



## **Guide de l'utilisateur des microdonnées**

### **Enquête sur les changements à l'égard de l'emploi**

#### **Cohortes 1 à 10**

**Janvier 1995 à septembre 1998**



Statistique  
Canada

Statistics  
Canada

**Canada**



## Table des matières

<b>1.0</b>	<b>Introduction</b>	<b>5</b>
<b>2.0</b>	<b>Contexte</b>	<b>7</b>
<b>3.0</b>	<b>Objectifs</b>	<b>9</b>
<b>4.0</b>	<b>Concepts et définitions</b>	<b>11</b>
<b>5.0</b>	<b>Méthodologie de l'enquête</b>	<b>13</b>
5.1	Population visée	13
5.2	Plan de sondage	13
5.3	Taille de l'échantillon selon la province	15
<b>6.0</b>	<b>Collecte des données</b>	<b>17</b>
<b>7.0</b>	<b>Traitement des données</b>	<b>19</b>
7.1	Saisie des données	19
7.2	Vérification	19
7.3	Codage des questions ouvertes	20
7.4	Création de variables dérivées	21
7.5	Pondération	21
7.6	Suppression de renseignements confidentiels	21
<b>8.0</b>	<b>Qualité des données</b>	<b>23</b>
8.1	Taux de réponse	23
8.2	Erreurs relatives à l'enquête	24
8.2.1	Base de sondage	24
8.2.2	Collecte des données	25
8.2.3	Traitement des données	25
8.2.4	Non-réponse	25
8.2.5	Mesure de l'erreur d'échantillonnage	25
<b>9.0</b>	<b>Lignes directrices pour la totalisation, l'analyse et la diffusion de données</b>	<b>27</b>
9.1	Lignes directrices pour l'arrondissement d'estimations	27
9.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation	28
9.3	Estimations catégoriques	28
9.3.1	Totalisation d'estimations catégoriques	28
9.4	Lignes directrices pour l'analyse statistique	29
9.5	Lignes directrices pour la diffusion de coefficients de variation	30
<b>10.0</b>	<b>Calcul de la variance</b>	<b>33</b>
10.1	Importance du calcul de la variance	33
10.2	Calcul de la variance et du coefficient de variation	33
10.3	Comment utiliser le Module d'extraction de coefficients de variation pour des estimations catégoriques	33
10.3.1	Exemples d'utilisation du Module d'extraction de coefficients de variation pour des estimations catégoriques	35
10.4	Comment utiliser le Module d'extraction de coefficients de variation pour obtenir des limites de confiance	37
10.4.1	Exemple d'utilisation du Module d'extraction de coefficients de variation pour obtenir des limites de confiance	38
10.5	Comment utiliser le Module d'extraction de coefficients de variation pour effectuer un test t	39

10.5.1	Exemple d'utilisation le Module d'extraction de coefficients de variation pour effectuer un test t .....	39
10.6	Le Module d'extraction de coefficients de variation .....	40
<b>11.0</b>	<b>Pondération</b> .....	<b>41</b>
<b>12.0</b>	<b>Questionnaires</b> .....	<b>43</b>
<b>13.0</b>	<b>Cliché d'enregistrement à valeurs univariées</b> .....	<b>45</b>

## 1.0 Introduction

L'Enquête sur les changements à l'égard de l'emploi (ECEE) a été menée par Statistique Canada avec l'appui et la collaboration de Développement des Ressources humaines Canada (DRHC), maintenant appelé Ressources humaines et Développement social Canada. Cette enquête, menée auprès des personnes ayant pu subir un changement d'emploi, s'inspire d'une enquête précédente, à savoir l'Enquête canadienne par panel sur l'interruption de l'emploi (ECPIE) mais avec des modifications apportées au plan de sondage et au contenu afin de permettre une meilleure compréhension de cette population.

Ce manuel a été produit pour faciliter la manipulation du fichier de microdonnées portant sur les résultats de l'enquête. Toutes les questions concernant l'ensemble de données ou son utilisation devraient être adressées à :

### Statistique Canada

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

### Ressources humaines et Développement social Canada

Roger Hubley  
Unité de politique et de planification stratégique  
Évaluations A.-E.  
Téléphone : (819) 953-8503  
Télécopieur : (819) 953-8868  
Courriel : [roger.hubley@hrsdc-rhdcc.gc.ca](mailto:roger.hubley@hrsdc-rhdcc.gc.ca)



## **2.0 Contexte**

En 1993, le secteur de Macro-évaluation de Développement des ressources humaines Canada (DRHC) a mené, pour la première fois, une enquête par panel auprès des sans-emploi. L'Enquête canadienne par panel sur l'interruption de l'emploi de 1993 (ECPIE) a été conçue pour évaluer l'incidence du projet de loi C-113 sur certains aspects du Programme d'emploi et d'assurance (a.-e.) anciennement désigné assurance-chômage (a.-c.) et à recueillir des renseignements sur des sujets tels la recherche d'emploi, les actifs et les dettes, et l'utilisation de l'aide sociale. Deux cohortes ont été interviewées à trois reprises pendant les douze mois suivant la cessation d'emploi. L'ECPIE de 1995 a poursuivi ce travail tout en évaluant l'incidence du projet de loi C-17, entérinée le 31 mai 1994. Cette fois, deux cohortes ont été interviewées à deux reprises pendant les douze mois suivant la cessation d'emploi.

Le contenu de l'Enquête sur les changements à l'égard de l'emploi (ECEE) ressemble à celui des deux ECPIE antérieures, à la différence près qu'elle tient compte de la totalité de la population récemment sans-emploi; c'est-à-dire qu'on n'a pas admis d'exclusions basées sur la raison pour laquelle le Relevé d'emploi (RE) a été produit, et qu'on a inclut toutes les dates de fin d'emploi. La nouvelle enquête a pour but de contrôler et d'évaluer les effets du projet de loi C-12, qui a été adopté en partie en juillet 1996 et dont le reste est entré en vigueur en janvier 1997. L'enquête couvre les Relevés d'emploi de la période de douze mois précédant le mois de juillet 1996, la période de six mois entre la première et la deuxième série de modifications et la période de douze mois suivant l'entrée en vigueur du projet de loi C-12.





### **3.0 Objectifs**

L'objectif principal de l'Enquête sur les changements à l'égard de l'emploi (ECEE) est d'évaluer les effets du projet de loi C-12 modifiant le régime d'assurance-emploi, et de mesurer jusqu'à quel point ses objectifs ont été atteints. Ils étaient de récompenser l'effort à l'emploi, d'assurer l'octroi d'avantages adéquats en visant les personnes dans le plus grand besoin, de stimuler la création d'emploi, et de réhausser la perception du grand public de l'équité du régime d'assurance-emploi. L'enquête a touché à certains aspects de ces objectifs; par ailleurs, elle vise à mesurer l'impact de l'ensemble des mesures législatives.

L'enquête a pour objectifs secondaires de poursuivre la collecte des données de l'Enquête canadienne par panel sur l'interruption de l'emploi de 1993 et de 1995, dont les données démographiques des personnes et des ménages, les renseignements sur la recherche d'emploi et son résultat, les actifs et les dettes, les dépenses, et l'utilisation de l'assurance-emploi et de l'aide sociale.



## **4.0 Concepts et définitions**

Ce chapitre donne un aperçu des concepts et des définitions d'intérêt pour les utilisateurs. Les utilisateurs sont priés de se reporter au chapitre 12.0 de ce document où figure une copie des questionnaires d'enquête réellement employés.

### **Emploi**

On ne considère aux fins de cette enquête que les « travailleurs rémunérés » où l'employé avait travaillé pendant au moins une semaine ou plus.

### **Industrie et profession**

L'Enquête sur les changements à l'égard de l'emploi fournit des renseignements sur la profession et l'industrie des personnes qui occupent un emploi. Ces statistiques pour les fichiers de microdonnées s'appuient sur la Classification type des professions de 1980 et sur la Classification type des industries de 1980.

### **Emploi à plein temps**

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

### **Emploi à temps partiel**

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

### **Assurance-emploi**

Le Régime de l'assurance-emploi (a.-e.), autrefois connu sous le nom de Régime de l'assurance-chômage (a.-c.), est un programme du gouvernement fédéral qui fournit un revenu aux personnes qui viennent de subir une cessation d'emploi. Les conditions d'admissibilité et le montant des prestations varient selon la province dans laquelle est domiciliée le demandeur, le nombre de semaines de travail à son actif et son taux de traitement, entre autres.

### **Assistance sociale**

Les programmes d'assistance sociale, dont le Bien-être, sont des programmes municipaux qui fournissent un revenu aux ménages sans autre source de revenu. Les conditions d'admissibilité et le montant de prestations dépendent des circonstances des membres du ménage. Seule une personne par ménage a le droit de toucher des prestations de Bien-être, qui sont destinées à l'usage de tout le ménage, non de la personne seulement.

### **Relevé d'emploi**

Le Relevé d'emploi (RE) constitue l'un des formulaires par lesquelles le gouvernement fédéral administre le programme d'Assurance-emploi. Pour effectuer une demande de prestations, le demandeur doit être muni d'un Relevé d'emploi dûment rempli par son ancien employeur. Le Relevé contient les renseignements de base servant à déterminer l'admissibilité du demandeur et le montant de prestations auquel il ou elle a droit d'après les règlements sur a.-e. Lorsque l'employeur remplit le Relevé il en remet une copie à l'ancien employé et une autre à Développement des ressources humaines Canada.

### **Cessation d'emploi**

Une cessation d'emploi veut dire toute raison pour laquelle un travailleur aurait quitté un emploi (voir la définition d'« Emploi »), résultant en la remise d'un Relevé d'emploi. Ces raisons incluent les démissions, les renvois, les mises à pied, les congés non rémunérés (par exemple, les congés de maladie prolongés, les congés de maternité etc.), les retraites, et les périodes d'inactivités dues à un conflit industriel. Le Relevé d'emploi exige que l'employeur coche l'une des cases expliquant la cessation, fournies à cet effet. Si la raison de cessation n'est pas offerte ou si l'employeur n'est pas certain de celle-ci, il a le choix de cocher la case « Autre ».



## **5.0 Méthodologie de l'enquête**

### **5.1 Population visée**

L'échantillon de l'Enquête sur les changements à l'égard de l'emploi (ECEE) était composé d'individus qui :

- étaient représentatif de la population canadienne civile non institutionnalisée âgée de 15 et plus, résidant dans l'une des 10 provinces ou territoires du Canada;
- avaient vécu au moins une cessation d'emploi entre le 1 juillet 1995 et le 31 décembre 1997; et
- avaient reçu un Relevé d'emploi (RE) résultant de cette cessation d'emploi, quelle que soit la raison de la cessation d'emploi.

Cette période d'échantillonnage de 30 mois a été divisée en trimestres, au cours de chacun des trimestres de référence un échantillon distinct a été sélectionné.

Le fichier de la base de données des RE provenant de Développement des ressources humaines Canada (DRHC) a été utilisé comme base de sondage pour l'ECEE. Ce fichier est constamment mis à jour au fur et à mesure que de nouveaux relevés soient reçus. Comme certains employeurs tardent à envoyer les formulaires à DRHC, il se peut que la base ne soit pas complète au moment de la sélection de l'échantillon trimestriel. Les noms et les adresses provenant du Système national des services d'emploi (SNSE) de DRHC complètent cette base de sondage.

### **5.2 Plan de sondage**

Les enregistrements de RE ont été dédoublés en utilisant le numéro d'assurance sociale (NAS) de l'individu. Pour le premier trimestre de référence de l'ECEE (1<sup>er</sup> juillet 1995 au 30 septembre 1995) les enregistrements de RE ont aussi été classés en deux strates de population (« nouvelle » population cible) :

- ceux qui avaient reçu seulement un RE pendant le trimestre de référence; et
- ceux qui avaient reçu plus d'un RE au cours du trimestre de référence.

Pour les personnes ayant reçu plus d'un RE au cours du trimestre de référence, le RE ayant la rémunération assurable et des semaines de travail les plus élevées a été retenu comme étant le RE de référence dans cette strate de population.

Pour tous les trimestres de référence subséquents de l'ECEE (1<sup>er</sup> octobre 1995 au 31 décembre 1997) les enregistrements de RE ont été classés en strates de population supplémentaires (la population cible de « chevauchement ») :

- ceux qui avaient reçu au moins un RE au cours d'un trimestre de référence antérieur, remontant seulement qu'au 1<sup>er</sup> juillet 1995.

Les fichiers, au sein de chacune des trois strates de population, ont été triés selon la province et la raison de cessation d'emploi. Pour des fins d'échantillonnage, les Territoires du Nord-Ouest ont été regroupés avec le Québec, et le Yukon avec la Colombie-Britannique.

Les raisons de cessation d'emploi ont été regroupées en trois strates, tel que démontré dans le tableau ci-dessous.

Codes	Raisons pour cessation d'emploi
A	Pénurie de travail
K	Autre
C	Retour aux études
E	Départ volontaire
F	Retraite
G	Grossesse
L	A atteint 65 ans
N	Permis d'absence
B	Conflit industriel
D	Accident ou maladie
H	Travail partagé
J	Formation en apprentissage
M	Renvoi

Les échantillons de chaque province étaient de taille satisfaisante à produire des estimations approximativement de qualité comparable. Les tailles d'échantillons étaient initialement déterminées en utilisant les effets de plan et les taux de réponse de l'Enquête canadienne par panel sur l'interruption de l'emploi de 1995 (ECPIE) et ensuite à partir des résultats des deux premiers trimestres de référence de l'ECEE. Des mesures ont été prises afin d'assurer que les personnes sélectionnées pour participer à l'ECPIE de 1995 ou à d'autres enquêtes menées récemment par DRHC n'auraient pas à être interviewés au cours des deux prochaines années.

On a alors choisi les échantillons des populations cibles « nouvelles » et de « chevauchement », attribuées proportionnellement aux strates de plan, à l'aide d'échantillonnage aléatoire simple. L'« échantillon de chevauchement » ainsi sélectionné a ensuite été apparié aux échantillons antérieurs et les doubles ont été rejetés de l'« échantillon de chevauchement ». Pour la population cible de « chevauchement », l'échantillon a été augmenté par le facteur suivant :

$$\text{Facteur} = \frac{\text{Population cible de chevauchement}}{\text{Population cible de chevauchement} - \text{Doubles d'échantillons antérieurs}}$$

Puisqu'il était nécessaire de produire des estimations fiables semestriellement, on a doublé la taille de la base de sondage du 1<sup>er</sup> trimestre pour calculer la taille de l'échantillon, qui a ensuite été divisée par deux. Après avoir tiré cette moitié des deux échantillons semestriels, on a recalculé la taille de l'échantillon requise pour produire des estimations semestrielles en combinant les tailles des bases de sondage des deux trimestres de référence. On a ensuite soustrait la taille du 1<sup>er</sup> échantillon de cette valeur afin d'obtenir la taille de l'échantillon requise pour le 2<sup>e</sup> trimestre de référence.

Des estimations fiables annuelles pour 13 communautés sélectionnées étaient requises. Les limites géographiques de ces communautés ont été définies selon les codes postaux fournis par DRHC. À la suite de la première interview les répondants de l'enquête ont été reclassifiés en nouvelles communautés selon leurs réponses à la question portant sur le lieu de domicile pendant le mois précédant celui où le RE a été émis.

À noter que la procédure suivante a été suivie afin de sélectionner l'échantillon pour la communauté et l'échantillon provincial :

- d'abord, l'échantillon pour la communauté a été tiré; et
- ensuite, l'échantillon provincial a été sélectionné et à été attribué au prorata aux chiffres de la population cible de la province moins la base de la communauté.

### 5.3 Taille de l'échantillon selon la province

Le tableau qui suit montre le nombre de personnes incluses dans l'échantillon de l'ECEE pour chacune des combinaisons de cohortes selon la région. Ce tableau comprend les personnes qui étaient non répondantes à l'ECEE.

Provinces et régions	Taille de l'échantillon				
	Cohortes 1 et 2	Cohortes 3 et 4	Cohortes 5 et 6	Cohortes 7 et 8	Cohortes 9 et 10
Provinces de l'Atlantique	4 154	5 429	4 711	4 675	5 007
Québec	1 584	2 449	2 249	2 159	1 968
Ontario	1 724	2 619	2 575	2 355	1 981
Ouest	7 188	9 225	8 734	8 850	8 385
<b>Canada</b>	<b>14 650</b>	<b>19 722</b>	<b>18 269</b>	<b>18 039</b>	<b>17 341</b>





## 6.0 Collecte des données

Toutes les interviews ont été menées à partir du bureau central de Statistique Canada à Ottawa à l'aide du système d'interviews téléphoniques assistées par ordinateur (ITAO). Pour toutes les cohortes, la collecte des données s'est effectuée en deux phases, ou vagues. Seuls les répondants de la première vague ont été rejoints lors des interview de la deuxième vague. Les dates pour les 10 cohortes sont indiquées dans le tableau ci-dessous.

Chaque répondant a reçu une lettre au sujet de l'enquête à peu près deux semaines avant la période de collecte pour la cohorte à laquelle il faisait partie. La lettre expliquait les objectifs de l'enquête, fournissait un numéro d'appel sans frais à l'intention des répondants qui auraient quelques questions au sujet de l'enquête, et avisait ceux-ci qu'un intervieweur les appellerait dans l'avenir proche.

Juste avant la deuxième vague de la période de collecte pour chaque cohorte, une deuxième lettre a été envoyée aux répondants qui avaient subi l'interview de la première vague et qui avaient donné leur accord quant au partage de données avec Développement des ressources humaines Canada (DRHC). Celle-ci les rappelait des objectifs de l'enquête et les informait qu'un intervieweur les appellerait bientôt pour effectuer une interview de suivi.

Tous les intervieweurs ont suivi une séance de formation ayant pour but de les familiariser avec les objectifs et les concepts de l'enquête, avec le questionnaire ITAO, ainsi que des procédures de bases de l'interview téléphonique. Les intervieweurs ont tenté de communiquer avec tous les répondants dans la langue officielle de leur choix.

La participation à l'enquête était facultative. Lorsqu'un répondant refusait de fournir une partie ou la totalité de l'information demandée, l'intervieweur devait alors contacter à nouveau le répondant pour tenter d'obtenir l'information désirée. Lorsque le répondant était temporairement absent ou que l'interview ne pouvait être réalisée en raison de difficultés d'ordre linguistique ou autre, l'intervieweur devait rappeler le répondant à une date ultérieure. Les réponses par personne interposée n'étaient pas admises.

### Périodes d'interviews indiquant les dates de périodes de cessation d'emploi des Relevés d'emploi (RE) et les périodes de référence de collecte pour les cohortes 1 à 10

Cohortes	Dates de périodes de cessation d'emploi des RE	Interviews de la 1 <sup>re</sup> vague	Période de référence - Interviews de la 1 <sup>re</sup> vague	Interviews de la 2 <sup>e</sup> vague	Période de référence - Interviews de la 2 <sup>e</sup> vague
1	Juill. à sept. 1995	Juill. 1996	Janv. 1995 à juill. 1996	Mars 1997	Juill. 1996 à mars 1997
2	Oct. à déc. 1995	Sept. 1996	Avril 1995 à sept. 1996	Juin 1997	Oct. 1996 à juin 1997
3	Janv. à mars 1996	Nov. 1996	Juill. 1995 à nov. 1996	Oct. 1997	Nov. 1996 à oct. 1997
4	Avril à juin 1996	Févr. 1997	Oct. 1995 à févr. 1997	Déc. 1997	Févr. 1997 à déc. 1997
5	Juill. à sept. 1996	Mai 1997	Janv. 1996 à mai 1997	Mars 1998	Juin 1997 à mars 1998
6	Oct. à déc. 1996	Sept. 1997	Avril 1996 à sept. 1997	Juin 1998	Sept. 1997 à juin 1998
7	Janv. à mars 1997	Nov. 1997	Juill. 1996 à nov. 1997	Oct. 1998	Nov. 1997 à oct. 1998
8	Avril à juin 1997	Févr. 1998	Oct. 1996 à févr. 1998	Déc. 1998	Févr. 1998 à déc. 1998
9	Juill. à sept. 1997	Mai 1998	Janv. 1997 à mai 1998	Mars 1999	Mai 1998 à mars 1999
10	Oct. à déc. 1997	Sept. 1998	Avril 1997 à sept. 1998	Juin 1999	Sept. 1998 à juin 1999



## 7.0 Traitement des données

Le principal produit de l'Enquête sur les changements à l'égard de l'emploi (ECEE) 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'un questionnaire utilisant les techniques d'une interview téléphonique assistée par ordinateur (ITAO). L'emploi d'un questionnaire ITAO réduit les délais et coûts de traitement associés à la saisie des données, aux erreurs de transcription et à la transmission des données.

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

### 7.2 Vérification

La première étape du traitement d'enquête a consisté à remplacer les codes de non-réponse par un ensemble de codes standardisés utilisés par la Division des enquêtes spéciales. Il était permis de répondre « Refus » ou « Ne sait pas » à la plupart des questions si le répondant était incapable de donner une estimation ou s'il était incapable ou refusait de répondre à la question.

La deuxième étape du traitement d'enquête a consisté à attribuer des codes particuliers aux questions sans réponse pour indiquer si le répondant devait répondre ou non à la question (se référer au tableau ci-dessous pour les codes). Le code « Enchaînement valide » est attribué si la personne ne fait pas partie de la population visée par cette question. Le code « Non déclaré » est attribué dans les cas suivants :

- a. Il n'est pas certain que la personne fait partie de la population visée par la question.
- b. La personne fait partie de la population visée par la question mais n'a pas répondu à cette dernière.

On est en présence d'une situation de type **a.** « Non déclaré » quand un répondant répond « Ne sait pas » ou « Refus » à une question servant à déterminer si l'on doit poser une question subséquente. Pour assurer le bon déroulement de l'interview, on sautera habituellement cette question subséquente étant donné qu'on ne sait pas si elle s'applique ou non. On est en présence d'une situation de type **b.** « Non déclaré » quand une partie de la population visée ne répond pas à une question à cause d'une erreur de spécification ou de programmation associée à l'application de la collecte de données.

Types de non-réponse	Codes de non-réponse	Ce que cela implique :
Enchaînement valide	6, 96, 996, etc.	D'après les réponses aux questions précédentes, cette question ne s'applique pas au répondant
Ne sait pas	7, 97, 997, etc.	La question a été posée au répondant qui a répondu « Ne sait pas ».
Refus	8, 98, 998, etc.	La question a été posée au répondant qui a répondu « Refus ».
Non déclaré	9, 99, 999, etc.	La question n'a pas été posée au répondant parce qu'il a répondu « Ne sait pas » ou « Refus » à une question précédente qui déterminait si cette question s'appliquait ou non. On ne sait pas si le répondant faisait partie de la population visée par cette question.

Veillez noter que les valeurs particulières des codes de non-réponse sont fonction de la longueur de la variable. Par exemple, on attribue le code « 9 » à un élément qui a été codé comme « Non déclaré ». Le code « 99 » est attribué quand il s'agit d'une variable à deux chiffres et « 999 », d'une variable à trois chiffres, etc.

### 7.3 Codage des questions ouvertes

Un certain nombre de données élémentaires ont été consignées sur le questionnaire ITAO sous forme de questions ouvertes. Quelques-unes de celles-ci ont été recodées par la suite.

Pour obtenir les données sur l'industrie et la profession correspondant à chaque emploi occupé par le répondant pendant les périodes de référence, on a recueilli les renseignements suivants à l'aide du questionnaire : nom de l'employeur, genre d'entreprise, d'industrie ou de service de l'employeur, genre de travail effectué et tâches ou responsabilités habituelles du répondant à un poste précis. Ces données ont été utilisées pour attribuer les codes d'industrie à chaque emploi à l'aide de la version 1980 de la Classification type des industries (CTI) de Statistique Canada. Les codes de profession ont été attribués à l'aide de la version 1980 de la Classification type des professions (CTP) de Statistique Canada. Les renseignements sur le nom de l'employeur, etc. ont d'abord été traités par le système de codage automatisé conçu pour l'Enquête sur la population active. Quand on obtenait un appariement avec la base de données des descriptions, on a attribué le code CTI 1980 pertinent et/ou le code CTP 1980 pertinent. Les enregistrements restants partiellement codés ou non codés ont été codés manuellement par le personnel des codeurs d'expérience. Tous les enregistrements codés ont été liés au fichier des données d'enquête.

De même, les réponses aux questions portant sur l'origine ethnique ou les minorités visibles ont été enregistrées par les intervieweurs pendant les entrevues puis traitées par un système de codage automatisé. Les enregistrements restants partiellement codés ou non codés ont été codés manuellement par le personnel des codeurs d'expérience. Tous les enregistrements codés ont été liés au fichier de données d'enquête.

## 7.4 Création de variables dérivées

Un certain nombre de données élémentaires incluses dans les fichiers de microdonnées ont été calculées ou regroupées en combinant des postes sur le questionnaire pour faciliter l'analyse des données. Ces variables sont identifiées par la lettre « D » ou « G » au troisième caractère du nom de la variable. Les utilisateurs devraient consulter les cinq fichiers qui renferment les clichés d'enregistrement à valeurs univariées, cités au chapitre 13.0 de ce guide, pour plus de détails.

## 7.5 Pondération

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

La phase de la pondération est une étape où l'on calcule ce nombre (ou poids) pour chaque enregistrement. Ce poids, qui figure dans le fichier de microdonnées à grande diffusion, **doit** servir à calculer des estimations significatives à partir de l'enquête. Si, par exemple, le nombre d'individus qui travaillaient à un emploi permanent avant leur changement d'emploi, doit être estimé, cette opération s'effectue en sélectionnant les enregistrements se référant aux individus incluses à l'intérieur de l'échantillon qui présentent cette caractéristique et en additionnant les poids inscrits dans ces enregistrements.

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

## 7.6 Suppression de renseignements confidentiels

Il convient de souligner que les fichiers de microdonnées à « grande diffusion » (FMGD) peuvent différer des fichiers « maîtres » de l'enquête que conserve Statistique Canada. Ces différences sont habituellement le résultat de mesures prises pour protéger l'anonymat des répondants à une enquête. Voici une liste des mesures prises pour protéger l'anonymat des répondants de l'ECEE.

- Réduction des identificateurs géographiques – Tout renseignement renvoyant à l'adresse du répondant ou à l'adresse de son employeur a été exclu du FMGD. Les fichiers de microdonnées à grande diffusion ne contiennent pas d'indicateurs géographiques sous le niveau de la région.
- Enlèvement des identificateurs directs – Tout renseignement renvoyant directement au nom d'un répondant a été exclu du FMGD.
- Enlèvement des variables contenant des renseignements confidentiels – Toute variable contenant des renseignements de nature personnelle ou confidentielle qui auraient aidé à identifier les répondants à l'enquête a été enlevée.
- Enlèvement de l'information en double – Quand une variable du fichier maître contenait la même information qu'une autre variable du même fichier, une seule de ces variables a été conservée dans le FMGD. On s'est efforcé de conserver autant de variables qui ont été dérivées des données obtenues à l'aide des questions originales d'enquête pendant le traitement.
- Réduction de nombreuses variables relatives aux ménages – Le fichier maître de

l'enquête contient plusieurs variables pour chaque membre du ménage du répondant. Cette information a été réduite en dérivant de nouvelles variables pour les remplacer dans le FMGD.

- Groupement des catégories de variables – Les catégories de variables ont été groupées dans le but de créer une série de données plus concentrée sur le fichier de microdonnées à grande diffusion. Cette démarche a également aidé à limiter le volume de renseignements pouvant être obtenus au sujet d'un répondant particulier pour l'identifier. Certaines des variables groupées sont les suivantes : âge du répondant, situation de famille, composition du ménage et niveau d'études le plus élevé. De plus, les variables industrie et profession ont été groupées pour contenir seulement les principaux groupes de classification.
- Plafonnement des variables – Les variables quantitatives ont été plafonnées. Un grand nombre des variables catégoriques qui ont été groupées ont également été plafonnées.
- Suppression de données – Il y avait certains agencements rares de variables dans l'échantillon et la population. Pour réduire le risque d'identification du répondant à cause de ces caractéristiques rares, on a supprimé certaines valeurs de variables.

## 8.0 Qualité des données

### 8.1 Taux de réponse

Les fichiers de microdonnées à grande diffusion de l'Enquête sur les changements à l'égard de l'emploi contiennent seulement les répondants qui ont rempli une partie minimale du questionnaire et qui ont également consenti à ce que l'on partage leurs réponses avec Développement des ressources humaines Canada (DRHC), Évaluation stratégique et suivi du rendement. Cependant, les répondants qui n'ont pas rempli la partie minimale du questionnaire ont été inclus dans la stratégie de pondération. Comme tel, il est important de connaître le taux de réponse à chacune des périodes de référence semestrielles de l'ECEE. Le tableau ci-dessous indique les taux de réponse.

Provinces et régions	Cohortes 1 et 2			Cohortes 3 et 4		
	Nombre échantillonné	Nombre de répondants	Taux de réponse (%)	Nombre échantillonné	Nombre de répondants	Taux de réponse (%)
Provinces de l'Atlantique	4 154	2 210	53	5 429	3 021	56
Québec	1 584	752	47	2 449	1 172	48
Ontario	1 724	758	44	2 619	1 179	45
Ouest	7 188	3 010	42	9 225	4 115	45
<b>Canada</b>	<b>14 650</b>	<b>6 730</b>	<b>46</b>	<b>19 722</b>	<b>9 487</b>	<b>48</b>

Provinces et régions	Cohortes 5 et 6			Cohortes 7 et 8		
	Nombre échantillonné	Nombre de répondants	Taux de réponse (%)	Nombre échantillonné	Nombre de répondants	Taux de réponse (%)
Provinces de l'Atlantique	4 711	2 757	59	4 675	2 378	51
Québec	2 249	1 190	53	2 159	1 027	48
Ontario	2 575	1 256	49	2 355	1 007	43
Ouest	8 734	3 968	45	8 850	3 301	37
<b>Canada</b>	<b>18 269</b>	<b>9 171</b>	<b>50</b>	<b>18 039</b>	<b>7 713</b>	<b>43</b>

Provinces et régions	Cohortes 9 et 10		
	Nombre échantillonné	Nombre de répondants	Taux de réponse (%)
Provinces de l'Atlantique	5 007	2 874	57
Québec	1 968	1 105	56
Ontario	1 981	996	50
Ouest	8 385	3 944	47
<b>Canada</b>	<b>17 341</b>	<b>8 919</b>	<b>51</b>

## 8.2 Erreurs relatives à l'enquête

Les estimations calculées à partir de cette enquête reposent sur un échantillon d'individus. 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

Le fichier administratif des Relevés d'emploi (RE) a été utilisé comme base d'échantillonnage. Ce fichier est mis à jour de façon continue au fur et à mesure que l'on reçoit des enregistrements RE. Comme certains employeurs peuvent tarder à présenter les formulaires à DRHC, il se peut que le cadre ne soit pas terminé au moment de la sélection trimestrielle de l'échantillon.

Les poids sont rajustés en fonction du dernier cadre à jour. Cependant, si les derniers déclarants ont des caractéristiques différentes de celles des premiers déclarants, les données pourraient être biaisées.

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



## **8.2.2 Collecte des données**

La formation des intervieweurs pour l'ECEE consistait à lire le Manuel de l'intervieweur, la pratique avec des cas simulés de l'ECEE sur ordinateur et une discussion avec l'intervieweur principal des questions de l'intervieweur avant le début de l'enquête. L'information fournie aux intervieweurs comprenait une description du contexte et des objectifs de l'enquête ainsi qu'un glossaire de terminologie et une série de questions et de réponses. Un tableau sommaire des Périodes d'interviews indiquant les dates de périodes de cessation d'emploi des Relevés d'emploi (RE) et les périodes de référence de collecte pour les cohortes 1 à 10 est inclus au chapitre 6.0 de ce document.

## **8.2.3 Traitement des données**

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

## **8.2.4 Non-réponse**

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

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

## **8.2.5 Mesure de l'erreur d'échantillonnage**

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

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

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

Par exemple, supposons que, d'après les résultats de l'enquête, l'on estime que 35,48 % des Canadiennes et des Canadiens qui avaient subi un changement d'emploi entre juillet et décembre 1995 avaient de l'assurance-maladie à leur emploi (variable E1QB23) et l'on constate que l'erreur-type de cette estimation est de 0,0095. Le coefficient de variation de l'estimation est donc calculé comme suit :

$$\left( \frac{0,0095}{0,3548} \right) \times 100 \% = 2,67 \%$$

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

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

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

### 9.1 Lignes directrices pour l'arrondissement d'estimations

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

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

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

Le plan d'échantillonnage utilisé pour l'Enquête sur les changements à l'égard de l'emploi (ECEE) n'était pas autopondéré. Lorsqu'ils produisent des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent appliquer le poids d'échantillonnage approprié.

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

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

## 9.3 Estimations catégoriques

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

Les estimations catégoriques sont des estimations du nombre ou du pourcentage de membres de la population visée par l'enquête possédant certaines caractéristiques ou faisant partie d'une catégorie définie. Le nombre de personnes qui fait du bénévolat ou du travail communautaire pour accroître leur possibilités d'emploi à un moment ou l'autre après la [date du Relevé d'emploi (date du RE)] ou la proportion de personnes qui le fait surtout à temps partiel constituent des exemples de telles estimations. Une estimation du nombre de personnes possédant une certaine caractéristique peut aussi être désignée une estimation d'un agrégat.

### Exemples de questions catégoriques :

- Q : Avez-vous fait du bénévolat ou du travail communautaire PRÉCISÉMENT pour accroître vos POSSIBILITÉS D'EMPLOI à un moment ou l'autre après le [date du RE]?
- R : Oui / Non
- Q : Ce (bénévolat/travail communautaire) était-il surtout à plein temps ou à temps partiel?
- R : Plein temps / Temps partiel

### 9.3.1 Totalisation d'estimations catégoriques

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

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

## 9.4 Lignes directrices pour l'analyse statistique

L'ECEE repose sur un plan d'échantillonnage complexe comportant une stratification, de multiples étapes de sélection ainsi que des probabilités inégales de sélection des répondants. L'utilisation des données provenant d'enquêtes aussi complexes présente des problèmes pour les analystes, parce que le plan d'enquête et les probabilités de sélection influent sur les procédures d'estimation et de calcul de la variance qui devraient être utilisées. Il faut utiliser les poids de l'enquête pour que les estimations et les analyses des données de l'enquête soient exemptes de biais.

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

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

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

- 1) sélectionner tous les répondants du fichier qui ont déclaré PIGF2 = homme;
- 2) calculer le poids MOYEN pour ces enregistrements en additionnant les poids originaux des personnes établis à partir du fichier de microdonnées pour ces enregistrements puis diviser cette somme par le nombre de répondants ayant déclaré PIGF2 = homme;
- 3) pour chacun de ces répondants, calculer un poids REMIS À L'ÉCHELLE égal au poids original de la personne divisé par le poids MOYEN;
- 4) effectuer l'analyse portant sur ces répondants en utilisant le poids REMIS À L'ÉCHELLE.

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

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

## **9.5 Lignes directrices pour la diffusion de coefficients de variation**

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

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

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

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

### Lignes directrices relatives au niveau de qualité de l'estimation

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





## 10.0 Calcul de la variance

L'Enquête sur les changements à l'égard de l'emploi (ECEE) est une enquête probabiliste, c'est-à-dire, qu'un échantillon a été sélectionné pour représenter la population cible. Une certaine variabilité est associée à la sélection aléatoire d'un échantillon. Cette variabilité est identifiée comme étant l'erreur d'échantillonnage tel qu'il a été décrit au chapitre 8.0. À cette erreur, s'ajoutent des corrections pour tenir compte de la non-réponse. La mesure de cette variabilité est connue comme étant la variance. Ce chapitre explique l'importance de calculer la variance et présente l'outil pour la calculer.

### 10.1 Importance du calcul de la variance

La variabilité ou variance d'une estimation est une bonne indication de la qualité d'une estimation. Une estimation avec une variance trop élevée est considérée comme non fiable. Afin de quantifier ce qui est une variance trop élevée, une mesure relative de la variabilité est utilisée, à savoir le coefficient de variation (CV). Le coefficient de variation est défini comme étant le ratio de la racine carrée de la variance sur l'estimation. La racine carrée de la variance est aussi connue sous le nom d'écart-type. L'utilisation du coefficient de variation plutôt que la variance permet à l'analyste de comparer sur une même échelle des estimations de magnitudes diverses. Ainsi, il est possible de quantifier la qualité de toute estimation avec le CV.

De plus, pour différents tests statistiques tels que des hypothèses sur la différence entre deux estimations, le calcul de la variance ou du CV est requis afin de déterminer si la différence est statistiquement différente ou non. Le calcul de la variance ou du CV est donc primordial.

### 10.2 Calcul de la variance et du coefficient de variation

L'échantillon de l'ECEE a été sélectionné de façon à pouvoir utiliser les outils créés par Statistique Canada pour obtenir des estimations précises de la variance pour des variables préétablies. Une base de données a été créée à l'aide de ces estimations. Le Module d'extraction de coefficients de variation (MECV) est un moyen convivial d'accéder aux données. Les estimations de la variance sont disponibles pour les totaux et les proportions. Pour de plus amples renseignements, consultez le Guide de l'utilisateur du MECV.

### 10.3 Comment utiliser le Module d'extraction de coefficients de variation pour des estimations catégoriques

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

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

Le coefficient de variation est disponible directement à partir du MECV en choisissant la *Taille* comme l'estimateur pour des variables préétablies.

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

Le coefficient de variation est disponible directement à partir du MECV en choisissant le *Ratio* comme l'estimateur pour des variables préétablies.

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

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

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

où  $\hat{X}_1$  est l'estimation 1,  $\hat{X}_2$  est l'estimation 2 et  $\alpha_1$  et  $\alpha_2$  sont les coefficients de variation de  $\hat{X}_1$  et  $\hat{X}_2$  respectivement. Le coefficient de variation de  $\hat{d}$  est donné par  $\sigma_{\hat{d}}/\hat{d}$ . Cette formule est exacte pour la différence entre des caractéristiques distinctes et non corrélées, mais n'est autrement qu'approximative.

**Règle 4 : Estimations de rapports**

Dans le cas où le numérateur est un sous-ensemble du dénominateur, le rapport devrait être converti en un pourcentage et la règle 2 appliquée. Cela s'appliquerait, par exemple, au cas où le dénominateur est le nombre d'employés qui ont subi un changement d'emploi et le numérateur, le nombre d'employés permanents qui ont subi un changement d'emploi pendant la période de référence.

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

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

où  $\alpha_1$  et  $\alpha_2$  sont les coefficients de variation de  $\hat{X}_1$  et de  $\hat{X}_2$  respectivement. Le coefficient de variation de  $\hat{R}$  est donné par  $\sigma_{\hat{R}}/\hat{R}$ . La formule tendra à surestimer l'erreur si  $\hat{X}_1$  et  $\hat{X}_2$  sont corrélés positivement et à la sous-estimer si  $\hat{X}_1$  et  $\hat{X}_2$  sont corrélés négativement.

**Règle 5 : Estimations de différences entre des rapports**

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

### 10.3.1 Exemples d'utilisation du Module d'extraction de coefficients de variation pour des estimations catégoriques

Les exemples ci-dessous utilisent les cohortes 1 et 2 de l'ECEE et sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter.

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

L'utilisateur devrait consulter l'Exemple 1 du Guide de l'utilisateur du MECV.

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

L'utilisateur devrait consulter l'Exemple 2 du Guide de l'utilisateur du MECV.

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

Supposons qu'un utilisateur estime que 37,3 % la proportion d'employés de sexe masculin qui ont subi un changement d'emploi avaient une assurance-maladie dans leur emploi et 32,9 % la proportion d'employés de sexe féminin qui ont subi un changement d'emploi avaient une assurance-maladie dans leur emploi. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

- 1) L'utilisation de l'estimateur *Ratio* du MECV de la même façon que celle décrite dans l'exemple 2 du Guide de l'utilisateur du MECV, donne un CV de l'estimation pour les hommes de 3,3 % et un CV de l'estimation pour les femmes de 4,6 %.
- 2) En utilisant la règle 3, l'erreur-type d'une différence ( $\hat{d} = \hat{X}_1 - \hat{X}_2$ ) est :

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

où  $\hat{X}_1$  est l'estimation 1 (hommes),  $\hat{X}_2$  est l'estimation 2 (femmes) et  $\alpha_1$  et  $\alpha_2$  sont les coefficients de variation de  $\hat{X}_1$  et de  $\hat{X}_2$  respectivement.

C'est-à-dire que l'erreur-type de la différence  $\hat{d} = 0,373 - 0,329 = 0,044$  est :

$$\begin{aligned} \sigma_{\hat{d}} &= \sqrt{[(0,373)(0,033)]^2 + [(0,329)(0,046)]^2} \\ &= \sqrt{(0,000152) + (0,000229)} \\ &= 0,020 \end{aligned}$$

- 3) Le coefficient de variation de  $\hat{d}$  est donné par  $\sigma_{\hat{d}} / \hat{d} = 0,020 / 0,044 = 0,455$ .
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 45,5 %. La différence entre les estimations est considérée inacceptable et Statistique Canada recommande de ne pas publier cette estimation. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant

la lettre I (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les utilisateurs subséquents en garde contre les hauts taux d'erreur associés à l'estimation.

#### Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime que 565 638 le nombre d'employés de sexe masculin qui ont subi un changement d'emploi avaient une assurance-maladie dans leur emploi et 361 818 le nombre d'employés de sexe féminin qui ont subi un changement d'emploi avaient une assurance-maladie dans leur emploi. L'utilisateur est intéressé à comparer l'estimation des femmes à celle des hommes sous la forme d'un rapport. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation d'un rapport, où le numérateur de l'estimation ( $\hat{X}_1$ ) est le nombre d'employés de sexe féminin qui ont subi un changement d'emploi et qui avaient une assurance-maladie dans leur emploi. Le dénominateur de l'estimation ( $\hat{X}_2$ ) est le nombre d'employés de sexe masculin qui ont subi un changement d'emploi et qui avaient une assurance-maladie dans leur emploi..
- 2) Reportez-vous à l'estimateur de *Taille* dans le MECV.
- 3) Le numérateur de cette estimation de rapport est 361 818. En utilisant le MECV, le coefficient de variation pour cette estimation est 4,7 %.
- 4) Le dénominateur de cette estimation de rapport est 565 638. En utilisant le MECV, le coefficient de variation pour cette estimation est 3,4 %.
- 5) Le coefficient de variation approximatif de l'estimation du rapport est donc donné par la règle 4, qui est :

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

où  $\alpha_1$  et  $\alpha_2$  sont les coefficients de variation de  $\hat{X}_1$  et  $\hat{X}_2$  respectivement.

C'est-à-dire que :

$$\begin{aligned} \alpha_{\hat{R}} &= \sqrt{(0,047)^2 + (0,034)^2} \\ &= \sqrt{0,002209 + 0,001156} \\ &= 0,058 \end{aligned}$$

- 6) Le rapport obtenu entre les employés de sexe féminin qui ont subi un changement d'emploi et avaient une assurance-maladie dans leur emploi et les employés de sexe masculin qui ont subi un changement d'emploi et avaient une assurance-maladie dans leur emploi est 361 818 / 565 638, c'est-à-dire 0,64 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1). Le coefficient de variation de cette estimation est 5,8 %, ce qui fait qu'on peut la diffuser sans réserve.

### Exemple 5 : Estimations de différences de rapports

Supposons qu'un utilisateur estime que le rapport entre les employés de sexe féminin qui ont subi un changement d'emploi et avaient une assurance-maladie dans leur emploi et les employés de sexe masculin qui ont subi un changement d'emploi et avaient une assurance-maladie dans leur emploi s'établit à 0,64 tandis qu'il est de 0,62 pour les employés de sexe féminin par rapport aux employés de sexe masculin ayant un Régime d'assurance-dentaire. L'utilisateur est intéressé à comparer les deux rapports pour voir s'il y a une différence statistique entre ceux-ci. Comment l'utilisateur détermine-t-il le coefficient de variation de cette différence?

- 1) Tout d'abord calculez le coefficient de variation approximatif pour le rapport de l'assurance-maladie ( $\hat{R}_1$ ) et le rapport du Régime d'assurance-dentaire ( $\hat{R}_2$ ) tel qu'il est décrit dans l'exemple 4. Le CV approximatif pour le rapport de l'assurance-maladie est 5,8 % et 6,1 % pour celui du Régime d'assurance-dentaire.

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

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

où  $\alpha_1$  et  $\alpha_2$  sont les coefficients de variation de  $\hat{R}_1$  et  $\hat{R}_2$  respectivement. C'est-à-dire que l'erreur-type de la différence  $\hat{d} = 0,64 - 0,62 = 0,02$  est :

$$\begin{aligned} \sigma_{\hat{d}} &= \sqrt{[(0,64)(0,058)]^2 + [(0,62)(0,061)]^2} \\ &= \sqrt{(0,001378) + (0,001430)} \\ &= 0,053 \end{aligned}$$

- 3) Le coefficient de variation de  $\hat{d}$  est donné par  $\sigma_{\hat{d}} / \hat{d} = 0,053 / 0,02 = 2,65$ .
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 265 %. La différence entre les estimations est considérée inacceptable et Statistique Canada recommande de ne pas la publier. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre I (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les utilisateurs subséquents en garde contre les hauts taux d'erreur associés à l'estimation.

## 10.4 Comment utiliser le Module d'extraction de coefficients de variation pour obtenir des limites de confiance

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

Si l'échantillonnage de la population est répété indéfiniment, chaque échantillon menant à un nouvel intervalle de confiance pour une estimation, l'intervalle englobera alors dans

95 % des échantillons la valeur vraie de la population.

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

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

Des intervalles de confiance pour une estimation peuvent être calculés directement à partir du MECV, en déterminant d'abord le coefficient de variation de l'estimation  $\hat{X}$  (ou l'erreur-type) puis en utilisant les formules suivantes pour le convertir à un intervalle de confiance ( $IC_{\hat{x}}$ ) :

$$\text{En utilisant le coefficient de variation } IC_{\hat{x}} = (\hat{X} - t\hat{X}\alpha_{\hat{x}}, \hat{X} + t\hat{X}\alpha_{\hat{x}}),$$

$$\text{En utilisant l'erreur-type } IC_{\hat{x}} = (\hat{X} - t\hat{\sigma}_{\hat{x}}, \hat{X} + t\hat{\sigma}_{\hat{x}})$$

Où  $\alpha_{\hat{x}}$  est le coefficient de variation déterminé de  $\hat{X}$ , et  $\hat{\sigma}_{\hat{x}}$  est l'erreur-type déterminé de  $\hat{X}$ , et

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

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

#### **10.4.1 Exemple d'utilisation du Module d'extraction de coefficients de variation pour obtenir des limites de confiance**

Un intervalle de confiance de 95 % pour la proportion estimée d'employés de sexe féminin qui ont subi un changement d'emploi et avaient une assurance-maladie dans leur emploi par rapport aux employés de sexe masculin qui ont subi un changement d'emploi et avaient une assurance-maladie dans leur emploi (d'après l'exemple 4 à la section 10.3.1 serait calculé comme suit :

$$\hat{X} = 0,64 \%$$

$$t = 2$$

$\alpha_{\hat{x}}$  = 5,8 % (0,058 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.

$$IC_{\hat{x}} = \{0,64 - (2) (0,64) (0,058), 0,64 + (2) (0,64) (0,058)\}$$

$$IC_{\hat{x}} = \{0,64 - 0,07, 0,64 + 0,07\}$$

$$IC_{\hat{x}} = \{0,57, 0,71\}$$

Avec un intervalle de confiance de 95 %, on peut dire que le rapport entre les employés de sexe féminin et les employés de sexe masculin qui ont subi un changement d'emploi et avaient une assurance-maladie dans leur emploi est 0,57 et 0,71.

## 10.5 Comment utiliser le Module d'extraction de coefficients de variation pour effectuer un test t

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

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

Si  $t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$  se situe entre -2 et 2, aucune conclusion à propos de la différence entre les

caractéristiques n'est alors justifiée au niveau de signification de 5 %. Si, cependant, ce rapport est inférieur à -2 ou supérieur à +2, la différence observée est significative au niveau de 0,05. C'est-à-dire que la différence entre les estimations est significative.

### 10.5.1 Exemple d'utilisation le Module d'extraction de coefficients de variation pour effectuer un test t

Supposons que l'utilisateur désire tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y a pas de différence entre la proportion d'employés de sexe masculin qui ont subi un changement d'emploi et avaient une assurance-maladie dans leur emploi par rapport aux employés de sexe féminin qui ont subi un changement d'emploi et avaient une assurance-maladie dans leur emploi. D'après l'exemple 3 à la section 10.3.1, il s'est avéré que l'erreur-type de la différence entre ces deux estimations était 0,020. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0,373 - 0,329}{0,020} = \frac{0,044}{0,020} = 2,2$$

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

## **10.6 Le Module d'extraction de coefficients de variation**

Le fichier CVEM – MECV.xls contient le module d'extraction de coefficients de variation (MECV). Les cinq fichiers des bases de données doivent d'être dans le même répertoire que le fichier CVEM – MECV.xls.

tab\_C0102.xls

tab\_C0304.xls

tab\_C0506.xls

tab\_C0708.xls

tab\_C0910.xls

Le fichier ECEE\_Cohorte1à10\_GuF\_MECV.pdf renferme le Guide de l'utilisateur pour le MECV.



## 11.0 Pondération

Le facteur de pondération définitif rattaché à chaque enregistrement du fichier de microdonnées à grande diffusion (FMGD) de l'Enquête sur les changements à l'égard de l'emploi (ECEE) constitue un facteur de pondération combiné basé sur les facteurs de pondération de chacun des trimestres de référence de l'ECEE. Tous les enregistrements du fichier représentent des répondants de l'ECEE ayant accordé leur permission qu'on communique les données à leur sujet à Développement des ressources humaines Canada (DRHC); tous ces enregistrements présentent ainsi donc un facteur de pondération non nul positif. Les facteurs de pondération sont basés sur le fait qu'un répondant faisait partie ou non de l'échantillon d'une collectivité ou de l'échantillon provincial, mais les calculs de la pondération sont effectués au moyen des mêmes principes.

Voici la démarche suivie pour l'obtention des facteurs de pondération finals de l'enquête :

- Désignation et classification des répondants de l'enquête. On classe une personne comme répondant si elle répond à une partie donnée du questionnaire. Parmi ces personnes, seules celles ayant permis qu'on communique leurs réponses à DRHC sont considérées comme des répondants du point de vue de la pondération.
- Rajustement des poids d'échantillonnage d'après la non-réponse.
- Calibrage des poids. Ce calibrage tient compte du fait que la base de sondage pourrait ne pas être complète au moment de la sélection de chaque trimestre.
- Rajustement des poids en fonction du groupe de domaines. On crée un fichier de base de sondage semestrielle à numéro d'assurance sociale (NAS) unique et on subdivise tous les NAS en domaines. Les groupes de domaines tiennent compte de la répétition des NAS dans les fichiers de la base de sondage de chaque trimestre de référence et correspondent à l'une de ces trois possibilités : le NAS figure seulement dans la base de sondage du premier trimestre de référence de l'ECEE; il figure seulement dans la base de sondage du deuxième trimestre de référence; ou il figure dans les deux bases de sondage.
- Calibrage combiné des poids. Ce calibrage tient compte de la différence de taille entre les fichiers des bases de sondage semestrielles (un seul enregistrement par NAS) et les bases de sondage semestrielles. Il s'agit du facteur de pondération final figurant dans le FMGD.

Il faut utiliser le facteur de pondération final pour tirer des estimations significatives de l'enquête. Par exemple, s'il faut estimer le nombre de personnes dont l'emploi a pris fin en raison d'une mise à pied au cours des 12 derniers mois, on le fera en sélectionnant les enregistrements se rapportant à ces personnes dans l'échantillon présentant cette caractéristique, puis en effectuant la somme des facteurs de pondération introduits dans ces fichiers.

Les utilisateurs de ces fichiers sont priés de prendre note du concept clé suivant concernant les estimations des combinaisons de cohortes et celles des cohortes individuelles : pour l'ECEE, l'unité d'analyse est la personne, et non pas la cessation d'emploi. Pour les cohortes individuelles, les estimations de l'ECEE représentaient le nombre de personnes qui avaient vécu au moins une cessation d'emploi au cours d'une période donnée de trois mois. Pour les combinaisons de cohortes, les chiffres correspondaient au nombre de personnes qui avaient vécu une cessation d'emploi pendant une période donnée de six mois.

Le nombre total de personnes ayant vécu au moins une cessation d'emploi au cours d'une période de six mois **n'est pas** le même que le total des personnes qui ont vécu une cessation d'emploi dans la première cohorte, additionné au nombre de personnes qui ont vécu une cessation d'emploi dans la deuxième cohorte; le nombre de personnes qui vivent une cessation d'emploi dans deux cohortes consécutives est appréciable.

Ce « chevauchement » a été repéré grâce au jumelage des fichiers de la base de sondage de l'ECEE fournis par Ressources humaines et Développement des compétences Canada (RHDC) pour les combinaisons de cohortes.

Une fois qu'on a amalgamé les fichiers de la base de sondage pour les combinaisons de cohortes, la suppression des doublons a entraîné une réduction de la taille des fichiers et des dénombrements totaux pondérés. Cette baisse des dénombrements pondérés représente les personnes qui ont vécu une cessation d'emploi pendant les deux périodes de trois mois.

## **12.0 Questionnaires**

Les questionnaires de l'Enquête sur les changements à l'égard de l'emploi (ECEE) qui ont servi à recueillir des données pour l' ECEE ont subi un certain nombre de changements au cours des différentes périodes d'interviews de l'enquête. Les trois fichiers ci-dessous renferment les trois versions du questionnaire français.

ECEE\_C01\_W1\_QuestF.pdf

ECEE\_C02 à 04\_W1\_QuestF.pdf

ECEE\_C05 à 10\_W1\_QuestF.pdf



### **13.0 Cliché d'enregistrement à valeurs univariées**

Consulter les cinq fichiers ci-dessous qui renferment les clichés d'enregistrement à valeurs univariées des fichiers de microdonnées pour l'Enquête sur les changements à l'égard de l'emploi (ECEE).

ECEE1995\_C0102\_W1\_LvCds.pdf

ECEE\_Jan-Juin\_1996\_C0304\_W1\_LvCds.pdf

ECEE\_Juillet-Dec\_1996\_C0506\_W1\_LvCds.pdf

ECEE\_Jan-Juin\_1997\_C0708\_W1\_LvCds.pdf

ECEE\_Juillet-Dec\_1997\_C0910\_W1\_LvCds.pdf