

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

## **Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage**

2008



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
5.2.1	Stratification primaire	13
5.2.2	Types de régions	13
5.2.3	Stratification secondaire	14
5.2.4	Délimitation et sélection des grappes	14
5.2.5	Sélection des logements	15
5.2.6	Sélection des personnes	15
5.3	Taille de l'échantillon	15
5.4	Renouvellement de l'échantillon	16
5.5	Modifications apportées au plan de l'Enquête sur la population active pour l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage	16
5.6	Taille de l'échantillon par province pour l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage	17
<b>6.0</b>	<b>Collecte des données</b>	<b>19</b>
6.1	Supervision et contrôle de qualité	19
6.2	Non-réponse à l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage	19
<b>7.0</b>	<b>Traitement des données</b>	<b>21</b>
7.1	Saisie des données	21
7.2	Vérification	21
7.3	Codage des questions ouvertes	21
7.4	Imputation	21
7.5	Création de variables dérivées	22
7.6	Pondération	22
<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	23
8.2.1	Base de sondage	24
8.2.2	Collecte des données	24
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	Définitions de types d'estimations : catégoriques et quantitatives .....	28
9.3.1	Estimations catégoriques.....	28
9.3.2	Estimations quantitatives .....	29
9.3.3	Totalisation d'estimations catégoriques .....	29
9.3.4	Totalisation d'estimations quantitatives .....	29
9.4	Lignes directrices pour l'analyse statistique .....	30
9.5	Lignes directrices pour la diffusion de coefficients de variation.....	31
9.6	Seuils pour la diffusion des estimations pour l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage .....	33
<b>10.0</b>	<b>Tables de variabilité d'échantillonnage approximative</b> .....	<b>35</b>
10.1	Comment utiliser les tables de coefficients de variation pour des estimations catégoriques.....	36
10.1.1	Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques .....	37
10.2	Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance .....	42
10.2.1	Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance .....	43
10.3	Comment utiliser les tables de coefficients de variation pour effectuer un test t.....	43
10.3.1	Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t .....	43
10.4	Coefficients de variation pour des estimations quantitatives.....	44
10.5	Tables des coefficients de variation.....	44
<b>11.0</b>	<b>Pondération</b> .....	<b>45</b>
11.1	Procédures de pondération pour l'Enquête sur la population active .....	45
11.2	Procédures de pondération pour l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage .....	46
<b>12.0</b>	<b>Questionnaire</b> .....	<b>49</b>
<b>13.0</b>	<b>Cliché d'enregistrement à valeurs univariées</b> .....	<b>51</b>

## **1.0 Introduction**

Statistique Canada a mené l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage (EACA) en avril et mai 2008 avec l'appui et la collaboration du Conseil canadien sur l'apprentissage. 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)

### Conseil canadien sur l'apprentissage

Tracey Lavin  
#1805-701 West Georgia  
Vancouver (Colombie-Britannique) V7Y 1C6  
Téléphone: 604-662-3101 poste 314  
Courriel : [tlavin@ccl-cca.ca](mailto:tlavin@ccl-cca.ca)



## **2.0 Contexte**

L'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage (EACA) a été parrainée par le Conseil canadien sur l'apprentissage (CCA). Le CCA est un organisme national indépendant à but non lucratif qui a pour mandat de favoriser et d'appuyer la recherche en vue d'améliorer tous les aspects de l'apprentissage et de suivre les progrès du Canada à chacune des étapes de l'apprentissage, de la petite enfance au marché du travail, et plus loin encore.

Statistique Canada et le CCA mènera une enquête annuelle sur les attitudes à l'égard de l'apprentissage afin d'évaluer les besoins, les opinions et les connaissances des Canadiens au sujet de l'apprentissage et de l'éducation. L'EACA est la première enquête canadienne qui porte sur de vastes sujets touchant les attitudes des Canadiens à l'égard de l'apprentissage.

Les résultats de l'enquête intéresseront beaucoup ceux qui participent à la planification des politiques et des services des programmes et des initiatives actuels et futurs des secteurs public et privé dans le domaine de l'apprentissage et de l'éducation. L'enquête nous aidera à cerner les lacunes et pourrait mener à une meilleure compréhension des attitudes de la population canadienne à l'égard de l'apprentissage.





### **3.0 Objectifs**

L'objectif de l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage (EACA) est de recueillir les opinions, les attitudes et les attentes des Canadiens à l'égard de questions d'apprentissage.

L'enquête est centrée sur quatre domaines qui correspondent à des thèmes d'apprentissage d'importance capitale pour le Conseil canadien sur l'apprentissage (CCA) :

1. apprentissage durant la petite enfance;
2. apprentissage structuré (primaire, secondaire et postsecondaire);
3. apprentissage des adultes;
4. apprentissage sur la santé.

Un cinquième domaine vient compléter les thèmes précités, c'est-à-dire le capital social, qui enrichira le travail en cours sur l'indice composite de l'apprentissage du CCA. L'indice composite de l'apprentissage est une combinaison d'indicateurs qui mesure les conditions d'apprentissage favorables au bien-être économique et social des Canadiens.



## 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 du questionnaire d'enquête réellement employé(s).

Le questionnaire a été conçu sur la base des thèmes d'apprentissage et divisé en modules comme il est expliqué ci-après.

### **Introduction à l'enquête**

Tous les répondants choisis ont été informés de l'objectif de l'enquête, des utilisateurs des données et de l'usage prévu de celles-ci, de la nature volontaire de l'enquête et de ses dispositions de confidentialité. De plus, ils ont été priés d'autoriser la communication de l'information recueillie au Conseil canadien sur l'apprentissage.

### **Population active (LF)**

On demande à tous les répondants sélectionnés leur âge, leur sexe, leur état matrimonial, leur niveau de scolarité le plus élevé, leur activité sur le marché du travail et leur inscription à une école, un collège/cégep (collèges d'enseignement général et professionnel) ou une université.

### **Questions de sélection (SQ) et liste des enfants (CHD)**

On demande à tous les répondants sélectionnés s'ils ont des enfants, et s'il en est ainsi combien d'enfants. Pour les répondants sélectionnés qui ont des enfants de 0 à 24 ans, on présente un tableau qui permet d'indiquer le nom de chaque enfant, le sexe, l'âge, si l'enfant demeure au sein du ménage et si l'enfant a déjà été régulièrement sous les soins d'un des services de garde d'enfants entre 0 et 5 ans. Le tableau peut servir pour un maximum de 10 enfants.

### **Apprentissage des adultes (AL)**

Les questions du module Apprentissage des adultes mesurent les attitudes envers l'importance de l'apprentissage continu. Les questions évaluent la participation du répondant dans l'apprentissage lié au travail et l'apprentissage par intérêt personnel de même que les raisons de faire cet apprentissage. Les questions portent aussi sur les opinions du répondant sur les compétences en littératie et l'apprentissage nécessaire connexe. L'apprentissage désigne toute activité structurée ou non visant à accroître la connaissance.

### **Apprentissage structuré (SL)**

Les questions du module Apprentissage structuré évaluent le niveau de satisfaction envers divers aspects de l'enseignement et de l'apprentissage dans les écoles primaires, les écoles secondaires et les établissements d'enseignement postsecondaire. Les questions visent aussi à obtenir les attitudes envers les enjeux, tels que la taille des classes, les devoirs, l'intimidation et l'immersion française

### **Apprentissage durant la petite enfance (EC)**

Les questions du module Apprentissage durant la petite enfance portent sur les attitudes envers les services de garde dans le contexte de l'apprentissage durant la petite enfance. Les questions évaluent l'utilisation des divers types de services de garde et la manière dont les fournisseurs de service procèdent à l'apprentissage durant la petite enfance. L'information a été recueillie pour le plus jeune enfant dans la liste des enfants, âgé de 0 à 12 ans, qui a déjà été dans un service de garde d'enfants..

### **Apprentissage relié à la santé (HL)**

On pose à tous les répondants des questions sur la façon dont ils s'informent sur la santé en général. On leur demande où ils trouvent de l'information sur des problèmes de santé particuliers et, dans une moindre mesure, ce qui motive les gens à s'informer sur les questions de santé. Les questions sur la santé et l'apprentissage viseront également à déterminer les moyens à prendre pour permettre aux Canadiens d'avoir un accès meilleur ou plus grand aux renseignements sur la santé. La présente

section comprend quatre scénarios de défi de santé. Un défi a été attribué aléatoirement à chaque répondant.

**Capital social (SC)**

Tous les répondants ont été interrogés sur les ressources disponibles au sein de groupes sociaux. Capital social s'entend de la disponibilité de ressources à l'intérieur de groupes sociaux, par exemple familles, collectivités, entreprises, clubs sociaux. Le capital social représente la valeur que les gens tirent de leur interaction avec d'autres. De l'information concluante indique que les personnes qui ont accès à du capital social sont en bonne posture pour profiter pleinement des occasions d'apprentissage qui s'offrent à elles. Par conséquent, lors du contrôle des facteurs pertinents à l'enseignement et à l'apprentissage qui seront inclus dans l'indice composite de l'apprentissage, il importe de prendre en considération l'état du capital social chez les Canadiens. Les questions en la matière offrent des renseignements de base sur le capital social au Canada.

**Données démographiques (DM)**

Ce module permet de recueillir de l'information sociodémographique sur le répondant et sa famille. Cela comprend le pays de naissance, le statut d'immigration, la date d'arrivée au Canada, l'origine ethnique et l'appartenance ancestrale, la langue parlée à la maison, la compétence dans les deux langues officielles, le nombre de personnes formant le ménage, le revenu et ses sources.

## **5.0 Méthodologie de l'enquête**

L'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage (EACA) a été menée aux mois d'avril et mai 2008 auprès d'un sous-échantillon de personnes sélectionnées dans des logements issus de 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 5.4<sup>1</sup>. Les sections 5.5 et 5.6 décrivent comment l'EACA s'est écartée du plan de base de l'EPA. Les tailles d'échantillon initiales de l'EACA sont présentées à la section 5.6 par province ou territoire.

### **5.1 Population visée**

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

### **5.2 Plan de sondage**

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

#### **5.2.1 Stratification primaire**

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

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

#### **5.2.2 Types de régions**

Les strates primaires (les intersections de RE et de REAE) sont en outre désagrégées en trois types de régions : rurales, urbaines et éloignées. Les régions rurales et urbaines sont généralement fondées sur les définitions du recensement de régions urbaines et de

---

<sup>1</sup> Une description détaillée du plan de l'EPA figure dans la publication de Statistique Canada intitulée *Méthodologie de l'Enquête sur la population active du Canada*, n° 71-526-XPB au catalogue.

régions rurales, à quelques exceptions près destinées à permettre la formation de strates dans certaines régions. Les régions urbaines vont des RMR les plus grandes jusqu'aux villages les plus petits classés suivant le Recensement de 1991 dans la catégorie des régions urbaines (de 1 000 habitants ou plus), tandis que les régions rurales se composent des régions non désignées régions urbaines ni régions éloignées.

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

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

### **5.2.3 Stratification secondaire**

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

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

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

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

### **5.2.4 Délimitation et sélection des grappes**

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

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

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

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

### **5.2.5 Sélection des logements**

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

### **5.2.6 Sélection des personnes**

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

## **5.3 Taille de l'échantillon**

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

L'échantillon mensuel de l'EPA se compose d'environ 60 000 logements. Après en avoir exclu les logements trouvés vacants, les logements démolis ou ceux convertis à des fins non résidentielles, ceux n'abritant que des personnes inadmissibles, les logements en construction et les logements saisonniers, il reste à peu près 54 000 logements occupés par une ou par plusieurs personnes admissibles. On recueille de l'information aux fins de l'EPA à partir de ces logements sur environ 102 000 civils âgés de 15 ans et plus.

#### **5.4 Renouvellement de l'échantillon**

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

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

#### **5.5 Modifications apportées au plan de l'Enquête sur la population active pour l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage**

On a utilisé pour l'EACA trois groupes de renouvellement de l'échantillon de l'EPA dont la dernière entrevue avait eu lieu aux mois d'octobre et décembre 2007, et février 2008. Pour les besoins de l'EACA on a établi la couverture de l'EPA au niveau des ménages. Cependant, contrairement à l'EPA, pour laquelle des données sont recueillies pour tous les membres admissibles d'un ménage, l'EACA n'a donné lieu à la cueillette d'information qu'après d'un seul membre présélectionné d'un ménage. Les réponses par personne interposée n'étaient pas permises pour l'EACA.

De plus les ménages contenant uniquement des personnes âgées de moins de 18 ans ou de plus de 74 ans ont été exclus de l'EACA. Le tableau à la section 5.3 présente les tailles d'échantillon par province et territoire.



## **5.6 Taille de l'échantillon par province pour l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage**

Le tableau qui suit montre le nombre de personnes sélectionnées pour répondre à l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage. Ce tableau comprend des ménages qui étaient non répondants à l'EPA.

<b>Province</b>	<b>Taille de l'échantillon</b>
Terre-Neuve-et-Labrador	660
Île-du-Prince-Édouard	569
Nouvelle-Écosse	718
Nouveau-Brunswick	670
Québec	1 210
Ontario	1 380
Manitoba	720
Saskatchewan	670
Alberta	890
Colombie-Britannique	1 123
<b>Canada</b>	<b>8 610</b>



## **6.0 Collecte des données**

La collecte des données aux fins de l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage (EACA) a eu lieu entre le 1<sup>er</sup> avril et le 31 mai 2008. L'EACA a été effectuée auprès d'une personne par ménage sélectionnée au hasard par l'entremise d'un questionnaire interview téléphonique assistée par ordinateur (ITAO). La sélection au hasard a été effectuée avant le début de la collecte des données lors de la sélection de l'échantillon (voir section 5.5).

Puisque la période de collecte des données pour l'EACA a suivi de plusieurs mois la dernière interview des ménages issus de l'EPA, les intervieweurs ont procédé à un exercice de dépistage lorsque le répondant sélectionné pour l'EACA avait déménagé ou lorsque le numéro de téléphone dans nos dossiers n'était plus valide ou en service.

### **6.1 Supervision et contrôle de qualité**

Tous les intervieweurs 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 et les procédures de l'enquête. Ils ont aussi pour fonction de contrôler régulièrement le travail des intervieweurs et d'examiner les documents complétés par ces derniers. Les intervieweurs principaux sont, à leur tour, supervisés par les gestionnaires du programme dans chacun des bureaux régionaux de Statistique Canada.

### **6.2 Non-réponse à l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage**

Les intervieweurs ont pour consigne de faire tous les efforts raisonnables pour obtenir des interviews avec le répondant sélectionné. Lorsqu'une personne refuse au départ de participer, le bureau régional envoie à l'adresse du logement une lettre soulignant l'importance que revêtent l'enquête et la collaboration du ménage. Cette lettre est suivie d'un deuxième appel. Dans les cas où l'appel de l'intervieweur survient à un moment inopportun, on fixe un rendez-vous pour un rappel à un moment plus opportun. Lorsqu'il n'y a personne dans le logement, l'intervieweur rappelle à de nombreuses reprises. Quelles que soient les circonstances, on ne remplace jamais des personnes échantillonnées par d'autres individus en cas de non-réponse.



## **7.0 Traitement des données**

Le principal produit de l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage (EACA) est un fichier de microdonnées « épuré ». Ce chapitre présente un bref résumé des phases de traitement inhérentes à la production de ce fichier.

### **7.1 Saisie des données**

Les intervieweurs saisissent directement les réponses aux questions de l'enquête au moment de l'interview à l'aide d'une version automatisée du questionnaire. L'emploi d'un questionnaire automatisé réduit les délais et coûts de traitement associés à la saisie des données, aux erreurs de transcription et à la transmission des données. On soumet les données des réponses à un cryptage pour en préserver le caractère confidentiel, puis on les transmet par modem au bureau régional concerné de Statistique Canada. Les données sont par la suite transmises à Ottawa, au moyen d'une ligne protégée, pour y être traitées davantage.

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

### **7.2 Vérification**

La première étape du traitement d'enquête effectué au bureau central a été de remplacer les valeurs « hors limites » incluses dans le fichier de données par des blancs. Ce processus a été conçu pour faciliter les vérifications ultérieures.

Les erreurs dans le déroulement du questionnaire, où l'on a relevé des questions qui ne s'appliquaient pas au répondant (et auxquelles on n'aurait donc pas dû répondre) et qui renfermaient des réponses, constituaient le premier type d'erreurs traitées. Dans ces cas, une vérification par ordinateur a éliminé automatiquement les données superflues en suivant l'ordre du questionnaire dicté par les réponses à des questions antérieures et subséquentes, parfois.

Le second type d'erreurs traitées avait trait à un manque d'information dans les questions pour lesquelles le répondant aurait dû répondre. Pour ce type d'erreur, un code de non-réponse ou « non déclaré » était attribué au poste.

### **7.3 Codage des questions ouvertes**

Quelques données élémentaires ont été consignées sur le questionnaire par les intervieweurs sous forme de questions ouvertes. L'enquête contenait au total deux questions partiellement ou complètement ouvertes au sujet du pays de naissance et de l'origine ethnique.

### **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 survient 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). Il y a non-réponse ponctuelle 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 survient 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.

De plus amples renseignements sur le processus d'imputation sont donnés au chapitre 8.0 (Qualité des données).

## **7.5 Création de variables dérivées**

Un certain nombre de données élémentaires incluses dans le fichier de microdonnées ont été calculées en combinant des postes sur le questionnaire pour faciliter l'analyse des données. Ces variables sont énumérées dans la liste des codes. LFSTAT est une variable dérivée (VD) importante qui détermine la situation d'activité du répondant en fonction des réponses données aux questions du module sur la population active (LF). EDUCLEV, autre exemple de VD, sert à établir le plus haut niveau de scolarité atteint par le répondant. Diverses autres VD sont conçues en vue de caractériser des variables continues et de faciliter la collecte d'information sur les enfants du répondant (âge et sexe).

## **7.6 Pondération**

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

La phase de la pondération est une étape où l'on calcule ce nombre (ou poids) pour chaque enregistrement. Ce poids, qui figure dans le fichier de microdonnées, **doit** servir à calculer des estimations significatives à partir de l'enquête. Si, par exemple, le nombre de personnes ayant suivi un programme ou une formation reliés au travail durant les 12 dernier mois, doit être estimé, cette opération s'effectue en sélectionnant les enregistrements se référant aux personnes incluses à l'intérieur de l'échantillon qui présentent cette caractéristique et en additionnant les poids inscrits dans ces enregistrements.

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

## 8.0 Qualité des données

### 8.1 Taux de réponse

Le tableau qui suit renferme un résumé des taux de réponse aux questionnaires de l'Enquête sur la population active (EPA) et à celui de l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage (EACA).

Province	Ménages sélectionnés de l'EPA	Ménages répondants à l'EPA	Taux de réponse à l'EPA* (%)	Ménages sélectionnés de l'EACA	Ménages répondants à l'EACA	Taux de réponse à l'EACA** (%)	Taux de réponse global à l'EACA*** (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(3)*(6)
Terre-Neuve-et-Labrador	971	909	93,6	660	423	64,1	60,0
Île-du-Prince-Édouard	713	666	93,4	569	354	62,2	58,1
Nouvelle-Écosse	1 461	1 362	93,2	718	499	69,5	64,8
Nouveau-Brunswick	1 361	1 278	93,9	670	471	70,3	66,0
Québec	4 978	4 660	93,6	1 210	794	65,6	61,4
Ontario	7 795	7 174	92,0	1 380	857	62,1	57,2
Manitoba	1 866	1 711	91,7	720	456	63,3	58,0
Saskatchewan	1 863	1 695	91,0	670	427	63,7	58,0
Alberta	2 735	2 456	89,8	890	550	61,8	55,5
Colombie-Britannique	3 185	2 898	91,0	1 123	657	58,5	53,2
<b>Canada</b>	<b>26 928</b>	<b>24 809</b>	<b>92,1</b>	<b>8 610</b>	<b>5 488</b>	<b>63,7</b>	<b>58,7</b>

**Nota :** Les chiffres de l'EPA s'appliquent aux ménages alors que les chiffres de l'EACA s'appliquent aux personnes sélectionnées à l'intérieur des ménages (seulement une personne est sélectionnée par ménage). Le taux de réponse global de l'EACA est fondé sur tous les enregistrements des ménages de l'EPA. Les ménages répondants de l'EPA comprennent les répondants qui ont été reportés du mois précédent.

\* Le taux de réponse à l'EPA est le nombre de ménages ayant répondu au questionnaire de l'EPA exprimé sous forme de pourcentage du nombre de ménages sélectionnés pour l'EPA.

\*\* Le taux de réponse à l'EACA est le nombre de personnes ayant répondu au questionnaire de l'EACA et qui ont accepté de partager les données recueillies lors de l'entrevue, exprimé sous forme de pourcentage du nombre de ménages ayant été sélectionnés pour l'EACA.

\*\*\* Le taux de réponse global à l'EACA est le nombre de personnes ayant répondu au questionnaire et accepté de partager les données recueillies lors de l'entrevue de l'EACA exprimé sous forme de pourcentage du nombre de ménages ayant répondu au questionnaire de l'EPA et sélectionnés pour l'EACA.

### 8.2 Erreurs relatives à l'enquête

Les estimations calculées à partir de cette enquête reposent sur un échantillon de ménages. Des estimations légèrement différentes auraient pu être obtenues si un recensement complet avait

été effectué en reprenant le même questionnaire et en faisant appel aux mêmes intervieweurs, superviseurs, méthodes de traitement, etc. que ceux effectivement utilisés dans l'enquête. L'écart entre les estimations découlant de l'échantillon et celles que donnerait un dénombrement complet réalisé dans des conditions semblables est appelé erreur d'échantillonnage de l'estimation. Des erreurs qui ne sont pas liées à l'échantillonnage peuvent se produire à presque toutes les étapes des opérations d'enquête. Les intervieweurs peuvent avoir mal compris les instructions, les enquêtés peuvent se tromper en répondant aux questions, les réponses peuvent être mal saisies sur le questionnaire et des erreurs peuvent survenir lors du traitement et de la totalisation des données. Ces erreurs sont toutes des exemples d'erreurs non dues à l'échantillonnage.

Sur un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations calculées à partir de l'enquête. Toutefois, les erreurs systématiques contribuent à biaiser les estimations de l'enquête. Énormément de temps et d'efforts ont été consacrés à réduire les erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été prises à chacune des étapes du cycle de collecte et de traitement des données afin de contrôler la qualité des données. Ces mesures comprenaient le recours à des intervieweurs hautement qualifiés, une formation poussée des intervieweurs concernant les procédures d'enquête et du questionnaire, l'observation des intervieweurs en vue de cerner les problèmes liés à la conception du questionnaire ou à une mauvaise compréhension des instructions, des procédures visant à s'assurer que les erreurs de saisie des données étaient réduites au minimum ainsi que des vérifications de la qualité du codage et de contrôle ayant pour but d'attester la logique du traitement.

### **8.2.1 Base de sondage**

Comme l'EACA était une enquête supplémentaire à l'EPA, sa base de sondage était constituée de ménages répondants à l'EPA. Toute non-réponse à l'EPA avait un impact sur la base de sondage de l'EACA. La qualité des variables d'échantillon dans la base était très élevée. L'échantillon de l'EACA était composé de trois groupes de renouvellement sortants de l'EPA. Les critères de sélection utilisés pour l'EACA (tel que le groupe de renouvellement) n'étaient manquants pour aucun des enregistrements de l'EPA.

Il est à noter que la base de sondage de l'EPA exclut environ 2 % de tous les ménages dans les 10 provinces du Canada. La base de sondage de l'EACA 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.

Les cas d'interviews sur place assistées par ordinateur (IPAO) de l'EPA ont été automatiquement mis « hors du champ de l'enquête », sans aucune procédure de dépistage malgré le fait qu'une adresse était disponible. Un certain niveau de dépistage a cependant été effectué pour atteindre les cas d'interviews téléphoniques assistées par ordinateur (ITAO) quand le numéro de téléphone n'était plus valide ou hors service.

### **8.2.2 Collecte des données**

La formation des intervieweurs pour l'EACA consistait à lire le Manuel des procédures et le Manuel de l'intervieweur, la pratique avec des cas simulés de l'EACA 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. La période de collecte allait de la semaine du 1<sup>er</sup> avril au 31 mai 2008.



### **8.2.3 Traitement des données**

Au cours du traitement des données de l'EACA, certains enregistrements ont été rejetés parce que des répondants n'ont pas accepté de partager leurs réponses avec le Conseil canadien sur l'apprentissage. Il y avait 65 enregistrements de ce genre et ceux-ci ont été codés comme cas de non-réponse.

Le traitement des données de l'EACA 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 cas de non-réponse totale ont été traités 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 de 2006, l'on estime que 18,65 % des personnes de 20 ans et plus ont au moins un enfant âgé entre 0 et 8 ans et l'on constate que l'erreur-type de cette estimation est de 0,0135. Le coefficient de variation de l'estimation est donc calculé comme suit :

$$\left( \frac{0,0135}{0,1865} \right) \times 100 \% = 7,24 \%$$

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 attitudes des Canadiens à l'égard de l'apprentissage (EACA) n'était pas autopondéré. Lorsqu'ils produisent des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent appliquer les poids d'enquête appropriés.

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

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

## **9.3 Définitions de types d'estimations : catégoriques et quantitatives**

Avant de discuter de la façon dont on peut totaliser et analyser les données de l'EACA, 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'EACA.

### **9.3.1 Estimations catégoriques**

Les estimations catégoriques sont des estimations du nombre ou du pourcentage de membres de la population visée par l'enquête possédant certaines caractéristiques ou faisant partie d'une catégorie définie. Le nombre de personnes ayant suivi une formation reliée au travail durant la dernière année et le nombre de personnes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années, 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: Pendant les 12 derniers mois, avez-vous suivi un programme menant à l'obtention d'un diplôme ou d'un certificat ou avez-vous assisté à des cours, à des ateliers ou à des séminaires ou reçu une autre formation liée à un emploi ou à une carrière?

R: Oui / Non

Q: Est-ce que [prénom de l'enfant] a déjà été régulièrement sous les soins d'un des services de garde d'enfants suivants?

R: Une garderie (centre de la petite enfance) / Garde dans la maison d'une autre personne par une personne non apparentée / Garde dans la maison de l'enfant par une personne non apparentée / Garde par une personne apparentée autre qu'un parent / Tout autre service de garde régulier autre qu'un parent / L'enfant n'avait jamais été gardé régulièrement entre sa naissance et 5 ans

### 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 d'enfants des personnes de 20 ans et plus. Le numérateur est une estimation du nombre total d'enfants des personnes de 20 ans et plus et son dénominateur est le nombre de personnes de 20 ans et plus ayant au moins un enfant.

#### Exemples de questions quantitatives :

Q : Combien d'enfants avez-vous? Veuillez inclure les enfants dont vous avez la garde légale ainsi que ceux qui ne demeurent pas avec vous, mais avec qui vous êtes encore en contact.

R : |\_|\_| enfants

Q : Combien tous les membres de votre ménage ont-ils reçu en revenu total de toutes les sources mentionnées avant impôts et retenues, pour l'année se terminant le 31 décembre 2007?

R : |\_|\_|\_|\_|\_|\_| dollars

### 9.3.3 Totalisation d'estimations catégoriques

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

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

### 9.3.4 Totalisation d'estimations quantitatives

On peut obtenir des estimations de quantités à partir du fichier de microdonnées en multipliant la valeur de la variable qui nous intéresse par le poids final de chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements qui nous intéressent. Pour obtenir, par exemple, une estimation du nombre total d'enfants des personnes de 18 ans à 74 ans, multipliez la valeur déclarée à la question SQ\_Q02 (Nombre d'enfants du répondant, qu'ils demeurent ou non avec lui) par le poids final de l'enregistrement et additionnez cette quantité pour tous les enregistrements où la variable SQ\_Q01 = 1 (ont des enfants).

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 d'enfants des personnes de 18 à 74 ans:

- a) estimez le nombre total d'enfants ( $\hat{X}$ ) tel qu'il est décrit ci-dessus,
- b) estimez le nombre total de personnes de 18 à 74 ans qui ont des enfants ( $\hat{Y}$ ) incluses dans cette catégorie en additionnant les poids finals de tous les enregistrements où la variable SQ\_Q01 = 1, puis
- c) divisez l'estimation a) par l'estimation b) ( $\hat{X} / \hat{Y}$ ).

## 9.4 Lignes directrices pour l'analyse statistique

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

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

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

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

- 1) sélectionner tous les répondants du fichier qui ont déclaré LF\_Q01 = 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 LF\_Q01 = 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'EACA, les utilisateurs devraient premièrement déterminer le niveau de qualité de cette estimation. Les niveaux de qualité sont *acceptable*, *médiocre* et *inacceptable*. Les erreurs d'échantillonnage et non dues à l'échantillonnage, dont il a été question au chapitre 8.0, influencent la qualité des données. Aux fins du présent document, cependant, on ne déterminera le niveau de qualité d'une estimation qu'à partir d'une erreur d'échantillonnage dont rend compte le coefficient de variation indiqué à l'intérieur du tableau qui figure ci-dessous. Les utilisateurs devraient néanmoins s'assurer de lire le chapitre 8.0 pour être plus pleinement informés des caractéristiques relatives à la qualité de ces données.

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

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

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

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

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



## 9.6 Seuils pour la diffusion des estimations pour l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage

Le tableau ci-dessous fournit une indication de la précision des estimations des prévisions démographiques ainsi que les seuils de diffusion associés aux trois niveaux de qualité de l'estimation présentés à la section précédente. Ces seuils proviennent des tables de coefficients de variation (CV) dont il sera question au chapitre 10.0.

Par exemple, d'après le tableau, la qualité d'une estimation pondérée de 18 500 personnes possédant une caractéristique donnée à Terre-Neuve-et-Labrador est médiocre.

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

Provinces et régions selon le lieu de résidence en 2008	CV acceptable 0,0 à 16,5 %	CV médiocre 16,6 à 33,3 %	CV inacceptable > 33,3 %
Provinces de l'Atlantique	64 000 et plus	16 000 à < 64 000	moins de 16 000
Terre-Neuve-et-Labrador	47 000 et plus	12 500 à < 47 000	moins de 12 500
Île-du-Prince-Édouard	14 500 et plus	4 000 à < 14 500	moins de 4 000
Nouvelle-Écosse	72 500 et plus	19 500 à < 72 500	moins de 19 500
Nouveau-Brunswick	63 500 et plus	17 000 à < 63 500	moins de 17 000
Québec	535 000 et plus	141 500 à < 535 000	moins de 141 500
Ontario	909 000 et plus	241 000 à < 909 000	moins de 241 000
Provinces des Prairies	270 500 et plus	70 000 à < 270 500	moins de 70 000
Manitoba	109 500 et plus	30 000 à < 109 500	moins de 30 000
Saskatchewan	91 500 et plus	25 000 à < 91 500	moins de 25 000
Alberta	344 500 et plus	94 500 à < 344 500	moins de 94 500
Colombie-Britannique	356 500 et plus	95 500 à < 356 500	moins de 95 500
<b>Canada</b>	<b>612 500 et plus</b>	<b>153 500 à &lt; 612 500</b>	<b>moins de 153 500</b>



## 10.0 Tables de variabilité d'échantillonnage approximative

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

Les coefficients de variation sont calculés à l'aide de la formule de la variance pour un échantillonnage aléatoire simple et en y incorporant un facteur qui reflète la nature du plan d'échantillonnage, qui est à plusieurs degrés et qui prévoit la formation de grappes. Ce facteur, appelé l'effet du plan, a été déterminé en calculant premièrement les effets du plan pour une gamme étendue de caractéristiques, puis en choisissant parmi ceux-ci une valeur modérée (habituellement le 75<sup>e</sup> percentile) à utiliser à l'intérieur des tables de CV qui s'appliqueraient ensuite à l'ensemble entier des caractéristiques.

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

Provinces et régions	Effet du plan	Taille de l'échantillon	Population
Provinces de l'Atlantique	1,87	1 747	1 693 586
Terre-Neuve-et-Labrador	1,63	423	378 641
Île-du-Prince-Édouard	1,65	354	99 845
Nouvelle-Écosse	1,64	499	672 946
Nouveau-Brunswick	1,70	471	542 154
Québec	2,28	794	5 608 374
Ontario	2,55	857	9 225 758
Provinces des Prairies	2,88	1 433	3 938 066
Manitoba	2,00	456	787 316
Saskatchewan	1,89	427	655 085
Alberta	2,40	550	2 495 665
Colombie-Britannique	2,25	657	3 189 416
<b>Canada</b>	<b>3,97</b>	<b>5 488</b>	<b>23 655 200</b>

Tous les coefficients de variation inclus dans les tables de variabilité d'échantillonnage approximative sont approximatifs et donc non officiels. Des estimations de la variance réelle pour des variables précises peuvent être obtenues auprès de Statistique Canada, contre remboursement des frais. Étant donné que le CV approximatif est une estimation prudente, l'utilisation de la variance réelle estimée pourrait faire passer l'estimation d'un niveau de qualité à un autre. Par exemple, une estimation *médiocre* pourrait devenir *acceptable* si elle était fondée sur le calcul du CV exact.

Rappelez-vous que : Si le nombre d'observations sur lesquelles une estimation est basée est inférieur à 30, l'estimation pondérée est très probablement inacceptable et Statistique Canada recommande de ne pas diffuser une telle estimation, quelle que soit la valeur du coefficient de variation.

## 10.1 Comment utiliser les tables de coefficients de variation pour des estimations catégoriques

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

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

Le coefficient de variation dépend uniquement de la taille de l'estimation elle-même. Dans la table de variabilité d'échantillonnage approximative pour la région géographique appropriée, repérez le nombre estimé dans la colonne la plus à gauche (intitulée « Numérateur du pourcentage ») et suivez les astérisques (le cas échéant) jusqu'au premier chiffre rencontré. Ce chiffre est le coefficient de variation approximatif.

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

Le coefficient de variation d'une proportion estimée ou d'un pourcentage estimé dépend à la fois de la taille de la proportion ou du pourcentage et de la taille du total sur lequel la proportion ou le pourcentage repose. Les proportions estimées ou les pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque la proportion ou que le pourcentage repose sur un sous-groupe de la population. La proportion, par exemple, de personnes ayant au moins un enfant âgé entre 0 et 8 ans est plus fiable que le nombre estimé de personnes ayant au moins un enfant âgé entre 0 et 8 ans. (Remarquez que dans les tables la valeur des coefficients de variation diminue lorsqu'on les lit de gauche à droite.)

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

Lorsque la proportion ou que le pourcentage repose sur un sous-ensemble de la population totale (p. ex., comme membres d'un groupe d'âge particulier), on devrait faire référence à la proportion ou au pourcentage (dans le haut de la table) et au numérateur de la proportion ou du pourcentage (dans la colonne de gauche de la table). L'intersection de la rangée et de la colonne appropriées donne le coefficient de variation.

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

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

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

où  $\hat{X}_1$  est l'estimation 1,  $\hat{X}_2$  est l'estimation 2 et  $\alpha_1$  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 de personnes ayant au moins un enfant âgé entre 0 et 8 ans et le numérateur, le nombre de personnes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années.

Dans le cas où le numérateur n'est pas un sous-ensemble du dénominateur, comme dans l'exemple du rapport du nombre d'hommes ayant un enfant âgé entre 0 et 8 ans comparativement au nombre de femmes ayant un enfant âgé entre 0 et 8 ans, 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.

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

Les exemples ci-dessous utilisent des données du fichier de l'EACA de 2006 et sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter. Veuillez noter que les données utilisées dans ces exemples diffèrent des véritables résultats de l'enquête et tiennent seulement lieu de guide.

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

Supposons qu'un utilisateur estime que 4 493 502 de personnes avaient au moins un enfant âgé entre 0 et 8 ans. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Reportez-vous à la table des coefficients de variation pour le CANADA.
- 2) L'agrégat estimé (4 493 502) ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, c'est-à-dire 4 000 000.
- 3) On trouve le coefficient de variation pour un agrégat estimé en se reportant à la

première entrée autre que des astérisques sur cette rangée, c'est-à-dire 7,0 %.

- 4) Le coefficient de variation approximatif de l'estimation est donc 7,0 %. Le résultat selon lequel il y avait 4 493 502 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1) de personnes ayant au moins un enfant âgé entre 0 et 8 ans peut être publié sans réserve.

L'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage, 2006													
Tables de la variabilité d'échantillonnage approximative - Canada													
NUMÉRATEUR DU POURCENTAGE (en milliers)	POURCENTAGE ESTIMÉ												
	0,1%	1,0%	2,0%	5,0%	10,0%	15,00%	20,0%	...	35,0%	40,0%	50,0%	70,0%	90,0%
1	496,3	494,1	491,6	484,0	471,1	457,8	444,1	...	400,3	384,6	351,1	272,0	157,0
2	350,9	349,4	347,6	342,2	333,1	323,7	314	...	283,1	272,0	248,3	192,3	111,0
3	286,5	285,2	283,8	279,4	272,0	264,3	256,4	...	231,1	222,1	202,7	157,0	90,7
4	248,2	247,0	245,8	242,0	235,5	228,9	222,1	...	200,2	192,3	175,6	136,0	78,5
5	222,0	221,0	219,8	216,4	210,7	204,7	198,6	...	179,0	172,0	157,0	121,6	70,2
6	202,6	201,7	200,7	197,6	192,3	186,9	181,3	...	163,4	157,0	143,3	111,0	64,1
7	187,6	186,7	185,8	182,9	178,0	173	167,9	...	151,3	145,4	132,7	102,8	59,3
8	175,5	174,7	173,8	171,1	166,5	161,9	157	...	141,5	136,0	124,1	96,2	55,5
9	165,4	164,7	163,9	161,3	157,0	152,6	148	...	133,4	128,2	117,0	90,7	52,3
10	156,9	156,2	155,4	153,0	149,0	144,8	140,4	...	126,6	121,6	111,0	86,0	49,7
11	149,6	149,0	148,2	145,9	142,0	138	133,9	...	120,7	116,0	105,9	82,0	47,3
12	143,3	142,6	141,9	139,7	136,0	132,2	128,2	...	115,6	111,0	101,4	78,5	45,3
.	.	.	.	.	.	.	.	...	.	.	.	.	.
.	.	.	.	.	.	.	.	...	.	.	.	.	.
.	.	.	.	.	.	.	.	...	.	.	.	.	.
450	****	****	23,2	22,8	22,2	21,6	20,9	...	18,9	18,1	16,6	12,8	7,4
500	****	****	****	21,6	21,1	20,5	19,9	...	17,9	17,2	15,7	12,2	7,0
750	****	****	****	17,7	17,2	16,7	16,2	...	14,6	14,0	12,8	9,9	5,7
1 000	****	****	****	15,3	14,9	14,5	14	...	12,7	12,2	11,1	8,6	5,0
1 500	****	****	****	****	12,2	11,8	11,5	...	10,3	9,9	9,1	7,0	4,1
2 000	****	****	****	****	10,5	10,2	9,9	...	9,0	8,6	7,9	6,1	3,5
3 000	****	****	****	****	****	8,4	8,1	...	7,3	7,0	6,4	5,0	2,9
4 000	****	****	****	****	****	****	7,0	...	6,3	6,1	5,6	4,3	2,5
5 000	****	****	****	****	****	****	****	...	5,7	5,4	5,0	3,8	2,2
6 000	****	****	****	****	****	****	****	...	5,2	5,0	4,5	3,5	2,0
7 000	****	****	****	****	****	****	****	...	4,8	4,6	4,2	3,3	1,9
8 000	****	****	****	****	****	****	****	...	4,5	4,3	3,9	3,0	1,8
9 000	****	****	****	****	****	****	****	...	****	4,1	3,7	2,9	1,7
10 000	****	****	****	****	****	****	****	...	****	****	3,5	2,7	1,6
12 500	****	****	****	****	****	****	****	...	****	****	****	2,4	1,4
15 000	****	****	****	****	****	****	****	...	****	****	****	2,2	1,3
20 000	****	****	****	****	****	****	****	...	****	****	****	****	1,1

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

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

Supposons qu'un utilisateur estime à  $1\,920\,409 / 4\,493\,502 = 42,7\%$  la proportion de personnes ayant au moins un enfant âgé entre 0 et 8 ans qui fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Reportez-vous à la table des coefficients de variation pour le CANADA.
- 2) Parce que l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c'est-à-dire les personnes ayant au moins un enfant âgé entre 0 et 8 ans), il faut utiliser à la fois le pourcentage (42,7 %) et la portion numérateur du pourcentage (1 920 409) pour déterminer le coefficient de variation.
- 3) Le numérateur, 1 920 409, ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, soit 2 000 000. De même, l'estimation du pourcentage ne figure dans l'en-tête d'aucune colonne; il faut donc utiliser la proportion qui s'en rapproche le plus, soit 40,0 %.
- 4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 8,6 %, est le coefficient de variation à employer.
- 5) Le coefficient de variation approximatif de l'estimation est donc 8,6 %. Le résultat selon lequel 42,7 % des personnes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années peut être publié sans réserve.

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

Supposons qu'un utilisateur estime à  $775\,148 / 1\,992\,794 = 38,9\%$  la proportion de femmes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années par rapport à  $1\,145\,261 / 2\,500\,708 = 45,8\%$  la proportion d'hommes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

- 1) L'utilisation de la table des coefficients de variation du CANADA de la même façon que celle décrite dans l'exemple 2, donne un CV de l'estimation pour les femmes de 14,0 % et un CV de l'estimation pour les hommes de 11,1 % .
- 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 (femmes),  $\hat{X}_2$  est l'estimation 2 (hommes) 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  $\hat{d} = 0,389 - 0,458 = -0,069$  est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0,389)(0,14)]^2 + [(0,458)(0,111)]^2} \\ &= \sqrt{(0,0029658) + (0,0025845)} \\ &= 0,0745\end{aligned}$$

- 3) Le coefficient de variation de  $\hat{d}$  est donné par  $\sigma_{\hat{d}} / \hat{d} = 0,0745 / 0,069 = 1,0797$ .
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 108,0 %. La différence entre les estimations est considérée inacceptable et Statistique Canada recommande de ne pas publier cette estimation. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre F (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les utilisateurs subséquents en garde contre les hauts taux d'erreur associés à l'estimation.

#### Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime à 775 148 le nombre de femmes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années, par rapport à 1 145 261 d'hommes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années. 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 de femmes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années. Le dénominateur de l'estimation ( $\hat{X}_2$ ) est le nombre d'hommes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années.
- 2) Reportez-vous à la table des coefficients de variation pour le CANADA.
- 3) Le numérateur de cette estimation de rapport est 775 148. Le chiffre qui s'en rapproche le plus est 750 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 17,7 %.
- 4) Le dénominateur de cette estimation de rapport est 1 145 261. Le chiffre qui s'en rapproche le plus est 1 000 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 15,3 %.
- 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 :



$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0,177)^2 + (0,153)^2} \\ &= \sqrt{0,031329 + 0,023409} \\ &= 0,23396\end{aligned}$$

- 6) Le rapport obtenu entre les femmes et les hommes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années est 775 148 / 1 145 261, c'est-à-dire 0,677 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1). Le coefficient de variation de cette estimation est 23,4 %, elle doit donc être considérée de qualité médiocre, même si on peut la diffuser. Elle devrait donc être signalées par la lettre E (ou un quelconque identificateur similaire) et être accompagnées d'une mise en garde avertissant les utilisateurs subséquents des niveaux élevés d'erreur associés à l'estimation.

### Exemple 5 : Estimations de différences de rapports

Supposons qu'un utilisateur estime que le rapport entre les femmes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années et les hommes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années s'établit à 1,06 pour le Québec tandis qu'il est de 0,71 pour les provinces de l'Atlantique. L'utilisateur est intéressé à comparer les deux rapports pour voir s'il y a une différence statistique entre ceux-ci. Comment l'utilisateur détermine-t-il le coefficient de variation de cette différence?

- 1) Tout d'abord calculez le coefficient de variation approximatif pour le rapport du Québec ( $\hat{R}_1$ ) et le rapport des provinces de l'Atlantique ( $\hat{R}_2$ ) tel qu'il est décrit dans l'exemple 4. Le CV approximatif pour le rapport du Québec est 29,4 % et 29,9 % pour celui des provinces de l'Atlantique.

- 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  $\hat{d} = 1,06 - 0,71 = 0,35$  est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(1,06)(0,294)]^2 + [(0,71)(0,299)]^2} \\ &= \sqrt{(0,097) + (0,045)} \\ &= 0,377\end{aligned}$$

- 3) Le coefficient de variation de  $\hat{d}$  est donné par  $\sigma_{\hat{d}} / \hat{d} = 0,377 / 0,35 = 1,077$ .
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 107,7 %. La différence entre les estimations est considérée inacceptable et Statistique Canada recommande de ne pas la publier. Cependant, si l'utilisateur

choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre F (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les utilisateurs subséquents en garde contre les hauts taux d'erreur associés à l'estimation.

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

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

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

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

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

Des intervalles de confiance pour une estimation peuvent être calculés directement à partir des tables de variabilité d'échantillonnage approximative, en déterminant d'abord à partir de la table appropriée le coefficient de variation de l'estimation  $\hat{X}$ , puis en utilisant la formule suivante pour le convertir à un intervalle de confiance ( $IC_{\hat{x}}$ ) :

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

où  $\alpha_{\hat{x}}$  est le coefficient de variation déterminé de  $\hat{X}$ , et

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

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

confiance.

### **10.2.1 Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance**

Un intervalle de confiance de 95 % pour la proportion estimée de personnes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années (d'après l'exemple 2 à la section 10.1.1 serait calculé comme suit :

$$\hat{X} = 42,7 \% \text{ (ou exprimé sous forme de proportion } 0,427)$$

$$t = 2$$

$$\alpha_{\hat{x}} = 8,6 \% \text{ (0,086 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.}$$

$$IC_{\hat{x}} = \{0,427 - (2) (0,427) (0,086), 0,427 + (2) (0,427) (0,086)\}$$

$$IC_{\hat{x}} = \{0,427 - 0,073, 0,427 + 0,073\}$$

$$IC_{\hat{x}} = \{0,354, 0,500\}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 35,4 % et 50,0 % des personnes ont au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années.

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

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

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

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

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

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

Supposons que l'utilisateur désire tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y a pas de différence entre la proportion de femmes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années et la proportion d'hommes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années. D'après l'exemple 3 à la section 10.1.1, il s'est avéré que l'erreur-type de la différence entre ces deux estimations était 0,0745. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_d} = \frac{0,389 - 0,458}{0,0745} = \frac{-0,069}{0,0745} = -0,926$$

Puisque  $t = -0,926$  est compris entre -2 et 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**

Il faudrait produire des tables spéciales afin de déterminer l'erreur d'échantillonnage d'estimations quantitatives, ce qui n'a pas été fait, parce que la plupart des variables pour l'EACA sont principalement de nature catégorique.

En général cependant, le coefficient de variation d'un total quantitatif sera supérieur au coefficient de variation de l'estimation de la catégorie correspondante (c'est-à-dire l'estimation du nombre de personnes retenues dans l'estimation quantitative). S'il est impossible de diffuser l'estimation de la catégorie correspondante, on ne pourra pas non plus communiquer l'estimation quantitative. Par exemple, le coefficient de variation du nombre d'enfants des personnes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années serait supérieur au coefficient de variation de la proportion correspondante des personnes ayant au moins un enfant âgé entre 0 et 8 ans qui a fréquenté un service de garde régulier, incluant une garderie au cours des cinq dernières années. Si, par conséquent, le coefficient de variation de la proportion est inacceptable (rendant la proportion non diffusable), il en sera de même du coefficient de variation de l'estimation quantitative correspondante (rendant cette estimation quantitative non diffusable).

Des coefficients de variation de telles estimations peuvent être calculés, au besoin, pour une estimation précise à l'aide d'une technique appelée pseudo-répétition, ce qui veut dire diviser les enregistrements inclus dans les fichiers de microdonnées en sous-groupes (ou répétitions) et déterminer la variation à l'intérieur de l'estimation de répétition en répétition. Les utilisateurs qui désirent calculer des coefficients de variation pour des estimations quantitatives peuvent communiquer avec Statistique Canada afin d'obtenir des conseils sur l'allocation d'enregistrements à des répétitions appropriées et sur les formules à employer à l'intérieur de ces calculs.

## **10.5 Tables des coefficients de variation**

Consulter le fichier EACA2008\_CVTabF.pdf pour les tables de coefficients de variation.

## **11.0 Pondération**

Puisque l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage (EACA) 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 la non-réponse et le facteur d'ajustement du rapport province-âge-sexe et du niveau infraprovincial. Chacun de ces facteurs est décrit ci-dessous.

#### **Poids de base**

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

#### **Sous-poids d'une grappe**

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

#### **Poids de stabilisation**

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

### **Non-réponse**

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

Dans d'autres cas, on contrebalance les poids de la non-réponse 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-réponse, qui sont définis par la région économique de l'assurance-emploi, le type de secteur et le groupe de renouvellement. L'ajustement repose sur l'hypothèse voulant que les ménages qui ont été interviewés représentent les caractéristiques de ceux qui auraient dû être interviewés se trouvant à l'intérieur d'un secteur de non-réponse.

### **Sous-poids de l'Enquête sur la population active**

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

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

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

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

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

## **11.2 Procédures de pondération pour l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage**

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

- 1) Probabilité de sélection

Le premier ajustement consiste à ajuster les poids des ménages sélectionnés pour faire partie de l'échantillon de l'EACA de façon à leur attribuer les poids des ménages non sélectionnés. Pour cela on multiplie à l'intérieur de chaque strate les poids des ménages sélectionnés par un facteur égal à l'inverse de leur probabilité de sélection, et on donne un poids égal à 0 aux autres ménages.

2) Nombre de personnes dans le ménage

Le deuxième ajustement est destiné à tenir compte de la taille des ménages sélectionnés afin de produire des estimations en termes de personnes, et non en termes de ménages. On multiplie pour cela le poids obtenu à l'étape précédente par le nombre de personnes appartenant à la population-cible présentes dans le ménage.

3) Nombre de groupes de rotation retenus

Cet ajustement ne s'applique qu'aux individus sélectionnés dans les provinces. Trois groupes de rotation, c'est-à-dire la moitié des six groupes de rotation nécessaire, de l'EPA ont été utilisés pour construire l'échantillon de l'EACA dans les provinces. Le troisième ajustement consiste donc à tenir compte de cette sélection en multipliant les poids obtenus à l'étape précédente par 2.

4) Ajustement pour la non-réponse

Le quatrième ajustement est destiné à tenir compte de la non-réponse survenue pendant l'enquête. Pour cela, on commence par former des groupes d'individus (répondants et non-répondants) présentant des probabilités de réponse similaires. Par la suite, dans chacun des groupes ainsi formés, on multiplie les poids des personnes ayant répondu par un facteur égale à :

$$\frac{\text{somme des poids des répondants} + \text{somme des poids des non-répondants}}{\text{somme des poids des répondants}}$$

5) Post-stratification

Le dernier ajustement est destiné à faire coïncider certains totaux produits au moyen des données recueillies lors de l'enquête à des estimations démographiques connues. Les répondants sont regroupés selon les variables province, groupe d'âge et sexe. Dans chaque catégorie, on multiplie les poids obtenus à l'étape précédentes par un facteur égal à :

$$\frac{\text{taille de la population dans la catégorie donnée}}{\text{somme des poids des répondants appartenant à cette catégorie}}$$

Le poids qui en résulte WTPS est le poids final qui figure dans le fichier de microdonnées de l'EACA.





## **12.0 Questionnaire**

Le questionnaire de l'Enquête sur les attitudes des Canadiens à l'égard de l'apprentissage (EACA) a servi à recueillir des données en avril et mai 2008. Le fichier EACA2008\_QuestF.pdf renferme le questionnaire français.



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

Consulter le fichier EACA2008\_Partagé\_LvCds.pdf pour le cliché d'enregistrement à chiffres univariés.