-2005_CVTabF.doc pour les tables de coefficients de variation.

11.0 Pondération

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

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

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

Poids de base

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

Sous-poids d'une grappe

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

Poids de stabilisation

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

des logements appartenant à la même région économique de l'assurance-emploi et au même groupe de renouvellement.

Non-réponse

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

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

Sous-poids de l'Enquête sur la population active

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

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

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

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

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

11.2 Procédures de pondération pour l'Enquête sur la couverture de l'assurance-emploi

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

- 1) Un ajustement qui tient compte de l'utilisation d'un sous-échantillon représentant les un sixième(s), plutôt que de l'échantillon complet de l'EPA. Dans le cas des mères, il s'agit des deux sixièmes.
- 2) Un ajustement qui tient compte du sous-échantillonnage de l'ECAE (voir la section 5.5.1).
- 3) Un ajustement qui tient compte des non-réponses additionnelles au questionnaire de l'enquête supplémentaire, c'est-à-dire des non-réponses au questionnaire de l'ECAE attribuables à des gens qui ont répondu au questionnaire de l'EPA ou pour qui des données de l'EPA du mois précédent ont été reportées. La procédure est similaire à l'ajustement des poids des non-réponses au questionnaire de l'EPA, mais les regroupements reposent sur des variables différentes. Ces variables sont la province, le type de répondant, le sexe et un regroupement de régions de l'assurance-emploi.
- 4) Un dernier ajustement est effectué à l'aide de deux sources externes et indépendantes ne se chevauchant pas. Ressources humaines et Développement social Canada fournit des chiffres estimatifs des bénéficiaires ordinaires avec et sans revenu. L'autre source est constituée des données de l'EPA, qui fournit des chiffres estimatifs des chômeurs (nom désaisonnalisés). L'ajustement est réalisé dans un processus de calibration qui garantit que les estimations produites avec les données de l'ECAE correspondent aux chiffres provenant des sources externes. Le poids final calibré équivaut au poids avant la calibration multiplié par le facteur nécessaire pour calibrer à la source indépendante applicable. La nouvelle partie de la population observée dans le contexte de l'ECAE, formée des mères d'enfant de moins de un an, n'est pas comprise dans ce processus de calibration.

Le poids qui en résulte WTPM (pour « final weight » en anglais) est le poids final qui figure dans le fichier de microdonnées de l'ECAE.

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 et 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 Les questionnaires de l'Enquête sur la couverture de l'assurance-emploi

Les questionnaires qui ont servi à recueillir des données pour l'Enquête sur la couverture de l'assurance-emploi (ECAE) sont :

ECAE2004_QuestF.pdf

ECAE2005_QuestF.pdf

13.0 Cliché d'enregistrement à valeurs univariées

Consulter les deux fichiers ci-dessous qui renferment les clichés d'enregistrement à valeurs univariées des fichiers de microdonnées pour l'Enquête sur la couverture de l'assurance-emploi (ECAE).

ECAE2004_Maître_LvCds.pdf

ECAE2005_Maître_LvCds.pdf