

Guide de l'utilisateur des microdonnées

Enquête sur la couverture de l'assurance-emploi

2020



Statistics
Canada

Statistique
Canada

Canada

Contents

1.0	Introduction	5
2.0	Contexte.....	6
3.0	Objectifs	7
4.0	Concepts et définitions	9
4.1	Concepts et définitions de l'Enquête sur la population active	9
4.2	Concepts et définitions de l'Enquête sur la couverture de l'assurance-emploi	10
5.0	Méthodologie de l'enquête	13
5.1	Population visée.....	13
5.2	Plan d'échantillonnage et taille de l'échantillon.....	13
5.3	Sélection de logements.....	14
5.4	Sélection des membres du ménage.....	14
5.5	Rotation de l'échantillon.....	14
5.6	Modifications apportées au plan de l'Enquête sur la population active pour l'Enquête sur la couverture de l'assurance-emploi	15
5.6.1	Population cible	15
5.6.2	Type 4 : Un cas particulier.....	15
5.6.3	Autres exclusions	15
5.7	Taille de l'échantillon selon la province pour l'Enquête sur la couverture de l'assurance-emploi 16	
6.0	Collecte des données	17
6.1	Réalisation des interviews dans le cadre de l'Enquête sur la population active	17
6.2	Supervision et contrôle de qualité.....	17
6.3	Non-réponse à l'Enquête sur la population active	17
6.4	Modifications apportées à la collecte des données de l'Enquête sur la couverture de l'assurance-emploi.....	18
6.5	Non-réponse à l'Enquête sur la couverture de l'assurance-emploi	18
7.0	Traitement des données	19
7.1	Saisie des données	19
7.2	Vérification et contrôle	19
7.3	Codage des questions ouvertes	20
7.4	Imputation	20
7.5	Création de variables dérivées.....	21
7.5.1	Regroupement des éléments de données continus	21

La plupart des éléments de données collectés en tant que variables continues ne figurent dans le fichier de microdonnées à usage public que sous forme de variables groupées. L'âge du répondant

(AGECAT), l'ancienneté dans l'emploi (TENURE_G) et le nombre de semaines travaillées pendant l'année de référence (WEEKSCAT) en sont des exemples.	21
Dans d'autres situations, les éléments de réponse catégoriels ont été regroupés pour créer des catégories significatives ou pour réduire le risque d'identifier des individus avec des ensembles de réponses uniques. C'est le cas pour le plus haut niveau de scolarité atteint (EDUC), la branche d'activité et la profession (SCIAN6 et CNP6), le mode de garde d'enfants prévu ou actuel (CHLDCARE), le régime de prestations parentales ou l'option choisie (BENPLAN - parents seulement), le type de famille économique (EFAMILY) et quelques autres.....	21
7.5.2 Regroupement de données de l'Enquête sur la population active et de l'Enquête sur la couverture de l'assurance-emploi	21
7.5.3 Regroupement de deux questions différentes ou plus.....	21
7.5.4 Classification de la protection par le régime d'assurance-emploi : la variable COV	22
7.6 Pondération	24
7.7 Suppression de renseignements confidentiels	25
8.0 Qualité des données	26
8.1 Taux de réponse.....	26
8.2 Erreurs relatives à l'enquête.....	26
8.2.1 Base de sondage	27
8.2.2 Collecte des données	27
En 2020, en raison des arrêts liés à COVID-19 en mars et avril, l'EICS a utilisé les groupes de rotation qui ont terminé leurs six mois dans l'EFT en juin, août, octobre et décembre.	27
8.2.3 Traitement des données	27
8.2.5 Non-réponse	29
8.2.6 Mesure de l'erreur d'échantillonnage	29
9.0 Lignes directrices pour la totalisation, l'analyse et la diffusion de données	30
9.1 Lignes directrices pour l'arrondissement d'estimations.....	30
9.2 Lignes directrices pour la pondération de l'échantillon en vue de la totalisation.....	31
9.3 Définitions de types d'estimations : catégoriques et quantitatives	31
9.3.1 Estimations catégoriques.....	31
9.3.2 Estimations quantitatives	32
9.3.3 Totalisation d'estimations catégoriques.....	32
9.3.4 Totalisation d'estimations quantitatives	32
9.4 Lignes directrices pour l'analyse statistique	33
9.5 Lignes directrices pour la diffusion de coefficients de variation	34
10.0 Pondération bootstrap dans les FMGD	35
10.1 Comment utiliser des poids bootstrap pour des estimations de variables qualitatives.....	35

10.1.1	Exemples d'utilisation de poids bootstrap pour des estimations de variables qualitatives	38
10.2	Comment utiliser des poids bootstrap pour obtenir des limites de confiance	41
10.2.1	Exemple d'utilisation de l'écart type pour obtenir des limites de confiance	42
10.2.2	Exemple d'utilisation des poids bootstrap pour obtenir des limites de confiance	42
10.3	Comment utiliser l'erreur-type bootstrap pour effectuer un test- <i>t</i>	43
10.3.1	Exemple d'utilisation des tables de coefficients de variation pour effectuer un test <i>t</i>	43
10.4	Coefficients de variation pour des estimations quantitatives	43
11.0	Pondération	44
11.1	Procédures de pondération pour l'Enquête sur la population active	44
11.2	Procédures de pondération pour l'Enquête sur la couverture de l'assurance-emploi	46

1.0 Introduction

L'Enquête sur la couverture de l'assurance-emploi (ECAE) est menée par Statistique Canada avec la collaboration et l'appui d'Emploi et Développement social Canada. Ce manuel a été produit pour faciliter la manipulation du fichier de micro-données 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

Centre de l'information sur le marché du travail

Téléphone: 1-866-873-8788

Courriel : statcan.labour-travail.statcan@canada.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.

Avant 2020, la population cible pour les parents était celle des mères d'un enfant d'un an ou moins au cours de la semaine de référence. À partir de 2020, la population cible pour les parents a été élargie aux parents d'un enfant de 18 mois ou moins pour inclure les pères dans les ménages sans parent de sexe féminin et les parents qui ont choisi de recevoir des prestations parentales prolongées.

En 2020, en raison des modifications apportées au programme d'assurance-emploi pendant la pandémie de COVID-19, seuls les parents de enfants âgés de 18 mois ou moins ont été échantillonnés.

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. L'enquête permet aussi l'étude de l'accès aux prestations de maternité ou parentales.

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'Enquête sur la population active (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 examinée 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 des parents 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.

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.

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

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 personnelle, 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.

- c) devaient commencer un nouvel emploi dans quatre semaines ou moins à compter de la semaine de référence et étaient disponibles pour travailler.

Inactifs

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

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. Le codage des industries est basé sur le Système de classification des industries en Amérique du Nord 2012 (SCIAN de 2012). Les codes des professions sont basés sur la Classification nationale des professions 2016 (CNP de 2016).

Pour l'ECAE, le codage des industries est basé sur le Système de classification des industries en Amérique du Nord 2012 (SCIAN de 2012). Les codes des professions sont basés sur la Classification nationale des professions 2016 (CNP de 2016).

Semaine de référence

La semaine civile complète (du dimanche au samedi) couverte par l'enquête sur les forces de travail chaque mois. Il s'agit habituellement de la semaine contenant le 15^e jour du mois. La collecte des données est réalisée au cours de la semaine suivante, appelée semaine d'enquête, et la situation de la population active 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 deux derniers mois;
- 5) les parents des enfants en bas âge de moins de 18 mois travaillant pendant la semaine de référence. Le type de sous-population détermine souvent les questions posées.

Parents

Parents

Dans l'ECAE, le terme « parent » renvoie à la mère ou au père (par naissance ou adoption) d'un enfant âgé de moins de 18 mois à un an pendant la semaine de référence de l'EPA. De nombreuses parents ne faisaient pas partie de l'échantillon avant 2000. Les parents travaillant à temps plein et les parents 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 parent d'un enfant en bas âge durant la semaine de référence de l'enquête (voir la définition des « parents » ci-dessus).

Échantillon initial

Population ciblée par l'ECAE avant qu'elle ne soit élargie de façon à englober toutes les parents 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 parents travaillant à temps partiel),
- type 3 (à l'exception les parents qui n'ont pas travaillé en deux ans) et
- type 4 (y compris les parents 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 jusqu'à sept semaines après celles de l'EPA, la semaine de référence de l'enquête est la même que celle de l'EPA.

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 « parents », 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 travailleur.

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 5.5³. Les sections 5.6 et 5.7 décrivent comment l'ECAE s'est écartée du plan de base de l'EPA.

5.1 Population visée

L'Enquête sur la population active (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. L'enquête est menée dans l'ensemble du pays, tant dans les provinces que dans les territoires.

Sont exclus du champ de l'enquête les personnes qui vivent dans les réserves et dans d'autres peuplements autochtones des provinces, les membres à temps plein des Forces armées canadiennes, les pensionnaires d'établissements institutionnels et les ménages situés dans des régions extrêmement éloignées où la densité de population est très faible. Les personnes exclues de l'enquête représentent environ 2 % de la population de 15 ans et plus.

Les estimations de l'EPA pour l'ensemble du pays sont établies à partir des données de l'enquête recueillies dans les provinces. Les résultats territoriaux de l'EPA ne sont pas pris en compte dans les estimations nationales, mais sont publiés séparément.

5.2 Plan d'échantillonnage et taille de l'échantillon

Tous les 10 ans, après le recensement décennal de la population, l'EPA est remaniée pour tenir compte de l'évolution des caractéristiques démographiques et des nouvelles définitions des limites géographiques. De 2005 à 2014, le plan d'échantillonnage de l'EPA se fondait sur l'information recueillie lors du Recensement de 2001 et reflétait la taille de la population, la répartition provinciale et les limites infra provinciales de 2001.

Depuis cette période, on a assisté à une importante croissance démographique, à une évolution des caractéristiques de la population et du marché du travail, ainsi qu'à un changement des limites des municipalités et des régions métropolitaines de recensement (RMR). Janvier 2015 a vu l'introduction d'un nouveau plan d'échantillonnage fondé sur les données et les limites géographiques du Recensement de 2011.

Les strates du plan d'échantillonnage de l'EPA, qui répartissent la population de façon à rendre l'échantillonnage plus efficace, se veulent homogènes en ce qui concerne certaines variables clés du marché du travail. Un autre avantage de la stratification est qu'elle stabilise la taille de l'échantillon par région. Toutefois, les strates deviennent moins efficaces à mesure que l'on s'éloigne de l'année de base du plan d'échantillonnage (c.-à-d. 2001 pour le plan en place de 2005 à 2014) et que les caractéristiques de la population et du marché du travail évoluent dans le temps.

Le plan d'échantillonnage de 2015 définit les nouvelles strates en s'appuyant sur les données du dernier recensement (2011) et de l'Enquête nationale auprès des ménages (2011) ainsi que sur des données

³ Pour des renseignements plus complets sur la méthodologie de l'EPA, voir la publication *Méthodologie de l'Enquête sur la population active*, no 71-526-X au catalogue.

administratives. La grande majorité des strates visent à améliorer l'efficacité du plan d'échantillonnage. Quelques strates sont réservées afin de cibler des sous-populations particulières. Des strates de revenus élevés sont construites dans la plupart des RMR afin de regrouper les régions qui comptent plus de ménages à revenu élevé. Ces strates spéciales ne sont pas traitées différemment durant l'échantillonnage de l'EPA, mais elles sont disponibles pour d'autres enquêtes, comme l'Enquête sur les dépenses des ménages. De même, des strates d'Autochtones sont construites dans les régions de la Saskatchewan, de l'Alberta et de la Colombie-Britannique qui comptent une prévalence plus importante d'Autochtones. Ces strates ont été construites en réponse aux exigences de clients externes qui ont financé un échantillon additionnel de l'EPA afin de cibler des groupes particuliers et des populations autochtones dans ces provinces.

L'échantillon est réparti entre les provinces et, à l'intérieur des provinces, entre les strates, de façon à répondre le mieux possible au besoin d'estimations fiables aux différents niveaux géographiques. Des études préliminaires indiquaient que, vu la taille actuelle de l'échantillon, il n'était plus possible de suivre les lignes directrices en matière de fiabilité qui ont guidé la répartition de l'échantillon en 2005. Les lignes directrices pour le plan d'échantillonnage de 2015 diffèrent donc légèrement de celles suivies en 2005.

5.3 Sélection de logements

Un plan d'échantillonnage à deux degrés est utilisé pour toutes les provinces, sauf pour l'Île-du-Prince-Édouard. À l'intérieur de chaque strate, les « grappes » sont définies comme étant des petits secteurs d'environ 230 ménages. Au cours de la première étape de l'échantillonnage, un nombre de grappes, typiquement six, est sélectionné. Pour chaque grappe sélectionnée, une liste de ses logements est créée. Pour la plupart des grappes, la liste est extraite du Fichier de l'univers des logements qui est créé et tenu à jour par Statistique Canada. Pour le reste des grappes, un intervieweur utilise une carte de la grappe pour se rendre sur place et dresser une liste de tous les logements. Au cours de la deuxième étape de l'échantillonnage, un échantillon de logements est sélectionné à partir de ces listes.

Étant donné les caractéristiques particulières de l'Île-du-Prince-Édouard, l'échantillonnage est fait en une seule étape utilisant une liste complète d'adresses pour toutes les strates et en sélectionnant les logements à partir de cette liste.

5.4 Sélection des membres du ménage

Les renseignements démographiques sont recueillis sur tous les membres du ménage qui considèrent le logement sélectionné comme étant leur lieu de résidence habituel. Les renseignements sur le marché du travail sont recueillis pour tous les membres civils du ménage âgés de 15 ans ou plus. Pour les personnes de 70 ans et plus, le fardeau de réponse est réduit en réutilisant leurs réponses de l'interview initiale au cours des cinq mois suivant l'enquête.

5.5 Rotation de l'échantillon

L'EPA utilise un plan d'échantillonnage avec renouvellement de panel, selon 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 rend le plan d'échantillonnage efficace pour estimer les variations d'un mois à l'autre. La rotation après six mois évite d'imposer un fardeau trop lourd aux répondants des ménages sélectionnés aux fins de l'enquête.

5.6 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 parents 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.

Note : En raison de changements apportés aux opérations de Statistique Canada en réponse aux mesures de santé publique COVID-19, le cycle de collecte d'avril a été reporté à septembre pour l'année de référence 2020 et l'échantillon a été limité aux parents de nourrissons âgés de 18 mois ou moins.

5.6.1 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 deux derniers mois;
- 5) les parents de nourrissons âgés de 18 mois ou moins, qui ont travaillé pendant 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.

5.6.2 Type 4 : Un cas particulier

Seuls les travailleurs à temps plein (Type 4) qui ont connu une interruption de travail au cours des deux derniers mois sont visés par l'enquête. Comme cette information n'est pas disponible à partir de l'EPA, on sélectionne les personnes avec une courte durée d'emploi et on détermine s'il y a eu ou non interruption d'emploi au début de l'interview de l'ECAE. L'enquête se termine alors pour les personnes qui n'ont pas connu d'arrêt de travail. 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.

5.6.3 Autres exclusions

Un maximum de 2 personnes par ménages est choisi afin de réduire le fardeau de réponse pour le ménage, sauf des ménages avec trois personnes choisies, toutes en chômage (la limite ici est trois).

Quelques cas sont exclus avant la sélection de l'échantillon: des répondants avec un nom, numéro de téléphone ou nom d'employeur invalide, des cas imputés de l'EPA, des personnes qui ne donnent pas la permission à Statistique Canada de les appeler, et des répondants âgés de 70 ans et plus.

5.7 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
	2020
Terre-Neuve-et-Labrador	31
Île-du-Prince-Édouard	41
Nouvelle-Écosse	65
Nouveau-Brunswick	52
Québec	289
Ontario	440
Manitoba	176
Saskatchewan	140
Alberta	218
Colombie-Britannique	152
Canada	1604

6.0 Collecte des données

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

La collecte des données aux fins de l'EPA a lieu tous les mois durant la période de 10 jours qui suit la semaine de référence de l'EPA. La semaine de référence correspond habituellement à la semaine au cours de laquelle tombe le 15^e jour du mois.

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.

Les interviews de l'EPA sont réalisées par téléphone par des intervieweurs travaillant dans des centres ITAO (Interview téléphonique assistée par ordinateur) d'un bureau régional ou par des visites personnelles d'un intervieweur sur place. Les visites personnelles d'un enquêteur de terrain ont été suspendues après mars 2020 en raison de la pandémie de Covid-19. Ils recueillent premièrement des renseignements sociodémographiques sur chaque membre du ménage, puis 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 canadiennes. Interviews suivantes sont réalisées par téléphone. Lors des interviews mensuelles subséquentes, l'intervieweur fait confirmer les renseignements sociodémographiques obtenus le premier mois, puis il recueille les renseignements relatifs à la population active pour le mois courant. Depuis 2015, les répondants ont aussi l'option de terminer l'enquête en ligne pour des interviews suivantes.

Pour un logement donné, 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 personne du ménage. 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.

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 vérifier régulièrement le déroulement des interviews. Les intervieweurs principaux sont eux-mêmes supervisés par les gestionnaires du programme de l'EPA.

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

Le taux de non-réponse à l'EPA s'établit le plus souvent à environ 10 % des ménages admissibles. Les intervieweurs ont pour consigne de faire tous les efforts raisonnables pour mener les interviews avec les membres des ménages admissibles. Lorsqu'une personne commence par refuser 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. Suit un deuxième appel ou une deuxième visite de l'intervieweur. Dans les cas où la visite ou l'appel de l'intervieweur tombe à un mauvais moment, on fixe un rendez-vous à un moment qui conviendra mieux. Quand il n'y a personne au logement, l'intervieweur

rappelle à de nombreuses reprises. Quelles que soient les circonstances, on ne remplace jamais un logement sélectionné par un autre logement en cas de non-réponse.

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. Une formule de pondération s'applique pour tenir compte des ménages non répondants.

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 dans les sept semaines suivant la dernière interview de l'EPA. Les ménages sont contactés par lettre, et ont la possibilité de remplir le questionnaire en ligne. Le suivi des non-réponses par des enquêteurs téléphoniques commence deux semaines après le début de la collecte en ligne. 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 reçu, 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 au moyen d'un réseau protégé, 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 vérifier ou 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 5.6.2). Un petit nombre d'autres enregistrements sont retirés après qu'a 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.

Un critère est défini pour supprimer les enregistrements de non-réponse. Dans le cadre de l'ECAE, le répondant doit avoir au moins répondu à une question requise pour établir la variable COV de la couverture par le régime d'assurance emploi (voir la section 7.5.4).

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 éléments de données du questionnaire ont été enregistrés dans un format ouvert. Dans l'EICS, le processus de codage attribue des codes standard aux descriptions de l'industrie et de la profession fournies par les répondants (Système de classification des industries de l'Amérique du Nord (SCIAN 2017), et Classification nationale des professions (CNP 2016)) et au pays de naissance. En outre, les champs "Autre, précisez" comportant un nombre important de réponses textuelles ont été examinés et codés selon les catégories existantes. Dans certains cas, de nouvelles catégories ont été créées pour faciliter l'analyse des informations textuelles. Il s'agissait d'éléments relatifs aux raisons de ne pas demander ou recevoir des prestations, à la branche d'activité, à la profession ou à la raison de ne pas retourner au travail.

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 du dictionnaire de données du fichier principal de l'enquête.

Il n'y a eu aucune imputation pour l'Enquête sur la couverture de l'assurance-emploi pour 2020.

7.5 Création de variables dérivées

Beaucoup d'éléments de données du fichier de microdonnées ont été calculés en combinant des éléments du 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.

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 dictionnaire de données associé au fichier de microdonnées comprend une note désignant tous les éléments du questionnaire qui ont été utilisés pour créer chacune des variables dérivées.

7.5.1 Regroupement des éléments de données continus

La plupart des éléments de données collectés en tant que variables continues ne figurent dans le fichier de microdonnées à usage public que sous forme de variables groupées. L'âge du répondant (AGECAT), l'ancienneté dans l'emploi (TENURE_G) et le nombre de semaines travaillées pendant l'année de référence (WEEKSCAT) en sont des exemples.

Dans d'autres situations, les éléments de réponse catégoriels ont été regroupés pour créer des catégories significatives ou pour réduire le risque d'identifier des individus avec des ensembles de réponses uniques. C'est le cas pour le plus haut niveau de scolarité atteint (EDUC), la branche d'activité et la profession (SCIAN6 et CNP6), le mode de garde d'enfants prévu ou actuel (CHLDCARE), le régime de prestations parentales ou l'option choisie (BENPLAN - parents seulement), le type de famille économique (EFAMILY) et quelques autres.

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

Les questions relatives à l'employeur et aux conditions d'emploi n'ont été posées dans l'EICS que si l'information n'était pas disponible dans l'Enquête sur les Forces de Travail (EFT). Dans l'EFT, ces questions portent sur l'emploi actuel ou, pour certains éléments, sur l'emploi précédent s'il a été occupé l'année précédente. L'EICS recherche cette information pour tous les répondants qui ont travaillé au cours des deux années précédentes. En général, on a utilisé le nom de variable utilisé dans le fichier de microdonnées de l'EPA (FTPT, HRLYEARN). Un grand nombre de ces variables liées à l'emploi ont été regroupées pour le fichier de microdonnées à usage public de l'EICS.

7.5.3 Regroupement de deux questions différentes ou plus

Des variables telles que la situation syndicale (UNIONCA), le type de régime de travail (WRKTYP), la demande de prestations d'assurance-emploi au cours des 12 derniers mois ou depuis le dernier mois de travail (CLAIM), les prestations d'assurance-emploi (BENEFIT), la raison pour laquelle le travailleur n'a pas reçu ou demandé de prestations d'assurance-emploi pour la semaine de référence ou depuis sa naissance ou son adoption (RNBENRW), les paiements supplémentaires reçus de l'employeur (ADDPAYM) et la recherche d'un emploi dans la collectivité ou la province (LOOKOUT) sont dérivées à l'aide de plus d'un élément du questionnaire.

Dans ces cas, l'algorithme utilisé pour créer la nouvelle variable est généralement assez intuitif. Par exemple, la variable sur le type de régime de travail est créée en combinant le statut à temps plein ou à

temps partiel, le statut d'emploi permanent ou temporaire et la raison de l'emploi temporaire et la catégorie de travailleur comme suit :

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, travailleur familial 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.

7.5.4 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 Emploi et Développement social Canada (EDSC). 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 EDSC 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 lors de la période sans emploi (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).
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 inférieur 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 de l'estimation dans un échantillon probabiliste tel que l'EFT est que chaque personne de l'échantillon "représente", outre elle-même, plusieurs autres personnes qui ne font pas partie de l'échantillon. Par exemple, dans un échantillon aléatoire simple de 2% de la population, chaque personne de l'échantillon représente 50 personnes de la population.

La phase de pondération est une étape qui permet de calculer, pour chaque enregistrement, quel est ce nombre. Ce poids apparaît dans le fichier de microdonnées et doit être utilisé pour obtenir des estimations significatives à partir de l'enquête. Par exemple, si le nombre de parents recevant des prestations de maternité ou parentales doit être estimé, on le fait en sélectionnant les enregistrements

se rapportant aux individus de l'échantillon présentant cette caractéristique et en additionnant les poids inscrits sur ces enregistrements.

Les détails de la méthode utilisée pour calculer ces poids sont présentés au chapitre 11.0.

7.7 Suppression de renseignements confidentiels

La loi interdit à Statistique Canada de publier des données qui divulgueraient des renseignements obtenus en vertu de la Loi sur la statistique et se rapportant à une personne, une entreprise ou un organisme identifiable, sans que cette personne, cette entreprise ou cet organisme en soit informé au préalable ou y consente par écrit. Diverses règles de confidentialité sont appliquées à toutes les données qui sont diffusées ou publiées afin d'empêcher la publication ou la divulgation de toute information jugée confidentielle. Si nécessaire, les données sont supprimées pour empêcher la divulgation directe ou résiduelle de données identifiées.

Il est à noter que les fichiers de microdonnées à " usage public " (FMUP) peuvent différer des fichiers " maîtres " de l'enquête détenus par Statistique Canada. Ces différences sont habituellement le résultat de mesures prises pour protéger l'anonymat des répondants individuels à l'enquête. Les mesures les plus courantes sont la suppression d'éléments de données et le regroupement de valeurs dans des catégories plus larges. Pour certaines variables susceptibles d'identifier des personnes, le FMGD peut avoir été traité avec une suppression locale, c'est-à-dire que certaines des valeurs du fichier maître peuvent avoir été codées comme " non déclarées " dans le FMGD.

Le fichier maître de l'enquête comprend des identificateurs géographiques pour les 10 provinces et pour les régions économiques de l'AE. Le FMGD ne contient pas d'identificateurs géographiques au-dessous du niveau provincial et certaines provinces ont été regroupées (c.-à-d. la région de l'Atlantique et le Manitoba avec la Saskatchewan). Le regroupement des provinces a été effectué afin d'éviter la suppression excessive de données sur des variables utiles.

Le fichier principal de l'enquête comprend l'âge précis du répondant alors que le FMGD ne contient que les groupes d'âge. De même, les variables non groupées telles que la branche d'activité et la profession, l'ancienneté de l'emploi, le nombre de mois depuis le dernier emploi et l'âge du bébé en mois (parents seulement) ne sont disponibles que dans le fichier principal de l'enquête.

Les utilisateurs de l'ESFP qui souhaitent avoir accès à des informations exclues des fichiers de microdonnées peuvent acheter des tabulations personnalisées. Les estimations produites seront communiquées à l'utilisateur, sous réserve de respecter les lignes directrices relatives à l'analyse et à la diffusion énoncées au chapitre 9.0 du présent document.

8.0 Qualité des données

8.1 Taux de réponse

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

Province	2020		
	Échantillon admissible	Réponse	Taux de réponse (%)
Terre-Neuve-et-Labrador	31	18	58
Île-du-Prince-Édouard	40	20	50
Nouvelle-Écosse	64	44	69
Nouveau-Brunswick	51	29	57
Québec	286	193	67
Ontario	426	288	68
Manitoba	167	112	67
Saskatchewan	135	96	71
Alberta	214	148	69
Colombie-Britannique	144	92	64
Canada	1558	1040	67

Note: 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 5.6.2 et 7.2).

8.2 Erreurs relatives à l'enquête

Les estimations calculées à partir de cette enquête reposent sur un sous-échantillon d'individus tirés de l'Enquête sur la population active. 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 un supplément à l'Enquête sur les forces de travail (EFT), la base de sondage utilisée était l'échantillon de l'EFT. Toute non-réponse à l'EFT a eu un impact sur la base de sondage de l'EICS. La qualité des variables d'échantillonnage de la base de sondage était très élevée. L'échantillon de l'EICS 2020 était composé de deux groupes de rotation issus de l'EFT pour la population " parents ".

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 plutôt 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 consistait à lire le manuel de l'intervieweur EICS, à s'exercer avec les cas de formation EICS sur l'ordinateur et à discuter de toute question avec les intervieweurs principaux avant le début de l'enquête. Une description du contexte et des objectifs de l'enquête était fournie, ainsi qu'un glossaire et une série de questions et réponses. Les intervieweurs ont commencé à recueillir les données de l'EICS deux semaines après la fin des périodes de collecte de l'EPA de mars, juin, octobre et décembre, soit en juillet, septembre, novembre et janvier respectivement. La collecte a duré environ cinq semaines pour chaque cycle de l'EICS.

En 2020, en raison des arrêts liés à COVID-19 en mars et avril, l'EICS a utilisé les groupes de rotation qui ont terminé leurs six mois dans l'EFT en juin, août, octobre et décembre.

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 pouvaient être jumelés aux enregistrements correspondants de l'EPA. Aucun enregistrement n'était 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 ayant le plus de renseignements 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. La répartition de chaque variable dérivée a été comparée à celles des éléments du questionnaire utilisés pour la créer. Une comparaison de la répartition sur la période allant de 2018 à 2019 a également été effectuée pour assurer la comparabilité chronologique.

En raison des changements apportés à la population de l'enquête et au questionnaire (section 5.6), il convient d'être prudent lors des comparaisons avec les données des années d'enquête précédentes.

8.2.4 Erreur de réponse à la question BK_Q05A

Des problèmes de réponse à la question BK_Q05A ("Avez-vous pris une pause d'une semaine ou plus [depuis la naissance ou l'adoption de votre enfant] ?") ont été identifiés chez les parents d'enfants âgés de 12 à 18 mois. Certains de ces parents qui ont indiqué qu'ils n'avaient pas pris d'arrêt de travail ont également déclaré qu'ils avaient reçu des prestations de maternité et/ou parentales et qu'ils n'avaient pas travaillé pendant qu'ils recevaient ces prestations. Comme BK_Q05A est utilisé pour aider à déterminer la durée du congé des parents et comme condition de flux pour le bloc "Travail après la

naissance (WA)", il y a eu une augmentation artificielle du nombre de parents ayant une durée de congé de 0 semaine et une diminution du nombre de réponses reçues pour le bloc WA.

Pour maintenir la cohérence des données de l'enquête, les réponses et les variables dérivées de la durée du congé (ALLEAVE, LEAVECAT) n'ont pas été ajustées ; cependant, il est recommandé d'utiliser le nombre de semaines de prestations reçues comme approximation de la durée du congé pour ces répondants.

En SAS, le code suivant peut être utilisé pour créer une variable dérivée pour la durée du congé (DURLEAVE) dans le fichier maître :

/ Si le répondant est un parent avec un enfant âgé de 12 mois ou plus*

qui n'a pas pris de congé et n'a pas travaillé tout en recevant des prestations,

la durée du congé est le nombre de semaines de prestations perçues./*

```
if PARENT = 1 and CHLDMTHS >= 12 and BK_05A = 2 and BE_75 = 2
    then DURLEAVE = BENWEEKS;
else DURLEAVE = ALLEAVE;
```

8.2.5 Non-réponse

L'une des principales sources d'erreurs non dues à l'échantillonnage dans les enquêtes est l'effet de la non-réponse sur les résultats de l'enquête. L'ampleur de la non-réponse varie de la non-réponse partielle (absence de réponse à une ou plusieurs questions) à la non-réponse totale.

La non-réponse totale se produit parce que l'enquêteur n'a pas pu contacter le répondant ou que celui-ci a refusé de participer à l'enquête.

La non-réponse totale a été traitée en ajustant le poids des individus qui ont répondu à l'enquête pour compenser ceux qui n'ont pas répondu. Dans la plupart des cas, il y a eu non-réponse partielle au questionnaire d'enquête lorsque le répondant n'a pas compris ou a mal interprété une question, a refusé d'y répondre ou ne pouvait se rappeler l'information demandée.

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

8.2.6 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.

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 champ du poids par ces progiciels.

9.3 Définitions de types d'estimations : catégoriques et quantitatives

Avant de discuter de la façon dont on peut totaliser et analyser les données de l'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 de congé pris après la naissance ou l'adoption d'un enfant. Le numérateur est une estimation du nombre total de mois de congé pris par tous les parents pour lesquels l'information est disponible (retour au travail déjà effectué ou projets connus) et son dénominateur est le nombre de parents prenant un congé d'une durée 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:

- a) additionnant les poids finals des enregistrements présentant la caractéristique qui nous intéresse pour le numérateur (\hat{X}),
- b) additionnant les poids finals des enregistrements présentant la caractéristique qui nous intéresse pour le dénominateur (\hat{Y}), puis en
- c) 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 d'intérêt par le poids final pour chaque enregistrement, puis en additionnant cette quantité sur tous les enregistrements d'intérêt. Par exemple, pour obtenir une estimation du nombre total de semaines de prestations d'assurance-emploi (AE) reçues par les parents d'un nourrisson qui sont déjà retournés au travail, il faut multiplier la valeur déclarée dans la variable dérivée BENWEEKS (semaines de prestations d'AE) par le poids final de l'enregistrement, puis additionner cette valeur à tous les enregistrements ayant PARENT = 1 et WORKNOW = 1 (parents d'un nourrisson âgé de 18 mois ou moins qui travaillent 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, le nombre moyen de semaines pour lesquelles les parents 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 des parents travaillant actuellement (\hat{Y}) incluses dans cette catégorie en additionnant les poids finals de tous les enregistrements où les variables PARENT = 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.

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 des répliques, plus particulièrement la méthode bootstrap. Cette méthode, qui se fonde sur une technique de réplification 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.

L'utilisateur du FMGD de l'ECAE peut également calculer ses propres estimations de la variance à l'aide d'un ensemble de poids bootstrap associé généré de manière à préserver la confidentialité. Ces poids bootstrap du FMGD peuvent être utilisés pour calculer des estimations de la variance pour les totaux, les proportions et les ratios, ainsi que pour des paramètres utilisés dans d'autres techniques d'analyse telles que la régression linéaire, la régression logistique et l'analyse de la variance. Voir la section 10.0 pour une description détaillée et des exemples.

En raison de la réduction de la taille de l'échantillon en 2020, les poids bootstrap utilisés pour produire les estimations officielles et les poids bootstrap accompagnant le FMGD sont tous deux générés à l'aide de la méthode bootstrap généralisée, qui ne nécessite qu'un ajustement aléatoire généré à partir d'une

distribution de probabilité. Traditionnellement, les poids bootstrap utilisés pour produire les estimations officielles du SCEI ont été générés à l'aide de la méthode de remise à l'échelle de Rao Wu Yue.

9.5 Lignes directrices pour la diffusion de coefficients de variation

Avant de diffuser et/ou de publier une estimation provenant du SCEI, les utilisateurs doivent d'abord déterminer le niveau de qualité de l'estimation. Les niveaux de qualité sont : acceptable, marginal et inacceptable. La qualité des données est affectée à la fois par les erreurs d'échantillonnage et les erreurs non dues à l'échantillonnage, comme l'indique le chapitre 8.0. Cependant, pour cet objectif, le niveau de qualité d'une estimation ne sera déterminé que sur la base de l'erreur d'échantillonnage telle que reflétée par le coefficient de variation, comme indiqué dans le tableau ci-dessous. Néanmoins, les utilisateurs doivent s'assurer de lire le chapitre 8.0 pour mieux connaître les caractéristiques de qualité de ces données.

Tout d'abord, il faut déterminer le nombre de répondants qui contribuent au calcul de l'estimation. Si ce nombre est inférieur à 5, l'estimation pondérée doit être considérée comme étant de qualité inacceptable.

Pour les estimations pondérées basées sur des tailles d'échantillon de 5 ou plus, les utilisateurs doivent déterminer le coefficient de variation de l'estimation et suivre les directives ci-dessous. Ces directives de niveau de qualité doivent être appliquées aux estimations pondérées arrondies.

Toutes les estimations peuvent être considérées comme publiables. Cependant, celles dont le niveau de qualité est marginal ou inacceptable doivent être accompagnées d'un avertissement pour mettre en garde les utilisateurs ultérieurs.

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 5 ou plus, et présentent de faibles coefficients de variation, de l'ordre de 0,0 à 15 %.</p> <p>Aucune mise en garde n'est requise.</p>
2) Médiocre	<p>Les estimations proviennent d'une taille d'échantillon de 5 ou plus, et présentent des coefficients de variation élevés, de l'ordre de 15 à 35 %.</p> <p>Ces estimations devraient être accompagnées de la mise en garde suivante :</p> <p>« À utiliser avec prudence »</p>

<p>3) Inacceptable</p>	<p>Les estimations proviennent d'une taille d'échantillon inférieure à 5, ou présentent des coefficients de variation très élevés, supérieurs à 35%.</p> <p>Statistique Canada recommande de ne pas diffuser d'estimations de qualité inacceptable. Ces estimations devraient être accompagnées de la mise en garde suivante :</p> <p>« Trop peu fiable pour être publié »</p>
------------------------	--

10.0 Pondération bootstrap dans les FMGD

Le fichier de poids bootstrap des fichiers de microdonnées à grande diffusion (FMGD) fourni avec les FMGD de l'ECAE contient 1 000 ensembles de poids bootstrap enregistrés en colonnes. Les colonnes sont intitulées $WRPMB$, où $b = 1, \dots, 1000$. Un ensemble de poids bootstrap est appelé répétition, de sorte que chaque colonne représente une répétition. En général, l'estimation de la variance à l'aide des poids bootstrap s'effectue en calculant l'estimation d'un paramètre pour chaque ensemble de poids bootstrap, puis en calculant la variance des estimations obtenues. Cette méthode est non seulement valide pour les totaux, les moyennes et les rapports, mais également pour les paramètres non linéaires comme les médianes et d'autres quantiles. Ces poids bootstrap peuvent être utilisés pour calculer une variété d'estimations de la variance fondées sur les données de l'ECAE. Dans la présente section, des exemples sont fournis pour illustrer les utilisations potentielles des poids bootstrap.

Il convient de noter que les poids bootstrap des FMGD sont générés afin de respecter la confidentialité. Par conséquent, alors que les résultats des estimations utilisant les poids bootstrap des FMGD sont comparables à ceux utilisant les poids bootstrap officiels, ils ne sont pas identiques et sont généralement moins précis. Des estimations de la variance pour des variables données à l'aide des poids bootstrap officiels peuvent être obtenues auprès de Statistique Canada moyennant un recouvrement des coûts.

10.1 Comment utiliser des poids bootstrap pour des estimations de variables qualitatives

En général, pour utiliser des poids bootstrap en vue d'estimer la variance d'une estimation d'un paramètre, un fichier d'analyse contenant les variables d'intérêt et toute variable de domaine en plus des poids bootstrap est d'abord utilisé. Il est possible de dériver un tel fichier en fusionnant, par exemple, le FMGD et le fichier de poids bootstrap. Une fois le fichier d'analyse créé, les étapes suivantes peuvent être utilisées pour calculer la variance bootstrap :

1. Calculer l'estimation de l'enquête, \hat{X} , à l'aide de la variable d'intérêt et des poids de sondage finaux.
2. Calculer l'estimation bootstrap, $\hat{X}^{*(b)}$, pour chaque répétition bootstrap b , à l'aide de la variable d'intérêt et des poids bootstrap de chaque répétition. Puisque l'ECAE fournit 1 000 répétitions, cette étape devrait fournir les 1 000 estimations bootstrap, $\hat{X}^{*(1)}, \dots, \hat{X}^{*(1000)}$.

3. Calculer la variance bootstrap des 1 000 estimations bootstrap à l'aide de la formule :

$$\widehat{\text{var}}(\hat{X}) = \sum_{b=1}^{1000} (\hat{X}^{*(b)} - \hat{X})^2 / 1000, \quad (1)$$

où \hat{X} est l'estimation du paramètre calculée à l'aide des poids finaux et $\hat{X}^{*(b)}$ est l'estimation bootstrap pour la répétition b . Si des estimations sont requises par domaine, tel que par province ou tranche d'âge, la procédure susmentionnée doit être appliquée séparément pour chaque niveau de domaine souhaité.

4. Calculer le coefficient de variation (CV) estimé \widehat{cv} avec l'équation :

$$\widehat{cv}(\hat{X}) = \frac{\sqrt{\widehat{\text{var}}(\hat{X})}}{\hat{X}} \quad (2)$$

Les règles qui suivent devraient permettre à l'utilisateur d'estimer les coefficients de variation à l'aide des poids bootstrap pour des estimations du nombre, de la proportion ou du pourcentage de membres de la population visée par l'enquête présentant une certaine caractéristique et pour des rapports et des différences entre de telles estimations. Dans toutes les règles, la notation suivante est utilisée :

- n La taille d'échantillon de la sous-population ou du domaine analysé.
- b L'indice des répétitions bootstrap; dans l'ECAE, $b = 1, \dots, 1000$.
- \hat{X} L'estimation du paramètre calculée à partir des poids finaux.
- $\hat{X}^{*(b)}$ L'estimation du paramètre b -th calculé à partir de la répétition b -th.
- w_i Le poids final de l'individu i -th.
- $w_i^{*(b)}$ Le poids bootstrap pour l'individu i -th pour la répétition b -th.
- I_i La fonction pour l'individu i -th, où $I_i = 1$ si l'individu présente la caractéristique d'intérêt et 0 dans le cas contraire.

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

L'estimation du nombre de personnes présentant une caractéristique est calculée avec l'équation :

$$\hat{X} = \sum_{i=1}^n w_i I_i. \quad (3)$$

L'estimation bootstrap du nombre de personnes présentant une caractéristique est calculée avec l'équation :

$$\hat{X}^{*(b)} = \sum_{i=1}^n w_i^{*(b)} I_i, \quad (4)$$

pour $b = 1, \dots, 1000$. La variance de \hat{X} est alors calculée comme dans l'équation (1).

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

L'estimation de la proportion de personnes présentant une caractéristique est calculée avec l'équation :

$$\hat{X} = \frac{\sum_{i=1}^n w_i I_i}{\sum_{i=1}^n w_i} \quad (5)$$

L'estimation bootstrap du nombre de personnes présentant une caractéristique est calculée avec l'équation :

$$\hat{X}^{*(b)} = \frac{\sum_{i=1}^n w_i^{*(b)} I_i}{\sum_{i=1}^n w_i^{*(b)}}, \quad (6)$$

pour $b = 1, \dots, 1000$. La variance de $\hat{\theta}$ est alors calculée comme dans l'équation (1).

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

L'estimation de la différence entre deux totaux est calculée avec l'équation :

$$\hat{d} = \hat{X}_1 - \hat{X}_2,$$

où \hat{X}_1 et \hat{X}_2 sont calculés comme dans l'équation (3) ou (5). L'estimation bootstrap de la différence entre deux totaux est calculée avec l'équation :

$$\hat{d}^{*(b)} = \hat{X}_1^{*(b)} - \hat{X}_2^{*(b)},$$

où $\hat{X}_1^{*(b)}$ et $\hat{X}_2^{*(b)}$ sont calculés comme dans l'équation (4) ou (6). La variance de \hat{d} est alors calculée comme dans l'équation (1).

Règle 4: Estimations de rapports

Cette règle est une généralisation de la règle 2, où le numérateur et le dénominateur peuvent être des quantités arbitraires. Dans le cas où le numérateur est un sous-ensemble du dénominateur, la règle 2 peut être appliquée; au cas où, par exemple, 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 par rapport au nombre de chômeurs qui reçoivent tout autre type de prestations, l'estimation du rapport peut être calculée avec l'équation :

$$\hat{R} = \frac{\hat{X}_1}{\hat{X}_2}, \quad (7)$$

où \hat{X}_1 et \hat{X}_2 sont calculés comme dans l'équation (3) ou (5).

L'estimation bootstrap de ce rapport est calculée avec l'équation :

$$\hat{R}^{*(b)} = \frac{\hat{X}_1^{*(b)}}{\hat{X}_2^{*(b)}}, \quad (8)$$

où $\hat{X}_1^{*(b)}$ et $\hat{X}_2^{*(b)}$ sont calculés comme dans l'équation (4) ou (6). La variance de \hat{R} est alors calculée comme dans l'équation (1).

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.

L'estimation de la différence entre des rapports est calculée avec l'équation :

$$\hat{\delta} = \hat{R}_1 - \hat{R}_2, \quad (8)$$

où \hat{R}_1 et \hat{R}_2 sont calculés comme dans l'équation (7).

Les estimations bootstrap de la différence de rapports sont calculées avec l'équation :

$$\hat{\delta}^{*(b)} = \hat{R}_1^{*(b)} - \hat{R}_2^{*(b)}, \quad (9)$$

où $\hat{R}_1^{*(b)}$ et $\hat{R}_2^{*(b)}$ sont calculés comme dans l'équation (8). La variance de $\hat{\delta}$ est alors calculée à l'aide de l'équation (1).

10.1.1 Exemples d'utilisation de poids bootstrap pour des estimations de variables qualitatives

Le calcul d'estimations de variance à l'aide de poids bootstrap est grandement facilité par l'utilisation d'un progiciel statistique, en particulier si celui-ci peut traiter des méthodes de réplcation. Les progiciels SAS et Stata ont, par exemple, la capacité d'estimer les variances à partir de poids bootstrap (Gagné et coll., 2014)⁴. De plus, Statistique Canada fournit un ensemble de macros SAS intitulées Bootvar pouvant calculer des variances pour de nombreux types de statistiques. Bootvar peut être demandé gratuitement auprès de Statistique Canada. Les exemples suivants sont inclus pour aider les utilisateurs à appliquer les règles précédentes. Les calculs ont été obtenus à l'aide de SAS 9.3.

Il convient de noter que ces exemples font référence à des estimations associées à la population cible typique de l'enquête et ne sont pas directement applicables à la population cible de 2020. Les méthodes restent toutefois applicables.

1^{er} exemple : Estimations du nombre de personnes présentant une caractéristique donnée (agrégats)

⁴ Gagné, C., G. Roberts, L. Keown. Estimation pondérée et estimation de la variance bootstrap pour analyser des données d'enquête : Comment les effectuer dans certains logiciels choisis? *Le Bulletin technique et d'information des Centres de données de recherche*, vol. 6 n° 1, p. 5-70.

Supposons qu'un utilisateur estime que 353 000 chômeurs ont reçu des prestations régulières d'assurance-emploi au cours de la semaine de référence. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

À l'aide des poids bootstrap, un utilisateur calculerait 1 000 estimations de personnes ayant reçu des prestations d'assurance-emploi régulières en utilisant les 1 000 ensembles de poids bootstrap. Appliquer les équations (1) et (2) à ces 1 000 estimations fournit un CV estimé de 0,0223.

2^e exemple : Estimations des proportions ou des pourcentages de personnes présentant une caractéristique donnée

Supposons qu'un utilisateur estime que $511\,700 / 606\,700 = 84,3\%$ 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?

- 1) Calculer 1 000 estimations de chômeurs admissibles à l'assurance-emploi et 1 000 estimations de chômeurs potentiellement admissibles à l'assurance-emploi à l'aide de poids bootstrap.
- 2) En utilisant les estimations de l'étape précédente, calculer les 1 000 estimations de la proportion de chômeurs potentiellement admissibles à l'assurance-emploi qui étaient admissibles à l'assurance-emploi.
- 3) Appliquer les équations (1) et (2) à ces 1 000 estimations fournit un CV estimé de 0,0206.
- 4) Le coefficient de variation estimé de cette estimation est donc 2,1 %. Le résultat selon lequel 84,3 % des chômeurs potentiellement admissibles à l'assurance-emploi étaient admissibles à l'assurance-emploi peut alors être publié sans réserve.

3^e exemple : Estimations de différences entre des agrégats ou des pourcentages

Supposons qu'un utilisateur estime que $163\,900 / 253\,300 = 64,7\%$ de la population ordinaire de chômeurs au Québec a cotisé à l'assurance-emploi, tandis que $250\,100 / 434\,700 = 57,5\%$ de la population ordinaire de chômeurs 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?

- 1) Pour chaque province, deux ensembles de 1 000 estimations doivent être calculés à l'aide des poids bootstrap : (i), l'estimation du nombre de membres de la population ordinaire de chômeurs et (ii), l'estimation du nombre de membres de la population ordinaire de chômeurs ayant cotisé à l'assurance-emploi.
- 2) Pour chaque province, calculer les 1 000 estimations de la proportion de la population ordinaire de chômeurs ayant cotisé à l'assurance-emploi à l'aide des estimations de l'étape (1).
- 3) En utilisant les deux ensembles de 1 000 estimations de l'étape (e), calculer les 1 000 estimations de la différence entre la proportion de la population ordinaire de chômeurs au Québec ayant cotisé à l'assurance-emploi et la proportion de la population ordinaire de chômeurs en Ontario ayant cotisé à l'assurance-emploi.

- 4) Appliquer les équations (1) et (2) à ces 1 000 estimations de différence fournit un CV estimé de 0,634.
- 5) Le coefficient de variation estimé de la différence entre ces estimations est donc de 63,4 %. Cette différence entre les estimations est jugée inacceptable, mais peut découler de la faible différence estimée des proportions de 0.072. Dans cette situation, une analyse plus poussée, comme le calcul d'intervalles de confiance, peut être utile.

4^e exemple : Estimations de rapports

Supposons qu'un utilisateur estime que 163 900 membres de la population ordinaire de chômeurs au Québec ont cotisé à l'assurance-emploi, par rapport à 253 100 membres de la population ordinaire de chômeurs en Ontario ayant cotisé à l'assurance-emploi. L'utilisateur souhaite comparer l'estimation pour le Québec par rapport à l'estimation pour 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 de rapports, où le numérateur de l'estimation (\hat{X}_1) est le nombre de chômeurs au Québec ayant cotisé à l'assurance-emploi; le dénominateur de l'estimation (\hat{X}_2) est le nombre de chômeurs en Ontario ayant cotisé à l'assurance-emploi.
- 2) Pour chaque province, calculer les 1 000 estimations du nombre de membres de la population ordinaire de chômeurs ayant cotisé à l'assurance-emploi.
- 3) À l'aide des résultats de l'étape (2), calculer les 1 000 estimations du rapport du nombre estimé pour le Québec au nombre estimé pour l'Ontario.
- 4) Appliquer les équations (1) et (2) à ces 1 000 estimations fournit un CV estimé de 0,075.

Le coefficient de variation estimé de l'estimation du rapport est 7,5 %; l'estimation est donc publiable sans réserve.

5^e exemple : 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 de chômeurs ayant cotisé à l'assurance-emploi et les personnes âgées de 25 à 44 ans de la population ordinaire de chômeurs ayant cotisé à l'assurance-emploi équivaut à 0,50 au Manitoba et en Saskatchewan et à 0,62 en Alberta. L'utilisateur souhaite comparer ces deux rapports pour savoir s'il existe une différence statistique entre eux. Comment l'utilisateur détermine-t-il le coefficient de variation de cette différence?

- 1) Calculer d'abord le rapport du Manitoba et de la Saskatchewan (\hat{R}_1) et celui de l'Alberta (\hat{R}_2) à l'aide de l'équation (7) et des poids finaux, ainsi que 1 000 rapports pour chaque région (respectivement $\hat{R}_1^{*(b)}$ et $\hat{R}_2^{*(b)}$ pour $b = 1, \dots, 1000$) à l'aide de l'équation (8) et des poids bootstrap.

- 2) À l'aide des résultats de l'étape (1), calculer les différences estimées $\hat{d} = \hat{R}_1 - \hat{R}_2$ et $\hat{d}^{*(b)} = \hat{R}_1^{*(b)} - \hat{R}_2^{*(b)}$, pour $b = 1, \dots, 1000$.
- 3) Utiliser les résultats de l'étape (2) et appliquer les équations (1) et (2) à ces 1 000 estimations fournit un coefficient de variation estimé de 2,409.
- 4) Le coefficient de variation estimé de la différence entre ces estimations est donc de 240,9 %. La différence entre les estimations est jugée inacceptable et Statistique Canada recommande de ne pas la publier. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée en utilisant la lettre F (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les futurs utilisateurs en garde contre les taux élevés d'erreur associés à cette estimation.

10.2 Comment utiliser des poids bootstrap 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 d'une estimation peuvent être calculés directement en déterminant d'abord l'erreur-type estimée, $\widehat{se}(\hat{X}) = \sqrt{\widehat{var}(\hat{X})}$, de l'estimation \hat{X} , puis en utilisant la formule suivante pour la convertir en intervalle de confiance $CI(\hat{X})$:

$$CI(\hat{X}) = (\hat{X} - t \cdot \widehat{se}(\hat{X}), \hat{X} + t \cdot \widehat{se}(\hat{X}))$$

où t est le point limite approximatif de la distribution normale. Selon le niveau de confiance, le tableau suivant peut être utilisé :

t	Niveau
1,0	si l'on désire un intervalle de confiance de 68 %
1,6	si l'on désire un intervalle de confiance de 90 %
2,0	si l'on désire un intervalle de confiance de 95 %
2,6	si l'on désire un intervalle de confiance de 99 %

Une autre façon de calculer des intervalles de confiance à l'aide de poids bootstrap est de déterminer les centiles de la distribution des estimations correspondant au niveau de confiance souhaité. Pour ce faire, trier d'abord les estimations provenant des 1 000 répétitions bootstrap en ordre croissant. Supposons que $\hat{X}_{(p)}$ est le centile p -th estimé de X . Alors, $\hat{X}_{(p)}$ est la valeur correspondant à l'estimation p -th des données triées. Pour obtenir les limites de confiance à l'aide de cette méthode, il suffit de calculer $\hat{X}_{(p)}$ et $\hat{X}_{(100-p)}$, les centiles p -th et $(100 - p)$ -th des répétitions bootstrap. L'intervalle de confiance obtenu est :

$$CI(\hat{X}) = (\hat{X}_{(p)}, \hat{X}_{(100-p)}).$$

Note: 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 de l'écart type pour obtenir des limites de confiance

À l'aide du tableau ci-dessus, un intervalle de confiance de 95 % pour la proportion estimée de chômeurs potentiellement admissibles à l'assurance-emploi qui étaient admissibles à l'assurance-emploi (d'après l'exemple 2 de la section 10.1.1) serait calculé comme suit :

$$\begin{aligned} \hat{X} &= 0.843 \\ t &= 2 \\ \widehat{se}(\hat{X}) &= 0.0173 \\ CI(\hat{X}) &= (0.843 - 2 \cdot 0.0173, 0.843 + 2 \cdot 0.0173) \\ &= (0.8084, 0.8776) \end{aligned}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 80,8 % et 87,8 % des chômeurs potentiellement admissibles à l'assurance-emploi étaient admissibles à l'assurance-emploi.

10.2.2 Exemple d'utilisation des poids bootstrap pour obtenir des limites de confiance

En utilisant les centiles bootstrap, l'intervalle de confiance de 95 % serait calculé comme suit :

$$\begin{aligned} \hat{X}_{(2.5)} &= 0.8111 \\ \hat{X}_{(97.5)} &= 0.8794 \\ CI(\hat{X}) &= (0.8111, 0.8794) \end{aligned}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 81,1 % et 87,9 % des chômeurs potentiellement admissibles à l'assurance-emploi étaient admissibles à l'assurance-emploi.

10.3 Comment utiliser l'erreur-type bootstrap pour effectuer un test-t

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

Supposons que \hat{X}_1 et \hat{X}_2 sont des estimations d'un échantillon pour deux caractéristiques d'intérêt. Supposons également que l'erreur-type de la différence $\hat{d} = \hat{X}_1 - \hat{X}_2$ est $se(\hat{d})$.

Si $t = \frac{\hat{X}_1 - \hat{X}_2}{se(\hat{d})}$ 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, en revanche, 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'existe pas de différence entre la proportion de la population ordinaire de chômeurs au Québec ayant cotisé à l'assurance-emploi et la proportion de la population ordinaire de chômeurs en Ontario ayant cotisé à l'assurance-emploi. D'après l'exemple 3 de la section 10.1.1, il s'est avéré que l'erreur-type de la différence entre ces deux estimations était 0,0455. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{se(\hat{d})} = \frac{0.647 - 0.575}{0.0455} = \frac{0.072}{0.0455} = 1.58$$

Puisque $t = 1,58$ est inférieur à 2, il faut en conclure qu'il n'existe pas de différence significative entre les deux estimations au niveau de signification de 0,05.

10.4 Coefficients de variation pour des estimations quantitatives

Pour des estimations quantitatives, les règles de la section 10.1 peuvent être adaptées pour déterminer l'erreur d'échantillonnage. Pour déterminer, par exemple, le coefficient de variation des gains horaires moyens habituels, on peut adapter la règle 2 en remplaçant la variable indicatrice I_i par une variable représentant les gains horaires habituels. Par exemple, si $z_i =$ gains horaires, adapter l'équation (5) permet d'obtenir :

$$\hat{X} = \frac{\sum_{i=1}^n w_i z_i}{\sum_{i=1}^n w_i},$$

une estimation de gains horaires moyens habituels. Voir également la section 9.3.4.

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'attaque 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. Les réponses des répondants ayant refusé la question sur le couplage des données ont été considérées comme des non-réponses.

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 5.6.1).
- 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. Emploi 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. En raison du changement de la population cible en 2020, cette étape n'a pas été effectuée sur les pondérations de 2020.

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.