

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

## **Enquête sur l'accès et le soutien à l'éducation et à la formation**

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>15</b>
5.1	Population cible.....	15
5.2	Base de sondage .....	15
5.2.1	Couverture de la population.....	15
5.3	Plan d'échantillonnage.....	15
5.3.1	Plan à deux phases de la partie Recensement .....	15
5.3.2	Stratification .....	16
5.3.3	Sélection des numéros de téléphone .....	16
5.3.4	Sélection des personnes.....	17
5.4	Taille de l'échantillon.....	17
<b>6.0</b>	<b>Collecte des données</b> .....	<b>19</b>
6.1	Supervision et contrôle de qualité.....	20
6.2	Non-réponse .....	21
<b>7.0</b>	<b>Traitement des données</b> .....	<b>23</b>
7.1	Saisie des données.....	24
7.2	Vérification .....	24
7.3	Codage des questions ouvertes .....	25
7.4	Imputation .....	25
7.5	Création de variables dérivées .....	26
7.6	Pondération.....	32
<b>8.0</b>	<b>Qualité des données</b> .....	<b>35</b>
8.1	Taux de réponse et taux de succès .....	35
8.2	Erreurs relatives à l'enquête .....	36
8.2.1	Base de sondage .....	36
8.2.2	Collecte des données .....	37
8.2.3	Traitement des données .....	38
8.2.4	Non-réponse .....	38
8.2.5	Mesure de l'erreur d'échantillonnage.....	40
<b>9.0</b>	<b>Lignes directrices pour la totalisation, l'analyse et la diffusion de données</b> .....	<b>43</b>
9.1	Lignes directrices pour l'arrondissement d'estimations .....	43
9.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation .....	44
9.3	Définitions de types d'estimations : catégoriques et quantitatives .....	44
9.3.1	Estimations catégoriques.....	44
9.3.2	Estimations quantitatives .....	44
9.3.3	Totalisation d'estimations catégoriques.....	45
9.3.4	Totalisation d'estimations quantitatives .....	45
9.4	Lignes directrices pour l'analyse statistique .....	46
9.5	Lignes directrices pour la diffusion de coefficients de variation.....	47
9.6	Seuils pour la diffusion des estimations.....	49

<b>10.0</b>	<b>Tables de variabilité d'échantillonnage approximative .....</b>	<b>51</b>
10.1	Comment utiliser les tables de coefficients de variation pour des estimations catégoriques.....	52
10.1.1	Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques .....	54
10.2	Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance.....	58
10.2.1	Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance .....	59
10.3	Comment utiliser les tables de coefficients de variation pour effectuer un test t.....	60
10.3.1	Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t .....	60
10.4	Coefficients de variation pour des estimations quantitatives.....	60
10.5	Tables des coefficients de variation.....	61
<b>11.0</b>	<b>Pondération .....</b>	<b>63</b>
<b>12.0</b>	<b>Questionnaire .....</b>	<b>67</b>
<b>13.0</b>	<b>Cliché d'enregistrement à valeurs univariées .....</b>	<b>69</b>

## **1.0 Introduction**

L'Enquête sur l'accès et le soutien à l'éducation et à la formation (EASEF) a été menée par Statistique Canada de juin à octobre 2008 avec la collaboration et l'appui de Ressources humaines et Développement des compétences Canada (RHDC). Le présent manuel a été produit pour faciliter la manipulation du fichier de microdonnées sur les résultats de l'enquête.

Toutes questions au sujet de l'ensemble de données ou de son utilisation devraient être adressées à :

### Statistique Canada

Division du tourisme et du Centre de la statistique de l'éducation

Services à la clientèle

Téléphone : 1-800-307-3382; 613-951-7608

Télécopieur : 613-951-4441

Courriel : [educationstats@statcan.gc.ca](mailto:educationstats@statcan.gc.ca)



## 2.0 Contexte

L'Enquête sur l'accès et le soutien à l'éducation et à la formation (EASEF) est une nouvelle enquête à participation volontaire réalisée auprès d'environ 72 000 ménages. L'objectif de l'enquête est d'évaluer la demande d'éducation et de formation au Canada dans le contexte de l'éducation permanente. L'EASEF a été financée par RHDCC.

L'enquête principale porte sur des enfants (âgés de 0 à 17 ans) pour lesquels la « personne la mieux renseignée » (PMR1) a été interviewée ou sur des répondants principaux (âgés de 18 à 64 ans) auprès desquels l'information a été recueillie directement. Un répondant principal est soit un jeune (âgé de 18 à 24 ans) ou un adulte (âgé de 25 à 64 ans). La collecte des données de l'enquête principale s'est déroulée du 16 juin au 6 octobre 2008. Un ensemble secondaire d'interviews a débuté le 11 juillet 2008 et s'est terminé le 10 octobre 2008 en vue de recueillir des renseignements auprès de la « personne la mieux renseignée » (PMR2) sur les jeunes. On a demandé aux jeunes et à la PMR2 qui avait répondu pour eux la permission de procéder au couplage de leurs réponses respectives à l'enquête.

L'EASEF regroupe trois anciennes enquêtes qui portaient sur les antécédents et les déterminants de l'accès à l'éducation postsecondaire, le rôle des prêts étudiants et de l'épargne dans le financement de l'éducation postsecondaire et la participation à des programmes d'éducation et de formation des adultes. L'EASEF remplace l'Enquête sur les approches en matière de planification des études (EAPE), l'Enquête sur la participation aux études postsecondaires (EPEP) et l'Enquête sur l'éducation et sur la formation des adultes (EEFA), qui ont été réalisées pour la dernière fois en 2002 et en 2003. Le regroupement de ces trois enquêtes permet de réaliser des économies d'échelle en ce qui concerne la gestion, l'élaboration et la mise en œuvre de l'enquête, ainsi que la diffusion des données.

L'EASEF accroît le potentiel analytique des trois enquêtes isolées, en ce qui concerne le contenu qui peut être abordé, les tailles d'échantillons permises compte tenu du budget et la possibilité d'étudier les corrélats pour l'ensemble des trois thèmes dans le contexte de l'éducation permanente. Les données recueillies dans le cadre de l'EASEF permettront de faire le suivi de l'état de préparation et de l'accès à l'éducation, d'évaluer l'efficacité des programmes publics reliés à l'éducation et d'élaborer des politiques pour répondre aux besoins de formation des Canadiens.



### 3.0 Objectifs

L'enquête mesure et fournit des indicateurs de:

- espoirs fondés par les parents/tuteurs sur l'éducation de leurs enfants
- valeurs et attitudes des parents/tuteurs face à l'éducation en général, et face aux études postsecondaires en particulier
- proportion de la population d'enfants âgés de 0 à 17 ans dont les parents/tuteurs ou autres membres de la famille épargnent actuellement en vue de leurs études postsecondaires
- proportion de la population d'enfants âgés de 0 à 17 ans dont les parents/tuteurs ont l'intention de commencer à épargner à une date ultérieure
- stratégies d'épargne
- cotisations au régime enregistré d'épargne-études (REEE)
- participation éventuelle de l'enfant au coût de ses études postsecondaires (salaires perçus au cours des études secondaires ou postsecondaires, prêts, subventions, bourses d'études, etc.)
- caractéristiques des familles qui n'épargnent pas et n'épargneront pas pour les études postsecondaires de leurs enfants, ainsi que la raison principale de cette absence d'épargne
- accès aux études postsecondaires
- incidence des prêts étudiants sur l'accès aux études postsecondaires
- incidence du statut socioéconomique des parents sur l'accès aux études postsecondaires
- caractéristiques du type de programme d'études postsecondaires poursuivies
- mécanismes de financement des études postsecondaires par les étudiants
- accès aux prêts étudiants, utilisation qui est faite de ces prêts et endettement des étudiants qui ont entrepris des études postsecondaires
- sensibilisation au Programme canadien pour l'épargne-études
- adéquation du financement des études en examinant les frais de scolarité, les autres coûts liés aux études et les principales dépenses mensuelles des personnes qui sont actuellement aux études
- profil sociodémographique des personnes qui ne poursuivent pas d'études postsecondaires, qui ont poursuivi ou qui poursuivent des études postsecondaires et qui ont reçu un prêt étudiant ou qui ont poursuivi ou qui poursuivent des études postsecondaires et qui n'ont pas reçu un prêt étudiant
- incidence et intensité de la participation des adultes à la formation formelle liée à un emploi
- profil du soutien des employeurs pour la formation formelle liée à un emploi
- aspects des activités de formation liée à un emploi, par exemple : les fournisseurs de la formation, les dépenses, le soutien financier, les motivations, les résultats et les difficultés qui se sont présentées pendant la formation
- obstacles qui empêchent des personnes de participer à la formation formelle qu'elles voudraient ou devraient suivre
- raisons qui expliquent le manque d'intérêt et de participation de la part des adultes à l'égard de la formation formelle



## 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 utilisé.

### **Bénévolat**

Le bénévolat se définit comme des activités non rémunérées effectuées pour des groupes ou des organismes tels que des écoles, des organismes religieux, des oeuvres de bienfaisance et des associations communautaires.

### **Bon d'études canadien (BEC)**

Une subvention versée par le gouvernement du Canada pour aider les familles à faible revenu à économiser pour les études postsecondaires de leurs enfants.

### **Bourse**

Somme allouée fondée sur des besoins financiers destinée à aider un étudiant à poursuivre ses études.

### **Bourse d'excellence, récompense ou prix**

Récompense monétaire, attribuée en général en raison des résultats scolaires exceptionnels, qui vise à aider un étudiant à poursuivre ses études.

### **Catégorie de la famille**

Catégorie d'immigrants parrainés par de proches parents ou des membres de la famille vivant au Canada.

### **Catégorie des immigrants économiques**

Catégorie d'immigrants sélectionnés en fonction de leurs compétences ou d'autres atouts permettant de contribuer à l'économie canadienne (comprend les travailleurs qualifiés, les investisseurs, les entrepreneurs et les travailleurs autonomes).

### **Catégorie des immigrants indépendants**

Catégorie d'immigrants qui possèdent les compétences nécessaires à l'exercice de certains emplois ou qui représentent un atout important pour le Canada. Ces personnes présentent une demande de leur propre chef ou ont des parents éloignés vivant au Canada.

### **Catégorie des réfugiés**

Catégorie de personnes qui demandent la protection du Canada.

### **Certificat de placement garanti (CPG)**

Titre attestant qu'une somme a été placée dans un établissement financier à un taux d'intérêt prédéterminé pour une durée déterminée habituellement entre un et cinq ans. En règle générale, l'investisseur ne peut exiger que l'établissement financier lui rembourse son titre avant l'échéance.

### **Citoyen canadien**

Un citoyen canadien est une personne qui est née au Canada, qui est née à l'étranger d'un parent citoyen canadien ou qui a acquis la citoyenneté canadienne par naturalisation (p. ex. un immigrant reçu qui a demandé la citoyenneté canadienne et l'a obtenue).

### **Compte en fiducie**

Compte ouvert dans une banque, une coopérative de crédit ou une société en fiducie dans lequel les dépôts sont faits sous forme de fonds en fiducie destiné à un bénéficiaire, même s'il n'existe aucun contrat officiel de fiducie.

**Couvert(e) par une convention collective ou par un contrat de travail négocié par un syndicat (mais non membre d'un syndicat)**

Un grand nombre d'employés occupent des emplois couverts par un contrat négocié par un syndicat sans être, pour une raison ou une autre, membre d'un syndicat et sans avoir signé leur carte syndicale, par exemple pour des motifs personnels ou religieux.

**Crédit d'impôt pour frais de scolarité - Montant relatif aux études - Montant pour manuels**

Le but de ces crédits est d'aider l'étudiant à réduire son impôt sur le revenu. La fraction du crédit d'impôt pour études que n'utilise pas l'étudiant peut être transférée à son conjoint, à son père ou à sa mère ou à un grand-parent, ou encore reportée à une année future.

**Crédit d'impôt pour intérêts payés sur des prêts étudiants**

Un montant pour les intérêts payés sur des prêts étudiants peut être demandé seulement par l'étudiant ou une personne apparentée ayant payé ces intérêts.

**École de métiers ou de formation professionnelle ou institut de technologie subventionné par l'État**

La différence entre cette catégorie et la catégorie « institut d'enseignement privé ou école commerciale privée » réside essentiellement sur la différence entre la nature « publique » et « privée » de l'institut.

**Emploi permanent**

Un emploi permanent est celui qui devrait durer aussi longtemps que l'employé le désire et aussi longtemps que la conjoncture économique le permettra. Un emploi non permanent est celui dont la date de cessation est prédéterminée ou qui se terminera dès qu'un projet précis aura été complété.

**Employé rémunéré**

Une personne qui travaille pour d'autres (c.à d. travaille pour un employeur) et reçoit un traitement ou un salaire.

**Employés (Note sur les employés)**

Les travailleurs autonomes et leurs partenaires d'affaires ne comptent pas à titre d'employés de l'entreprise.

**Exemption d'intérêts**

Programme conçu dans le but d'aider temporairement les emprunteurs qui ont de la difficulté à rembourser leur prêt d'études. Au cours de cette exemption, l'emprunteur n'a pas à effectuer de paiements afin de rembourser le prêt et le gouvernement fédéral paie les intérêts sur le prêt d'études. L'exemption est normalement approuvée pour une période de six mois, jusqu'à un maximum de 30 mois.

**Fondation collective de bourses d'études**

Une compagnie qui propose des régimes d'investissement communs conçus pour aider à payer des études postsecondaires.

**Fonds en fiducie**

Le fonds en fiducie est composé d'éléments d'actif appartenant à une fiducie détenue par des fiduciaires pour le compte de bénéficiaires qui ont reçu de l'argent à certaines conditions.

**Frais de scolarité**

Il s'agit d'une somme d'argent exigée par un établissement en retour de l'instruction qu'il dispense.

**Frais obligatoires**

Ce sont des frais payés à l'établissement d'enseignement qui n'incluent pas les frais de scolarité. Ces frais obligatoires couvrent par exemple les frais d'administration, les frais d'entretien, l'accès à des services comme la bibliothèque, les laboratoires ou le gymnase, l'affiliation à l'association étudiante et autres frais variés qui doivent être payés à l'établissement d'enseignement.

### **Immigrant reçu**

Un immigrant reçu désigne une personne à laquelle les responsables de l'immigration ont accordé le droit de vivre en permanence au Canada, mais qui n'a pas encore obtenu la citoyenneté canadienne. Cette personne est désignée « résident permanent » en vertu de la Loi sur l'immigration.

### **Obligations d'épargne**

Il s'agit d'obligations vendues au grand public par les gouvernements et qui rapportent un taux d'intérêt variable.

### **Membre d'un syndicat**

Pour être membre d'un syndicat, le répondant doit avoir adhéré à un syndicat et avoir signé une carte syndicale.

### **Prêt étudiant gouvernemental**

Prêt étudiant accordé dans le cadre d'un programme fédéral ou provincial/territorial qui permet au répondant de poursuivre ses études. Depuis mars 2001, le Programme canadien de prêts aux étudiants est géré directement par le gouvernement du Canada, par l'intermédiaire du Centre de service national de prêts aux étudiants. Exemples de prêt étudiant : le Programme canadien de prêts aux étudiants (PCPE), le Régime d'aide financière aux étudiantes et étudiants de l'Ontario (RAFEO), le Service des prêts et bourses aux étudiants du Québec, le Programme de prêts aux étudiants de l'Alberta, etc.

### **Programme**

Un programme est un ensemble de cours choisis à partir d'un plan de cours, d'un répertoire ou d'une liste. Les cours faisant partie d'un programme sont généralement suivis pour acquérir des unités (crédits) en vue de l'obtention d'un grade, d'un diplôme ou d'un certificat. Un programme comprend un ensemble de cours suivis dans le but d'obtenir un certificat, un diplôme et/ou un grade d'études et doit prendre trois mois ou plus à compléter.

### **Programme canadien pour l'épargne-études (PCEE)**

Programme du gouvernement du Canada qui octroie une subvention directement au REEE d'un bénéficiaire. Le programme ajoute de 20 % à 40% aux contributions versées chaque année à un REEE au nom d'un bénéficiaire admissible. Avec le revenu accumulé de l'investissement, la subvention du Programme canadien pour l'épargne-études pourra être versée à un étudiant dans le cadre des paiements d'aide aux études au moment des études postsecondaires.

### **Réduction de la dette**

Une mesure de gestion de la dette conçue pour aider à administrer le remboursement des prêts d'études dans l'éventualité où des difficultés financières à long terme sont éprouvées.

### **Régime enregistré d'épargne-études (REEE)**

Un REEE est un moyen d'épargner à l'abri de l'impôt en vue de financer les études postsecondaires d'un enfant. Lorsque l'enfant commence ses études postsecondaires, l'épargne accumulée dans le régime est utilisée pour payer les frais de scolarité ou d'autres dépenses liées aux études.

### **Régime enregistré d'épargne-retraite (REER)**

Un REER permet d'accumuler des épargnes en vue de la retraite. Les cotisations à ce régime sont déductibles d'impôt, avec certaines limites.

### **Remise de dette**

Le Programme canadien de prêts aux étudiants (PCPE) n'a pas de politique de remise de prêt. Toutefois, la plupart des provinces offrent cette mesure de gestion de la dette.

### **Subvention**

Somme monétaire accordée par un gouvernement ou une société à un étudiant à la condition que ce dernier accepte certaines conditions ou respecte certains engagements.

**Subvention canadienne pour l'épargne-études (SCEE)**

Une subvention versée par le gouvernement du Canada afin d'encourager les parents, la parenté et les amis à économiser en prévision des études postsecondaires d'un enfant.

**Travailleur autonome**

Une personne qui est le propriétaire et l'exploitant d'une entreprise et qui est payée directement par le client.

**Travailleur familial non rémunéré**

Une personne qui travaille sans salaire sur une ferme ou dans une entreprise qui est possédée ou exploitée par une personne apparentée vivant dans le même ménage.

## **5.0 Méthodologie de l'enquête**

Le présent chapitre décrit la population cible, la base de sondage et le plan d'échantillonnage de l'Enquête sur l'accès et le soutien à l'éducation et à la formation (EASEF).

### **5.1 Population cible**

La population cible de l'EASEF comprend toutes les personnes résidant au Canada de moins de 65 ans, excepté celles habitant dans les trois territoires du Nord et celles résidant en établissement.

### **5.2 Base de sondage**

La base de sondage de l'EASEF est une liste de numéros de téléphone créée en se servant de deux sources de numéros:

- les numéros de téléphone du Recensement de la population du Canada de 2006 et
- une liste de numéros de téléphone « résidentiels » compilée en se servant des fichiers administratifs des compagnies de téléphone.

La base de sondage a été construite en plusieurs étapes :

- obtenir un fichier contenant tous les ménages du Recensement
- éliminer les ménages du recensement n'ayant pas de numéro de téléphone valide (7 % des ménages), ceux ayant des numéros de téléphone en double (1,5 %) et ceux pour lesquels des dates de naissance étaient imputées (2 %).
- ajouter les numéros de téléphone provenant des fichiers administratifs : seuls les numéros de téléphone qui manquaient dans la partie Recensement de la base de sondage ont été ajoutés.

La base de sondage de l'EASEF consiste en une liste de numéros de téléphone distincts sans numéros en double. Elle peut par conséquent être traitée comme une base de sondage unique (par opposition à une base de sondage double) même si elle a été créée en se servant de deux sources.

#### **5.2.1 Couverture de la population**

La population étudiée dans le cadre de l'EASEF ne comprend pas les personnes n'ayant pas de téléphone ou possédant seulement un téléphone mobile. En outre, ne font pas non plus partie du champ de l'enquête les personnes dont le numéro de téléphone courant appartenait à un ménage dont les membres étaient tous âgés de 65 ans et plus selon le recensement. Enfin, la base de sondage ne couvre pas les ménages dont le numéro de téléphone ne figure ni sur la liste de numéros de téléphone du Recensement de 2006 ni dans les fichiers administratifs.

La couverture de la base de sondage est discutée plus en détail à la section 8.2.1.

### **5.3 Plan d'échantillonnage**

#### **5.3.1 Plan à deux phases de la partie Recensement**

Le Recensement de 2006 a été réalisé en utilisant un questionnaire abrégé (formule 2A) et un questionnaire complet (formule 2B/2D). Dans la plupart des régions du pays, un ménage sur cinq a été sélectionné au hasard pour recevoir le questionnaire 2B/2D; les autres ménages ont reçu le questionnaire 2A. La partie Recensement de l'échantillon de

l'EASEF est un sous-échantillon de l'échantillon 2B/2D du recensement. Cette partie de l'échantillon correspond donc à un plan à deux phases, la première phase étant l'échantillon 2B/2D du recensement et la deuxième, le sous-échantillon de l'EASEF.

Le tirage d'un sous-échantillon parmi les ménages du recensement qui ont rempli le questionnaire complet 2B/2D offrait l'avantage de disposer d'un plus grand nombre de variables du recensement pour procéder à la repondération pour corriger la non-réponse.

### 5.3.2 Stratification

Les numéros de téléphone qui figurent dans la base de sondage de l'EASEF ont été stratifiés selon la province, et chaque province a été stratifiée à son tour en six strates, comme le définit le tableau suivant :

Source du numéro de téléphone	Nom de la strate	Description
Recensement	Jeunes seulement	Ménages composés de jeunes (18 à 24 ans) seulement
	Jeunes, mixte	Ménages composés de jeunes et d'autres groupes d'âge
	Enfants, mixte	Ménages composés d'enfants (0 à 17 ans) et d'aucun jeune
	Adultes seulement	Ménages composés d'adultes (25 à 64 ans) seulement
	Personnes âgées seulement	Ménages composés de personnes âgées (65 ans et plus) seulement
Administrative	Administrative	Numéros de téléphone tirés de fichiers administratifs

Les cinq premières strates (jeunes seulement, jeunes mixte, enfants mixte, adultes seulement et personnes âgées seulement) sont des strates formées pour les numéros de téléphone du recensement. La date de naissance du fichier de recensement a été utilisée pour calculer l'âge de chaque membre du ménage et pour classer les ménages dans l'une des cinq strates.

La strate « Administrative » est une strate distincte pour les numéros de téléphone tirés des fichiers administratifs.

### 5.3.3 Sélection des numéros de téléphone

Pour la partie Recensement de la base de sondage, l'échantillon de numéros de téléphone a été sélectionné selon un plan d'échantillonnage systématique avec probabilité proportionnelle à la taille (PPT) dans chaque strate issue de l'échantillon 2B du recensement. La mesure de taille utilisée était le poids de recensement du ménage. L'échantillonnage PPT a été utilisé afin que tous les ménages compris dans une même strate aient le même poids de sondage.

Pour la strate « Administrative », les numéros de téléphone ont été sélectionnés selon un plan d'échantillonnage aléatoire simple sans remise (EASSR).

Les numéros de téléphone de l'échantillon sélectionné ont été envoyés sur le terrain pour la collecte. Le ménage rejoint en composant le numéro de téléphone sélectionné a été interviewé, qu'il s'agisse ou non du même ménage que celui enregistré dans la base de sondage ou qu'il ait encore ou non la même composition d'âge.

### 5.3.4 Sélection des personnes

Une fois le contact établi avec un ménage, la liste de tous les membres du ménage, y compris leur âge, était dressée et un membre du ménage dont l'âge était compris entre 0 et 64 ans était sélectionné au hasard pour participer à l'enquête. Afin de répondre aux exigences de l'enquête, la probabilité de sélection dépendait de l'âge : les jeunes de 18 à 24 ans étaient huit fois plus susceptibles d'être sélectionnés que les adultes de 25 à 64 ans, et cinq fois plus susceptibles que les enfants de 0 à 17 ans. Si tous les membres du ménage étaient âgés de 65 ans et plus, aucun membre du ménage n'était sélectionné et l'interview prenait fin.

## 5.4 Taille de l'échantillon

L'échantillon de l'EASEF comprenait environ 72 000 numéros de téléphone. Il a été réparti entre les strates de type de ménage de la façon suivante :

Strate	Taille de l'échantillon	Fraction d'échantillonnage
Jeunes seulement	1 360	0,0108
Jeunes, mixte	19 640	0,0108
Enfants, mixte	17 000	0,0072
Adultes seulement	14 000	0,0030
Personnes âgées seulement	0	0,0000
Administrative	20 000	0,0070
Total	72 000	

Comme l'indique le tableau, les ménages comprenant des jeunes selon les données du recensement ont été suréchantillonnés, tandis que ceux composés d'adultes seulement ont été sous-échantillonnés. Les ménages composés uniquement de personnes âgées n'ont pas été échantillonnés. La répartition de l'échantillon a été dictée par les exigences concernant la taille d'échantillon pour les divers groupes d'âge.

La méthode de répartition de Kish a été utilisée pour répartir l'échantillon entre les provinces. Une valeur de 0,67 a été choisie pour le paramètre de Kish, ce qui donne une répartition représentant un compromis entre une répartition égale et une répartition proportionnelle. Voici les tailles d'échantillon par province:

<b>Province</b>	<b>Taille de l'échantillon</b>
Terre-Neuve-et-Labrador	3 811
Île-du-Prince-Édouard	3 725
Nouvelle-Écosse	4 004
Nouveau-Brunswick	3 902
Québec	12 958
Ontario	20 864
Manitoba	4 140
Saskatchewan	4 022
Alberta	6 879
Colombie-Britannique	7 695
<b>Canada</b>	<b>72 000</b>

## 6.0 Collecte des données

La collecte des données de l'Enquête sur l'accès et le soutien à l'éducation et à la formation (EASEF) a eu lieu entre le 16 juin et le 10 octobre 2008 et a été effectuée par le biais d'interviews téléphoniques assistées par ordinateur (ITAO). Une lettre de présentation a été envoyée aux ménages sélectionnés environ deux semaines avant le début de la collecte.

Le système d'ITAO comprend un certain nombre de modules génériques qui peuvent être adaptés rapidement à la plupart des types d'enquêtes. Un module frontal contient un ensemble de codes de réponse standard qui permettent de traiter tous les résultats possibles des appels, ainsi que les scénarios correspondants qui doivent être lus par les intervieweurs.

Une approche standard a été utilisée pour présenter l'organisme, le nom et le but de l'enquête, les commanditaires de l'enquête, l'usage qui sera fait des résultats et la durée de l'interview. Nous avons expliqué aux répondants comment ils avaient été sélectionnés pour participer à l'enquête, que leur participation était volontaire et que l'information qu'ils fourniraient demeurerait strictement confidentielle. Des écrans d'aide étaient fournis aux intervieweurs pour les aider à répondre aux questions qui sont fréquemment posées par les répondants.

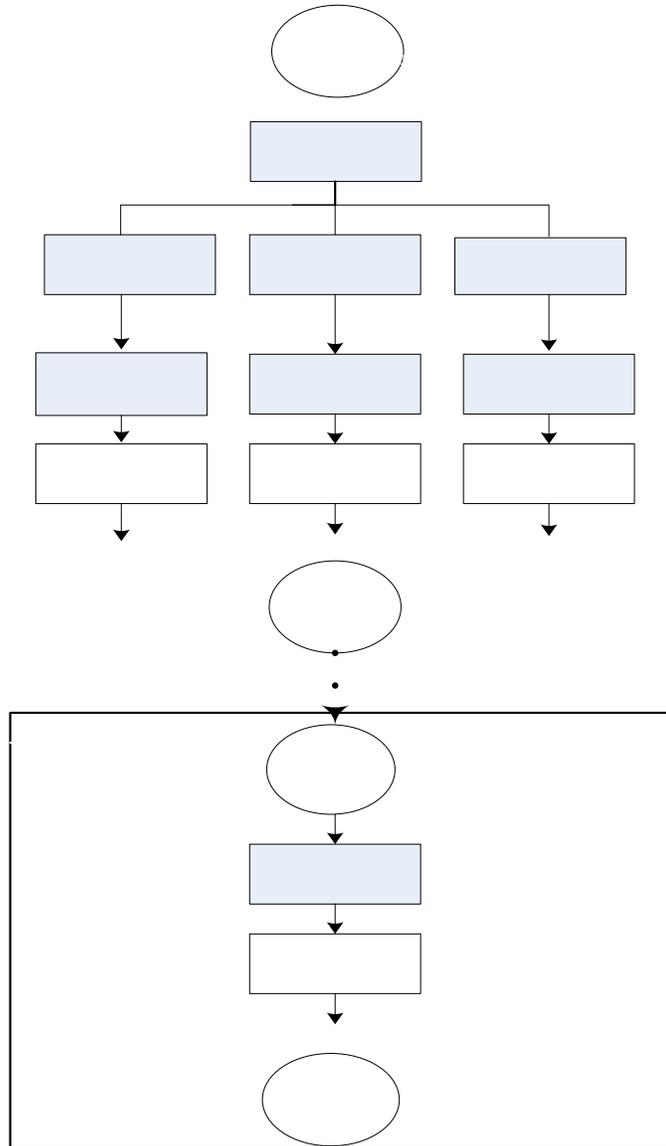
La collecte des données de l'EASEF a été exécutée en deux phases. L'information recueillie à la phase 1 portait sur les enfants sélectionnés (âgés de 0 à 17 ans), pour lesquels la « personne la mieux renseignée » (PMR1) était interviewée, ou sur les répondants principaux (âgés de 18 à 64 ans), auprès desquels l'information était recueillie directement. À la phase 2, des renseignements étaient recueillis auprès de la « personne la mieux renseignée » (PMR2) au sujet des répondants principaux âgés de 18 à 24 ans. Chaque phase de collecte a été réalisée en se servant d'applications d'ITAO différentes.

Au début de la phase 1, des renseignements socio-démographiques généraux étaient recueillis sur tous les membres des ménages sélectionnés auprès d'un membre adulte de ces ménages. Appelée réponse « par procuration », cette façon de procéder est appliquée parce qu'il serait trop coûteux et trop long de faire les appels répétés nécessaires pour obtenir l'information directement auprès de chaque répondant.

Une fois le module socio-démographique complété, une sélection au hasard était effectuée pour la phase 1 (EASEF principale) et l'intervieweur demandait à parler à la personne sélectionnée ou à la PMR1 (lorsqu'un enfant avait été sélectionné). Si cette personne n'était pas disponible, l'intervieweur fixait un moment opportun pour rappeler.

À la fin de l'interview d'un jeune, nous demandions à ce dernier de fournir les coordonnées d'une PMR2, c'est-à-dire la personne la mieux renseignée au sujet de l'état de préparation financière pour que ce jeune poursuive des études postsecondaires. La phase 2 de la collecte consistait à administrer le questionnaire de l'EASEF aux PMR2 après les avoir rejointes par téléphone.

La nécessité d'une deuxième phase a comporté des difficultés supplémentaires pour la programmation et la mise en œuvre de l'application ITAO pendant la collecte des données. Certains renseignements obtenus à la phase 1 devaient être disponibles pour la phase 2 et étaient transférés par le biais du fichier de l'échantillon. Ce processus n'a pas toujours réussi (voir 8.2.3 sur le Traitement des données).



La collecte des données a été effectuée par des employés spécialisés des bureaux de Statistique Canada d'Edmonton, de Sturgeon Falls, de Toronto, de Halifax, de Winnipeg et de Sherbrooke. Dans chaque bureau, un gestionnaire de projet gérait le personnel chargé des interviews et leur charge de travail. L'ordonnanceur automatisé utilisé par le système d'ITAO faisait en sorte que les cas soient affectés aléatoirement aux intervieweurs et que, pour ces cas, des appels soient faits à différents moments de la journée et différents jours de la semaine afin de maximiser la probabilité d'une prise de contact. Le nombre maximal de tentatives d'appel par cas était fixé à 25. Une fois le maximum atteint, un intervieweur principal passait le cas en revue et déterminait si d'autres appels seraient effectués.

## 6.1 Supervision et contrôle de qualité

Tous les intervieweurs de l'EASEF travaillaient sous la supervision d'un groupe d'intervieweurs principaux, qui avaient pour responsabilité de s'assurer que les intervieweurs connaissent bien les concepts et les procédures de l'EASEF et de contrôler régulièrement le travail des intervieweurs et d'examiner les documents complétés par ces derniers. Les intervieweurs principaux étaient, à leur tour, supervisés par les gestionnaires du programme de l'EASEF,

affectés dans chacun des bureaux régionaux de Statistique Canada.

## **6.2 Non-réponse**

Les intervieweurs avaient pour consigne de faire toutes les tentatives raisonnables en vue de réaliser les interviews de l'EASEF auprès des membres des ménages admissibles. Dans les cas où une personne refusait, au départ, de participer à l'EASEF, le bureau régional de Statistique Canada envoyait à l'adresse du logement une lettre insistant sur l'importance de l'enquête et de la collaboration du ménage. Dans les cas où l'intervieweur avait appelé à un mauvais moment, il prenait rendez-vous en vue de rappeler à un moment plus opportun. Dans les cas où il n'y avait personne à la maison, plusieurs appels ultérieurs étaient faits. Dans aucune circonstance les logements échantillonnés n'ont été remplacés par d'autres logements pour des raisons de non-réponse.

Afin d'accroître le taux de réponse, la période de collecte a été prolongée de deux semaines au-delà de la durée prévue au départ. Une carte postale attrayante a été créée et envoyée par la poste, du bureau central, aux ménages non répondants soulignant à quel point il était important de recueillir les données de l'enquête et invitant les ménages sélectionnés à participer à l'enquête. Bien que cet effort en vue d'obtenir des interviews semble avoir été fructueux, un certain nombre de ménages sont demeurés non répondants.



## 7.0 Traitement des données

Le principal produit de l'Enquête sur l'accès et le soutien à l'éducation et à la formation (EASEF) est un fichier de microdonnées « épuré ». Ce chapitre présente un bref résumé des phases de traitement inhérentes à la production de ce fichier.

Les données de sortie de l'EASEF ont été sauvegardées dans deux fichiers de données séparés : un fichier **principal** et un fichier du **cours**.

Le tableau ci-dessous énumère tous les modules de l'EASEF et indique la population comprise dans chaque module. Un répondant principal peut être un jeune (âgé de 18 à 24 ans) ou un adulte (âgé de 25 à 64 ans) et peut être identifié sur le fichier principal en fixant la valeur de la variable dérivée RESMR à 1. Les enregistrements ayant des données enfant/PMR1 peuvent être extraites en fixant RESPMK à 2, alors que RESPMK = 3 identifie les enregistrements ayant les données des jeunes et de leur PMR2. Le nom des variables (y compris les identificateurs d'imputation) pour lesquelles les renseignements concernent ou proviennent d'une PMR ont la lettre *P* à la fin. Par exemple, DWELCUR contient de l'information sur le type de logement de la résidence actuelle du répondant principal, alors que DWELCURP, qui a la même signification que DWELCUR, contient les données de la PMR1 ou PMR2.

Modules EASEF		Population		
Abbréviation	Titre des modules	Répondant principal	PMR1	PMR2
LP	Information apprentissage	X		
AL	Attitude à l'apprentissage	X	X	X
EM	Emploi le plus récent	X	X	X
SP	Bénévolat	X		
EH	Expériences au secondaire	X		
EA	Expériences au postsecondaire	X		
HE	Plus haut niveau de scolarité	X		
EC	Éducation formelle	X	EC_Q04	EC_Q04
MR	Programme plus récent	X		
EE	Durée éducation formelle	X		
EO	Autres caractéristiques éducation	X		
SE/FM	Soutien de l'employeur / Situation d'emploi	X		
CE	Coût de l'éducation formelle	X		
ES	Aide financière – épargnes	X		
ER	Prêt à rembourser éducation	X		
EN	Aide financière – non remboursable	X		
EJ	Étudiants actuels avec un prêt étudiant	X		
EK	Étudiants actuels sans prêt	X		
CR1/PN	Formation non formelle / Nom du cours	X		
CR2	Cours reliés à l'emploi	X		
CN	Cours reliés à l'emploi (sur fichier du cours)	X		
CO/FM	Coûts non formels / Situation d'emploi (sur fichier du cours)	X		
BE	Barrières	X		
FB	Information sur la famille	X	X	
CC	État de l'enfant		X	

Modules EASEF		Population		
Abbréviation	Titre des modules	Répondant principal	PMR1	PMR2
SA	Fréquentation scolaire de l'enfant		X	
CS	Expérience scolaire de l'enfant		X	
PH	Attentes et espérances des parents		X	X
PP	Participation parentale		X	
CT	Compétences en informatique	X	X	
IM	Immigration	X	X	
LG	Langue	X	X	X
LD	Incapacités	X		
FP	Planification financière		X	X
FO	Préparation financière au postsecondaire		X	X
SO	Parents qui épargnent pour d'autres enfants		X	X
SS	Épargnes des parents		X	X
SW	Retraits des épargnes		X	X
NS	Pas d'épargnes		X	X
PA	Sensibilisation d'épargnes		X	X
HL	Résidence principale	X	X	X
EL	Impact d'un prêt étudiant	X		
IE	Dépenses du ménage	X	X	X
ET	Aide financière	X	X	X
IN	Revenu	X	X	X

## 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. 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, parfois, subséquentes.

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**

Certains renseignements recueillis auprès des répondants doivent être codés afin que les données puissent être mieux analysées. Plusieurs systèmes de codage ont été utilisés pour l'EASEF.

#### **Classification des programmes d'enseignement (CPE)**

Le codage lié aux programmes d'enseignement a été effectué à l'aide des codes CPE (voir l'annexe A).

#### **Industrie et profession**

Le Système de classification des industries de l'Amérique du Nord (SCIAN) 2007 a été utilisé pour coder le genre d'entreprise (voir l'annexe B). La Classification nationale des professions pour statistiques (CNP-S) 2006 a été utilisée pour coder l'information sur le type de travail et l'activité principale des répondants (voir l'annexe C).

#### **Codes de pays**

La liste des codes de pays du Recensement de 2006 a été utilisée pour coder le pays d'origine (voir l'annexe D).

#### **Codes de langue**

La liste des codes de langue du Recensement de 2006 a été utilisée pour coder la langue parlée le plus souvent à la maison (voir l'annexe E).

#### **Origine ethnique**

La liste des codes d'origine ethnique du Recensement de 2006 a été utilisée pour coder l'origine ethnique et culturelle (voir l'annexe F).

#### **Autre - précisez**

Les questions accompagnées d'un choix de réponse contiennent souvent la catégorie finale « Autre - précisez ». Le texte qui y est inscrit est saisi. Ces réponses écrites sont examinées et peuvent être recodées dans l'une des catégories existantes. Si la réponse écrite correspond à l'une des catégories existantes de la question, celle-ci est établie à « Oui » et la catégorie « Autre - précisez » est établie à « Non ». De plus, de nouvelles catégories peuvent être ajoutées si un grand nombre de réponses écrites peuvent être regroupées.

### **7.4 Imputation**

L'imputation est le processus qui fournit des valeurs valides concernant les variables qui ont été retenues pour être modifiées, soit en raison de renseignements invalides, soit en raison de renseignements manquants. Les nouvelles valeurs sont établies de façon à préserver la structure sous-jacente des données et à garantir que les enregistrements qui en résultent passeront tous les contrôles requis. En d'autres mots, l'objectif n'est pas de reproduire les véritables valeurs des microdonnées mais plutôt d'établir des enregistrements intrinsèquement cohérents qui permettront de produire de bonnes estimations agrégées.

Nous pouvons faire la distinction entre trois types de non-réponse. La non-réponse complète est lorsque le répondant ne fournit pas le nombre minimal de réponses. Ces enregistrements sont supprimés et seront pris en compte lors du processus de pondération (voir le chapitre 11.0). La non-réponse ponctuelle est lorsque le répondant ne fournit pas une réponse à une question mais qu'il passe à la question suivante. Ce type de non-réponse est habituellement traité en utilisant le code « non déclaré » ou en ayant recours à l'imputation. Enfin, la non-réponse partielle est lorsque le répondant fournit le nombre minimal de réponses mais ne termine pas l'interview. Ces enregistrements peuvent être traités comme des cas de non-réponse complète ou ponctuelle.

Dans le cas de l'EASEF, l'imputation a été utilisée pour fournir les éléments d'information manquants aux enregistrements avec non-réponse ponctuelle. La méthode d'imputation par donneur a été utilisée pour la non-réponse complète des PMR2. De plus amples renseignements sur le processus d'imputation sont donnés au chapitre 8.0.

## 7.5 Création de variables dérivées

Un certain nombre d'éléments de données inclus dans le fichier de microdonnées ont été calculés en combinant des postes sur le questionnaire pour faciliter l'analyse des données. Ce qui suit est une liste des variables dérivées qui ont été créées. Veuillez vous référer au livre des codes pour une liste complète des variables dérivées et leurs valeurs.

VARIABLES D'IDENTIFICATION	
Variable	Description
RESMR	Type de répondant (jeune de 18 à 24 ans ou adulte de 25 à 64 ans)
RESPMK	Type de répondant (PMR de l'enfant de 0 à 17 ans ou PMR du jeune de 18 à 24 ans)
PMKHLD	Lieu de résidence de la PMR
MRSTERRP	Identificateur indiquant si le bon code a été utilisé pour l'état matrimonial de la PMR2 lors de la collecte

VARIABLES SOCIO-DÉMOGRAPHIQUES	
Variable	Description
DPROV	Code de la province pour l'adresse actuelle du répondant
REGION	Région dérivée de la province de résidence actuelle du répondant sélectionné
DWELCUR	Type de logement de la résidence actuelle du répondant principal
DWELCURP	Type de logement de la résidence actuelle de la PMR1 ou PMR2
DWELSTU	Type de logement de la résidence durant l'année scolaire
DVAGE	Âge du répondant sélectionné
AGEDEC	Âge du répondant sélectionné le 31 décembre 2007
IMMAGE	Âge du répondant sélectionné au moment de l'immigration
LANGUAGE	Langue parlée le plus souvent à la maison
RELSELP	Lien entre la PMR1 ou PMR2 et le répondant sélectionné
RELSELSP	Lien entre le conjoint ou partenaire de la PMR1 et le répondant

<b>VARIABLES SOCIO-DÉMOGRAPHIQUES</b>	
<b>Variable</b>	<b>Description</b>
CHDLT25	Nombre d'enfants de la PMR âgés de 24 ans ou moins qui habitent avec la PMR dans le ménage sélectionné
FAMSCHD	Structure de la famille de l'enfant sélectionné
FAMSTPA	Structure de la famille - passé
MOM_ED	Plus haut niveau de scolarité de la mère vivant dans le ménage sélectionné
DAD_ED	Plus haut niveau de scolarité du père vivant dans le ménage sélectionné
PAR_ED	Plus haut niveau de scolarité des parents vivant dans le ménage sélectionné
PMK1EDP	Plus haut niveau de scolarité de la PMR1
OTHERED	Plus haut niveau de scolarité de l'autre parent vivant à l'extérieur du ménage sélectionné
RESPINC	Salaire du répondant principal - groupé
RESPINCP	Salaire de la PMR - groupé
TOTINC	Salaires combinés du répondant principal et du conjoint ou partenaire - groupés
TOTINCP	Salaires combinés de la PMR et du conjoint ou partenaire - groupés

<b>CODAGE DE LA PROFESSION ET DE L'INDUSTRIE (FICHER PRINCIPAL)</b>	
<b>Variable</b>	<b>Description</b>
EM1CLSW	Catégorie de travailleur (répondant principal)
EM1CLSWP	Catégorie de travailleur (PMR)
EM2CLSWP	Catégorie de travailleur (conjoint ou partenaire de la PMR)
EM1COCC	Groupe de base pour le système de classification type des professions (au niveau de 4 chiffres) du répondant principal
EM1COCCP	Groupe de base pour le système de classification type des professions (au niveau de 4 chiffres) de la PMR
EM2COCCP	Groupe de base pour le système de classification type des professions (au niveau de 4 chiffres) du conjoint/partenaire de la PMR
EM1CIND	Groupes d'industrie (au niveau de 4 chiffres) du répondant principal
EM1CINDP	Groupes d'industrie (au niveau de 4 chiffres) de la PMR
EM2CINDP	Groupes d'industrie (au niveau de 4 chiffres) du conjoint/partenaire de la PMR
EM1NA18	Groupe d'industrie, répondant principal
EM1NA18P	Groupe d'industrie, PMR
EM2NA18P	Groupe d'industrie, conjoint/partenaire de la PMR
EM1NO2	Grand groupe des professions (agrégés aux 2 premiers chiffres du code CNP-S), répondant principal
EM1NO2P	Grand groupe des professions (agrégés aux 2 premiers chiffres du code CNP-S), PMR

<b>CODAGE DE LA PROFESSION ET DE L'INDUSTRIE (FICHER PRINCIPAL)</b>	
<b>Variable</b>	<b>Description</b>
EM2NO2P	Grand groupe des professions (agrégés aux 2 premiers chiffres du code CNP-S), conjoint/partenaire de la PMR
EM1NA2	Secteurs (agrégés aux 2 premiers chiffres du code SCIAN), répondant principal
EM1NA2P	Secteurs (agrégés aux 2 premiers chiffres du code SCIAN), PMR
EM2NA2P	Secteurs (agrégés aux 2 premiers chiffres du code SCIAN), conjoint/partenaire de la PMR
EM1NA3	Sous-secteurs (agrégés aux 3 premiers chiffres du code SCIAN), répondant principal
EM1NA3P	Sous-secteurs (agrégés aux 3 premiers chiffres du code SCIAN), PMR
EM2NA3P	Sous-secteurs (agrégés aux 3 premiers chiffres du code SCIAN), conjoint/partenaire de la PMR
EM1SECT	Secteur public ou privé, répondant principal
EM1SECTP	Secteur public ou privé, PMR
EM2SECTP	Secteur public ou privé, conjoint/partenaire de la PMR
EM1NO3	Sous-groupe des professions (agrégés aux 3 premiers chiffres du code CNP-S), répondant principal
EM1NO3P	Sous-groupe des professions (agrégés aux 3 premiers chiffres du code CNP-S), PMR
EM2NO3P	Sous-groupe des professions (agrégés aux 3 premiers chiffres du code CNP-S), conjoint/partenaire de la PMR
EM1NO10	Grande catégorie professionnelle, premier chiffre du code CNP-S, répondant principal
EM1NO10P	Grande catégorie professionnelle, premier chiffre du code CNP-S, PMR
EM2NO10P	Grande catégorie professionnelle, premier chiffre du code CNP-S, conjoint/partenaire de la PMR
FM1CIND	Groupes d'industrie (au niveau de 4 chiffres) - employeur qui offre du soutien pour le programme
FM1COCC	Groupe de base pour le système de classification type des professions (au niveau de 4 chiffres) - employeur qui offre du soutien pour le programme
FM1SECT	Secteur public ou privé - employeur qui offre du soutien pour le programme
FM1NA2	Secteurs (agrégés aux 2 premiers chiffres du code SCIAN) - employeur qui offre du soutien pour le programme
FM1NA3	Sous-secteurs (agrégés aux 3 premiers chiffres du code SCIAN) - employeur qui offre du soutien pour le programme
FM1NA18	Groupe d'industrie - employeur qui offre du soutien pour le programme
FM1NO2	Sous-groupe des professions (agrégés aux 2 premiers chiffres du code CNP-S) - employeur qui offre du soutien pour le programme
FM1NO3	Sous-groupe des professions (agrégés aux 3 premiers chiffres du code CNP-S) - employeur qui offre du soutien pour le programme
FM1NO10	Grande catégorie professionnelle, premier chiffre du code CNP-S - employeur qui offre du soutien pour le programme
FM2CIND	Groupes d'industrie (au niveau de 4 chiffres) - employeur qui offre du soutien pour le cours / séminaire / atelier

<b>CODAGE DE L'OCCUPATION ET DE L'INDUSTRIE (FICHIER DE COURS)</b>	
<b>Variable</b>	<b>Description</b>
FM2COCC	Groupe de base pour le système de classification type des professions (au niveau de 4 chiffres) - employeur qui offre du soutien pour le cours / séminaire / atelier
FM2SECT	L'emploi le plus récent appartient au secteur public ou privé - employeur qui offre du soutien pour le cours / séminaire / atelier
FM2NA2	Secteurs (agrégés aux 2 premiers chiffres du code SCIAN) - employeur qui offre du soutien pour le cours / séminaire / atelier
FM2NA3	Sous-secteurs (agrégés aux 3 premiers chiffres du code SCIAN) - employeur qui offre du soutien pour le cours / séminaire / atelier
FM2NA18	Groupe d'industrie - employeur qui offre du soutien pour le cours / séminaire / atelier
FM2NO2	Grand groupe des professions (agrégés aux 2 premiers chiffres du code CNP-S) - employeur qui offre du soutien pour le cours / séminaire / atelier
FM2NO3	Sous-groupe des professions (agrégés aux 3 premiers chiffres du code CNP-S) - employeur qui offre du soutien pour le cours / séminaire / atelier
FM2NO10	Grande catégorie professionnelle, premier chiffre du code CNP-S - employeur qui offre du soutien pour le cours / séminaire / atelier

<b>PLANIFICATION DE L'ÉDUCATION</b>	
<b>Variable</b>	<b>Description</b>
PAREXP	Indicateur pour désigner si le parent croit que l'enfant poursuivra des études postsecondaires
RESPELCP	Valeur totale des épargnes de la PMR dans un REEE pour l'enfant sélectionné à la fin de 2007
OTHSAV1P	Valeur totale des autres épargnes de la PMR pour l'enfant sélectionné à la fin de 2007
SELSAVOP	Indicateur de qui épargne actuellement pour les études postsecondaires de l'enfant sélectionné
SELWHPYP	Indicateur pour l'enfant sélectionné dont la PMR va payer ou aider à payer les études postsecondaires
SAVESTAP	Situation d'épargnes du ménage pour les études postsecondaires de l'enfant sélectionné
SAVESTOP	Situation d'épargnes du ménage pour les études postsecondaires des enfants autres que l'enfant sélectionné
SAVECHDP	Situation d'épargnes du ménage selon l'endroit où vivent les enfants
RESPSTAP	Situation d'épargnes dans un REEE du ménage pour les enfants vivant dans le ménage
SAVSELCP	Épargnes totales pour l'enfant sélectionné à la fin de 2007

<b>PARTICIPATION À L'ÉDUCATION</b>	
<b>Variable</b>	<b>Description</b>
SECSTAT	Situation par rapport au secondaire en juin 2008
PSSTAT	Situation par rapport au postsecondaire
HEDATT	Plus haut niveau de scolarité atteint
AGELFSEC	Âge a complété le secondaire ou quitté le primaire/secondaire
AGECOMPS	Âge au début du premier programme postsecondaire
DATLFSEC	Date a quitté le secondaire (ou primaire)
DATHLEV	Date à laquelle le répondant a atteint son plus haut niveau de scolarité
MOSECPS	Nombre de mois entre la fin du secondaire ou le départ du primaire/secondaire et le début des études postsecondaires
MOLFSEC	Nombre de mois entre la fin du secondaire ou le départ du primaire/secondaire et la date de l'enquête (30 juin 2008)
DATSTRPS	Date du début du premier programme postsecondaire
TYPEMRPG	Type du plus récent programme ou programme en cours
GOVSTAT	Situation de prêt étudiant du gouvernement
TOTEDEX	Dépenses scolaires totales pour le plus récent programme et durant la période de référence de juillet 2007 à juin 2008
TOTREPS	Sources de financement remboursables - total pour la période de référence de juillet 2007 à juin 2008
TOTNREPS	Sources de financement non remboursables - total pour la période de référence de juillet 2007 à juin 2008

<b>ÉDUCATION ET FORMATION</b>	
<b>Variable</b>	<b>Description</b>
PAIDHRS	Nombre d'heures rémunérées travaillées par semaine à l'emploi le plus récent (répondant principal)
PAIDHRSP	Nombre d'heures rémunérées travaillées par semaine à l'emploi le plus récent (PMR1 ou PMR2)
TKPM	Le répondant a suivi un programme pour des raisons d'emploi
TKNBPM	Nombre de programmes liés à l'emploi
TKTA	Le répondant a suivi de la formation liée à l'emploi
TKNBTA	Nombre de cours, ateliers ou séminaires liés à l'emploi
TAHRPM0	Durée du plus récent programme, en heures
TAHRCR0	Durée de la formation sélectionnée liée à l'emploi, en heures (sur le fichier du cours)
TAHRCR1	Durée de la première activité de formation liée à l'emploi, en heures
TAHRCR2	Durée de la deuxième activité de formation liée à l'emploi, en heures
TAHRCR3	Durée de la troisième activité de formation liée à l'emploi, en heures
TAHRCR4	Durée de la quatrième activité de formation liée à l'emploi, en heures

<b>ÉDUCATION ET FORMATION</b>	
<b>Variable</b>	<b>Description</b>
TAHR5	Durée de la cinquième activité de formation liée à l'emploi, en heures
TKHR	Durée de l'activité de formation liée à l'emploi - heures totales
TEPM	Soutien de l'employeur pour le plus récent programme

<b>CLASSIFICATION DES PROGRAMMES D'ENSEIGNEMENT (CPE) - PROGRAMMES</b>	
<b>Variable</b>	<b>Description</b>
MR1_C01	Code de la Classification des programmes d'enseignement - plus récent programme
MR2_C01	Code de la Classification des programmes d'enseignement - deuxième plus récent programme
MR3_C01	Code de la Classification des programmes d'enseignement - troisième plus récent programme
MR4_C01	Code de la Classification des programmes d'enseignement - quatrième plus récent programme
MR5_C01	Code de la Classification des programmes d'enseignement - cinquième plus récent programme
MR6_C01	Code de la Classification des programmes d'enseignement - sixième plus récent programme
MR7_C01	Code de la Classification des programmes d'enseignement - septième plus récent programme
MR8_C01	Code de la Classification des programmes d'enseignement - huitième plus récent programme
MR9_C01	Code de la Classification des programmes d'enseignement - neuvième plus récent programme
MR10_C01	Code de la Classification des programmes d'enseignement - dixième plus récent programme
MR1_G01	Code de la Classification des programmes d'enseignement (groupé) - plus récent programme
MR2_G01	Code de la Classification des programmes d'enseignement (groupé) - deuxième plus récent programme
MR3_G01	Code de la Classification des programmes d'enseignement (groupé) - troisième plus récent programme
MR4_G01	Code de la Classification des programmes d'enseignement (groupé) - quatrième plus récent programme
MR5_G01	Code de la Classification des programmes d'enseignement (groupé) - cinquième plus récent programme
MR6_G01	Code de la Classification des programmes d'enseignement (groupé) - sixième plus récent programme
MR7_G01	Code de la Classification des programmes d'enseignement (groupé) - septième plus récent programme
MR8_G01	Code de la Classification des programmes d'enseignement (groupé) - huitième plus récent programme
MR9_G01	Code de la Classification des programmes d'enseignement (groupé) - neuvième plus récent programme
MR10_G01	Code de la Classification des programmes d'enseignement (groupé) - dixième plus récent programme

<b>CLASSIFICATION DES PROGRAMMES D'ENSEIGNEMENT (CPE) - COURS</b>	
<b>Variable</b>	<b>Description</b>
CR1_C01	Code de la Classification des programmes d'enseignement – plus récent cours
CR1_C02	Code de la Classification des programmes d'enseignement - deuxième plus récent cours
CR1_C03	Code de la Classification des programmes d'enseignement - troisième plus récent cours
CR1_C04	Code de la Classification des programmes d'enseignement - quatrième plus récent cours
CR1_C05	Code de la Classification des programmes d'enseignement - cinquième plus récent cours
CR1_C06	Code de la Classification des programmes d'enseignement - sixième plus récent cours
CR1_C07	Code de la Classification des programmes d'enseignement - septième plus récent cours
CR1_C08	Code de la Classification des programmes d'enseignement - huitième plus récent cours
CR1_C09	Code de la Classification des programmes d'enseignement - neuvième plus récent cours
CR1_C10	Code de la Classification des programmes d'enseignement - dixième plus récent cours
CR1_G01	Code de la Classification des programmes d'enseignement (groupé) - plus récent cours
CR1_G02	Code de la Classification des programmes d'enseignement (groupé) - deuxième plus récent cours
CR1_G03	Code de la Classification des programmes d'enseignement (groupé) - troisième plus récent cours
CR1_G04	Code de la Classification des programmes d'enseignement (groupé) - quatrième plus récent cours
CR1_G05	Code de la Classification des programmes d'enseignement (groupé) - cinquième plus récent cours
CR1_G06	Code de la Classification des programmes d'enseignement (groupé) - sixième plus récent cours
CR1_G07	Code de la Classification des programmes d'enseignement (groupé) - septième plus récent cours
CR1_G08	Code de la Classification des programmes d'enseignement (groupé) - huitième plus récent cours
CR1_G09	Code de la Classification des programmes d'enseignement (groupé) - neuvième plus récent cours
CR1_G10	Code de la Classification des programmes d'enseignement (groupé) - dixième plus récent cours

## **7.6 Pondération**

Le principe qui sous-tend une estimation pour un échantillon probabiliste comme celui de l'EASEF 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 inscrites à des programmes à temps plein à une université au cours des 12 derniers 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 et taux de succès

Le tableau qui suit renferme un résumé des taux de réponse et les taux de succès de l'Enquête sur l'accès et le soutien à l'éducation et à la formation (EASEF).

Province	Numéros de téléphone envoyés à la collecte	Nombre estimé de numéros dans le champ de l'enquête	Taux de succès (%)	Répondants	Taux de réponse (%)	Répondants qui ont autorisé le partage des données	Taux de partage (%)
Terre-Neuve-et-Labrador	3 760	2 767	74	1 667	60	1 529	92
Île-du-Prince-Édouard	3 554	2 465	69	1 567	64	1 468	94
Nouvelle-Écosse	3 998	2 834	71	1 934	68	1 814	94
Nouveau-Brunswick	3 875	2 742	71	1 772	65	1 652	93
Québec	12 929	8 775	68	5 902	67	5 530	94
Ontario	20 801	14 313	69	9 061	63	8 636	95
Manitoba	4 120	2 757	67	1 867	68	1 768	95
Saskatchewan	3 987	2 895	73	1 806	62	1 727	96
Alberta	6 878	4 674	68	2 875	62	2 749	96
Colombie-Britannique	7 656	5 094	67	3 058	60	2 882	94
<b>Canada</b>	<b>71 558</b>	<b>49 305</b>	<b>69</b>	<b>31 509</b>	<b>64</b>	<b>29 755</b>	<b>94</b>

Les définitions des colonnes du tableau sont les suivantes :

#### Numéros de téléphone envoyés à la collecte

Ces nombres sont plus faibles que les chiffres de taille d'échantillon donnés à la section 5.4. Au départ, il était prévu d'envoyer 72 000 numéros de téléphone à la collecte, mais 442 ont été éliminés à cause d'un chevauchement avec d'autres enquêtes.

#### Nombre estimé de numéros dans le champ de l'enquête

Quand les numéros de téléphone sont revenus de la collecte, les cas ont été classés comme étant dans le champ de l'enquête, hors du champ de l'enquête ou non résolus. Les cas non résolus sont ceux pour lesquels les données ne sont pas suffisantes pour pouvoir les classer. Le nombre estimé de numéros de téléphone/ménages dans le champ de l'enquête est égal au nombre de cas qui ont été résolus comme étant dans le champ de l'enquête durant la collecte, plus une partie des cas non résolus.

#### Taux de succès

$$\text{Taux de succès} = \frac{\text{Nombre estimé de numéros dans le champ de l'enquête}}{\text{Numéros de téléphone envoyés à la collecte}}$$

#### Répondants

Nombre de cas considérés comme des répondants dans le fichier maître de Statistique Canada.

#### Taux de réponse

$$\text{Taux de réponse} = \frac{\text{Nombre de répondants}}{\text{Nombre estimé de numéros dans le champ de l'enquête}}$$

### **Répondants qui ont autorisé le partage des données**

Nombre de répondants qui ont donné la permission de partager leurs données avec Ressources humaines et Développement des compétences Canada (RHDCC). Il s'agit du nombre de cas dans le fichier de partage des données destiné à RHDCC.

### **Taux de partage**

$$\text{Taux de partage} = \frac{\text{Répondants qui ont autorisé le partage}}{\text{Nombre de répondants}}$$

## **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**

La couverture de la base de sondage de l'EASEF a été évaluée en divisant la somme des poids de sondage avant le calage par les chiffres de population. Les résultats selon le groupe d'âge sont les suivants :

Groupe d'âge	Couverture (%)
0 à 17	80
18 à 21	79
22 à 24	57
25 à 34	61
35 à 44	77
45 à 54	85
55 à 64	89
<b>Total</b>	<b>78</b>

La couverture est plus faible pour le groupe des 22 à 34 ans, ce que peuvent expliquer les taux de mobilité plus élevés et la plus forte proportion de personnes de ce groupe d'âge vivant dans des ménages dotés de téléphones mobiles uniquement.

Nous soupçonnions que la couverture des jeunes serait vraisemblablement plus faible parmi ceux vivant hors du domicile parental que parmi ceux vivant avec leurs parents. Pour le vérifier, nous avons calculé la proportion de jeunes vivant avec une personne au moins 18 ans plus âgée qu'eux en utilisant les répondants à l'EASEF et nous l'avons comparée à la même proportion calculée d'après les données du recensement. Les résultats sont les suivants :

Groupe d'âge	EASEF (%)	Recensement (%)
18 à 21	92	84
22 à 24	72	57

Le tableau montre que l'EASEF sur-représente les jeunes qui vivent avec un « parent » (une personne au moins 18 ans plus âgée qu'eux). Le tableau est fondé sur les poids avant calage.

Pour chaque groupe d'âge, les poids ont été calés sur les totaux démographiques. Le calage a également été utilisé pour corriger, dans une certaine mesure, la sur-représentation des jeunes vivant avec leurs parents.

## 8.2.2 Collecte des données

Les superviseurs de la collecte ont reçu une formation au bureau central à Ottawa et des employés du bureau central étaient disponibles pour répondre aux questions durant les séances de formation des intervieweurs. Un fichier contenant toutes les questions posées durant ces séances de formation ainsi que les réponses a été produit et distribué par la suite à tous les bureaux régionaux.

La formation des superviseurs et des intervieweurs consistait à lire le guide de l'intervieweur (ou du superviseur) afin de se familiariser avec les concepts et définitions de l'enquête. Une description du contexte et des objectifs a été fournie, ainsi qu'un glossaire et un ensemble de questions et réponses pour aider à répondre aux questions qui sont posées fréquemment par les répondants.

### 8.2.3 **Traitement des données**

La collecte de l'EASEF a été effectuée au moyen d'un instrument d'interviews téléphoniques assistées par ordinateur (ITAO). L'utilisation d'un instrument d'ITAO comporte deux avantages majeurs :

- elle assure que l'enchaînement correct des questions est suivi;
- elle permet de vérifier automatiquement les réponses incohérentes au moyen des contrôles compris dans l'application.

Au cours de la collecte, l'état matrimonial de la PMR2 qui vivait dans le ménage sélectionné a été recueilli dans le module démographique à la phase 1, et devait être transféré par l'application à la phase 2. Ce transfert n'a pas fonctionné et durant l'interview, le texte dynamique concernant le conjoint ou partenaire n'a pas bien fonctionné. Les questions touchées, parmi lesquelles la plus importante est le revenu combiné, peuvent être identifiées en utilisant la variable dérivée MRSTERRP – indicateur identifiant si le bon code d'état matrimonial a été utilisé pour les répondants PMR2 au cours de la collecte.

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

### 8.2.4 **Non-réponse**

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

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

Toutes les imputations ont été faites à l'aide de donneurs qui avaient été choisis au moyen d'une fonction de caractérisation. Pour tous les enregistrements contenant de la non-réponse ponctuelle ou de la non-réponse partielle (également appelés enregistrements receveurs), nous avons comparé certaines de leurs caractéristiques avec celles de tous les donneurs. Lorsque les caractéristiques étaient les mêmes entre un donneur et le receveur, une valeur était ajoutée au résultat de ce donneur. Le donneur ayant obtenu le résultat le plus élevé était réputé être le donneur « le plus semblable », et il était choisi pour fournir les éléments d'information manquants des non-répondants. S'il y avait plus d'un donneur avec le résultat le plus élevé, on procédait à une sélection aléatoire. Le bassin de donneurs était constitué de telle façon que la valeur imputée assignée au receveur, parallèlement à d'autres éléments d'information non imputés du receveur, serait encore acceptée à la vérification. Il convient de souligner que plus le taux d'imputation est faible, plus il est probable de trouver un donneur ayant des caractéristiques semblables à celles du receveur.

Le tableau qui suit donne la liste des variables qui ont été imputées, ainsi que le taux d'imputation pour chaque variable. La colonne intitulée « Enregistrements imputés (%) » donne la proportion (non pondérée) d'enregistrements qui contiennent une valeur imputée. Seuls les enregistrements contenant une valeur pour la variable ont été inclus dans le calcul (autrement dit, les champs vides valables ont été exclus). La colonne intitulée « Proportion de l'estimation imputée (%) » donne la proportion de l'estimation totale (pondérée) qui est basée sur des valeurs imputées.

Âge du répondant sélectionné	Nom de la variable	Description	Enregistrements imputés (%)	Proportion de l'estimation imputée (%)
Enfant âgé de 0 à 17 ans	SS_Q09P	Valeur du Régime enregistré d'épargne-études (REEE ) à la fin de l'année 2007	33	28
	SS_Q13P	Contributions totales au REEE en 2007	26	23
	SS_Q18P	Valeur des autres épargnes à la fin de l'année 2007 (exclure REEE)	33	29
	SS_Q20P	Contributions totales aux autres épargnes en 2007	32	24
	IN_Q02P	Revenu total provenant de toutes les sources avant impôts et autres retenues en 2007 de la personne la mieux renseignée (PMR)	26	27
	IN_Q04P	Revenu total de la PMR et de son (sa) conjoint(e) provenant de toutes sources, avant impôts et autres retenues en 2007	39	41
Jeune âgé de 18 à 24 ans	EE_Q14	Durée du plus récent programme	12	12
	CR2Q07_1 à CR2Q07_9 et CR2Q7_10	Durée des cours reliés au travail	7	10
	SS_Q09P	Valeur du REEE à la fin de l'année 2007	45	48
	SS_Q13P	Contributions totales au REEE en 2007	40	50
	SS_Q18P	Valeur des autres épargnes à la fin de l'année 2007 (exclure REEE)	51	58
	SS_Q20P	Contributions totales aux autres épargnes en 2007	46	53
	IN_Q02P	Revenu total provenant de toutes les sources avant impôts et autres retenues en 2007 de la PMR	44	51
Adulte âgé de 25 à 64 ans	EE_Q14	Durée du plus récent programme	8	7
	CR2Q07_1 à CR2Q07_9 et CR2Q7_10	Durée des cours reliés au travail	8	11

Des variables supplémentaires ont été imputées afin que les flux vers les variables énumérées dans le tableau qui précède soient corrects.

En ce qui concerne les cas où le répondant sélectionné avait 18 à 24 ans (jeune), des interviews de suivi ont été réalisées auprès du parent ou du gardien (PMR2). Dans 1 881 cas, une interview a été menée auprès du jeune, mais aucune donnée n'a été recueillie

auprès de la PMR2. La non-réponse aux interviews de suivi auprès de la PMR2 a été traitée par la méthode d'imputation par donneur: les cas concernant un jeune pour lequel il n'existait pas de données de la PMR2 ont été imputés en se servant de cas pour lesquels existaient des données déclarées par la PMR2. Les cas où toutes les variables de l'interview auprès de la PMR2 ont été imputées peuvent être identifiés à l'aide de l'indicateur d'imputation appelé IPMK2P. Le tableau qui suit donne les taux d'imputation pour les données de la PMR2 concernant le jeune :

Description	% de données de la PMR2 imputées
Jeune vivant dans le même ménage que la PMR2	20
Jeune vivant dans un autre ménage que la PMR2	55

Il convient de souligner que le taux d'imputation des données de la PMR2 est particulièrement élevé pour les jeunes vivant dans un autre ménage que leur PMR2. Le taux élevé signifie que nous étions moins susceptibles de trouver des donneurs ayant des caractéristiques semblables à celles des receveurs pour ces jeunes.

On a procédé à des imputations additionnelles pour 2 problèmes d'application d'ITAO. Le premier problème vient du fait qu'on n'a pas posé les questions IM\_Q15 et IM\_Q15P, utilisées pour dériver l'identificateur de visibilité minoritaire appelé VISMIN, aux personnes nées à l'extérieur du Canada, des États-Unis et du Groenland. L'identificateur de visibilité minoritaire a été imputé pour tous ces cas en utilisant le pays de naissance (IM\_Q05, IM\_C05P) et de l'information du Recensement de 2006. Le second problème, encouru durant la collecte, vient du fait qu'on n'a pas posé aux PMR des jeunes la question sur le revenu combiné, IN\_Q04P, dans tous les cas où le PMR avait un conjoint ou partenaire, et que le PMR et le jeune demeuraient dans le même ménage. La variable IN\_Q04P a été imputée par enregistrement donneur pour tous les cas où la PMR était un parent du jeune, et la PMR et son conjoint ou partenaire étaient des partenaires de sexe opposé. Le bassin de donneurs a été construit à partir des données du Recensement de 2006. L'imputation consistait à trouver un donneur du Recensement avec des caractéristiques semblables à ceux des receveurs de l'EASEF, et d'imputer une catégorie de revenu basée sur la valeur provenant du recensement et ajustée pour l'année de référence.

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

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

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

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

Par exemple, supposons que, d'après les résultats de l'enquête, l'on estime que 62,1 % d'enfants âgés de 0 à 17 ans ont une PMR1 qui espère que l'enfant poursuivra des études postsecondaires et qui a fait (éventuellement avec son(sa) conjoint(e)) des économies pour payer l'éducation postsecondaire de l'enfant. Supposons également que l'on constate que l'erreur-type de cette estimation est de 0,004. Le coefficient de variation de l'estimation est donc calculé comme suit :

$$\left( \frac{0,004}{0,621} \right) \times 100 \% = 0,64 \%$$

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 l'accès et le soutien à l'éducation et à la formation (EASEF) n'était pas autopondéré. Lorsqu'ils produisent des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent appliquer le poids d'enquête approprié.

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

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

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

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

### **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 d'enfants dont la PMR1 (personne la mieux renseignée) a épargné en prévision d'études postsecondaires ou la proportion d'adultes qui ont suivi une formation par enseignement à distance sont des exemples de ce genre d'estimation. Une estimation du nombre de personnes possédant une certaine caractéristique peut aussi être désignée comme une estimation d'un agrégat.

#### Exemples de questions catégoriques :

- Q : Est-ce que vous (et votre conjoint(e)) avez actuellement des épargnes accumulées en vue des études postsecondaires de/d'[nom de l'enfant]? Les épargnes comprennent les comptes bancaires, les CPG, les REEE, les REER, les fonds communs de placement, les fonds d'investissement, etc.
- R : Oui / Non
- Q : Quelle partie de cette formation avez-vous suivie par correspondance ou par une autre méthode de formation à distance?
- R : Aucune partie de la formation / Moins de la moitié / À peu près la moitié / Plus de la moitié / Toute la formation

### **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 la valeur moyenne des économies placées dans un Régime enregistré d'épargne-études (REEE) par enfant à la fin de 2007. Le numérateur est une estimation de la valeur totale des économies placées dans un REEE et son dénominateur est une estimation du nombre total d'enfants.

Exemples de questions quantitatives :

Q : Environ combien d'heures (rémunérées) par semaine (travailliez-vous/votre conjoint(e) travaillait-il/elle) habituellement à cet emploi (le plus récent)?

R : |\_|\_| heures

Q : Quelle était la valeur du REEE de/d[*nom de l'enfant*] à la fin de l'année 2007? Inclure les gains et l'intérêt, ainsi que la Subvention canadienne pour l'épargne-études (SCEE) et le Bon d'études canadien (BEC).

R : |\_|\_|\_|\_|\_|\_|\_| dollar

### 9.3.3 Totalisation d'estimations catégoriques

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

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

### 9.3.4 Totalisation d'estimations quantitatives

On peut obtenir des estimations de quantités à partir du fichier de microdonnées en multipliant la valeur de la variable qui nous intéresse par le poids final de chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements qui nous intéressent. Par exemple, pour obtenir une estimation de la valeur totale des économies placées dans un REEE à la fin de 2007 pour payer l'éducation postsecondaire des enfants dont la PMR1 espère qu'ils poursuivront des études postsecondaires, il faut multiplier la valeur déclarée à la question SS\_Q09P (valeur des REEE à la fin de 2007) par le poids final pour l'enregistrement, puis faire la somme de ces valeurs sur tous les enregistrements pour lesquels la valeur de PH\_Q01P est comprise entre 4 et 12 (postsecondaire).

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. Par exemple, pour estimer la valeur moyenne des économies placées dans un REEE à la fin de 2007 pour payer l'éducation postsecondaire des enfants dont la PMR1 espère qu'ils poursuivront des études postsecondaires,

- a) estimez la valeur totale des économies placées dans un REEE ( $\hat{X}$ ) tel qu'il est décrit ci-dessus,
- b) estimez le nombre d'enfants ( $\hat{Y}$ ) incluses dans cette catégorie en additionnant les poids finals de tous les enregistrements où la valeur de PH\_Q01P est comprise entre 4 et 12, puis
- c) divisez l'estimation a) par l'estimation b) ( $\hat{X} / \hat{Y}$ ).

## 9.4 Lignes directrices pour l'analyse statistique

L'EASEF 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 18 à 24 ans. Les étapes à suivre pour remettre à l'échelle les poids sont les suivantes :

- 1) sélectionner tous les répondants du fichier qui ont déclaré DVAGE = 18 à 24;
- 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 DVAGE = 18 à 24;
- 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 est recommandé d'utiliser la méthode bootstrap pour estimer les variances pour l'EASEF. Il s'agit d'une méthode de rééchantillonnage permettant de produire des estimations de variance valides pour les plans d'échantillonnage complexes et les estimateurs complexes. Dans le cas de

l'EASEF, nous avons produit deux ensembles de 1,000 poids bootstrap qui sont fournis avec les fichiers de données. L'un des ensembles de poids bootstrap doit être utilisé avec le fichier de données principal (niveau de la personne) et le second doit être utilisé avec le fichier de données au niveau du cours.

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

SAS: [http://www.statcan.gc.ca/rdc-cdr/bootvar\\_sas-fra.htm](http://www.statcan.gc.ca/rdc-cdr/bootvar_sas-fra.htm)

SPSS: [http://www.statcan.gc.ca/rdc-cdr/bootvar\\_spss-fra.htm](http://www.statcan.gc.ca/rdc-cdr/bootvar_spss-fra.htm)

Pour les estimations fondées sur des données imputées, il est recommandé de corriger le coefficient de variation à l'aide d'un facteur d'inflation, comme il suit :

$$CV\_corrige = \left( \frac{1}{\sqrt{1 - \text{Taux d'imputation}}} \right) \times CV ,$$

où CV est le coefficient de variation calculé en utilisant la méthode bootstrap.

## **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'EASEF, 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

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

Par exemple, d'après le tableau, la qualité d'une estimation pondérée de 5 000 personnes possédant une caractéristique donnée à Terre-Neuve-et-Labrador est médiocre.

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

**Tableau de seuils de diffusion – Fichier principal**

Provinces	CV acceptable 0,0 à 16,5 %	CV médiocre 16,6 à 33,3 %	CV inacceptable > 33,3 %
Terre-Neuve-et-Labrador	17 500 et plus	4 400 à < 17 500	moins de 4 400
Île-du-Prince-Édouard	4 900 et plus	1 200 à < 4 900	moins de 1 200
Nouvelle-Écosse	30 100 et plus	7 600 à < 30 100	moins de 7 600
Nouveau-Brunswick	24 300 et plus	6 100 à < 24 300	moins de 6 100
Québec	82 800 et plus	20 600 à < 82 800	moins de 20 600
Ontario	95 700 et plus	23 700 à < 95 700	moins de 23 700
Manitoba	37 200 et plus	9 400 à < 37 200	moins de 9 400
Saskatchewan	30 300 et plus	7 700 à < 30 300	moins de 7 700
Alberta	76 800 et plus	19 200 à < 76 800	moins de 19 200
Colombie-Britannique	94 600 et plus	23 700 à < 94 600	moins de 23 700
<b>Canada</b>	<b>82 000 et plus</b>	<b>20 200 à &lt; 82 000</b>	<b>moins de 20 200</b>

**Tableau de seuils de diffusion – Fichier du cours**

Provinces	CV acceptable 0,0 à 16,5 %	CV médiocre 16,6 à 33,3 %	CV inacceptable > 33,3 %
Terre-Neuve-et-Labrador	90 300 et plus	31 700 à < 90 300	moins de 31 700
Île-du-Prince-Édouard	18 600 et plus	5 600 à < 18 600	moins de 5 600
Nouvelle-Écosse	113 400 et plus	33 700 à < 113 400	moins de 33 700
Nouveau-Brunswick	67 400 et plus	19 400 à < 67 400	moins de 19 400
Québec	394 400 et plus	109 500 à < 394 400	moins de 109 500
Ontario	885 500 et plus	238 900 à < 885 500	moins de 238 900
Manitoba	250 900 et plus	87 700 à < 250 900	moins de 87 700
Saskatchewan	110 300 et plus	31 600 à < 110 300	moins de 31 600
Alberta	360 600 et plus	100 800 à < 360 600	moins de 100 800
Colombie-Britannique	401 300 et plus	115 400 à < 401 300	moins de 115 400
<b>Canada</b>	<b>666 100 et plus</b>	<b>168 600 à &lt; 666 100</b>	<b>moins de 168 600</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 l'accès et le soutien à l'éducation et à la formation (EASEF).

### Fichier principal

Provinces	Effet du plan	Taille de l'échantillon	Population
Terre-Neuve-et-Labrador	1,76	1 529	433 148
Île-du-Prince-Édouard	1,73	1 468	118 756
Nouvelle-Écosse	2,00	1 814	774 025
Nouveau-Brunswick	1,82	1 652	625 391
Québec	1,94	5 530	6 524 701
Ontario	2,05	8 636	11 081 441
Manitoba	1,93	1 768	965 974
Saskatchewan	1,85	1 727	803 193
Alberta	1,91	2 749	3 090 573
Colombie-Britannique	2,05	2 882	3 719 183
<b>Canada</b>	<b>2,37</b>	<b>29 755</b>	<b>28 136 385</b>

### Fichier du cours

Provinces	Effet du plan	Taille de l'échantillon	Population
Terre-Neuve-et-Labrador	4,71	262	227 105
Île-du-Prince-Édouard	2,79	314	76 005
Nouvelle-Écosse	3,12	385	494 733
Nouveau-Brunswick	2,18	332	347 374
Québec	3,62	736	2 578 207
Ontario	6,51	1 766	7 426 596
Manitoba	7,02	395	635 376
Saskatchewan	2,61	409	581 407
Alberta	3,20	612	2 238 423
Colombie-Britannique	3,67	560	2 068 718
<b>Canada</b>	<b>6,54</b>	<b>5 771</b>	<b>16 673 943</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 ou être calculées en se servant des poids bootstrap fournis avec le fichier de données. 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. Par exemple, la proportion d'enfants dont la PMR1 (personne la mieux renseignée) a fait des économies pour payer leurs études postsecondaires est plus fiable que le nombre estimé d'enfants dont la PMR1 a fait des économies pour payer leurs études postsecondaires. (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 le pourcentage repose sur un sous-ensemble de la population totale (p. ex., comme ses membres d'un sexe ou d'un groupe d'âge particulier), on devrait faire référence à la proportion ou au pourcentage (dans le haut de la table) et au numérateur de la proportion ou du pourcentage (dans la colonne de gauche de la table). L'intersection de la rangée et de la colonne appropriées donne le coefficient de variation.

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

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

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

où  $\hat{X}_1$  est l'estimation 1,  $\hat{X}_2$  est l'estimation 2 et  $\alpha_1$  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 jeunes et le numérateur, le nombre de jeunes qui poursuivent des études postsecondaires.

Dans le cas où le numérateur n'est pas un sous-ensemble du dénominateur, comme dans l'exemple du rapport du nombre d'étudiants du niveau postsecondaire de sexe féminin comparativement au nombre d'étudiants du niveau postsecondaire de sexe masculin, 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' EASEF et sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter.

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

Supposons qu'un utilisateur estime que 4 174 026 enfants de 0 à 17 ans ont une PMR1 qui espère que l'enfant poursuivra des études postsecondaires et qui a fait (éventuellement avec son (sa) conjoint(e)) des économies pour payer l'éducation postsecondaire de l'enfant (d'après les question FP\_Q01P ou FO\_Q01P). 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 calculés pour le fichier principal.
- 2) L'agrégat estimé 4 174 026 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 2,2 %.
- 4) Le taux d'imputation pour FP\_Q01P est de 1,9 %. Le facteur d'inflation pour l'imputation est d'environ 1,01, d'où le CV corrigé =  $1,01 \times 2,2 \% = 2,2 \%$  (voir la section 9.4). (Le taux d'imputation est suffisamment faible pour que le CV ne change pas.)
- 5) Le coefficient de variation approximatif de l'estimation est donc 2,2 %. Le résultat selon lequel il y avait 4 174 026 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1) enfants dont la PMR1 espère que l'enfant poursuivra des études postsecondaires et a fait des économies pour payer l'éducation postsecondaire de l'enfant est publiable sans aucune restriction.

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

Supposons qu'un utilisateur estime que  $4\,174\,026 / 6\,721\,149 = 62,1 \%$  d'enfants de 0 à 17 ans ont une PMR1 qui espère que l'enfant poursuivra des études postsecondaires et qui a fait (éventuellement avec son (sa) conjoint(e)) des économies pour payer l'éducation postsecondaire de l'enfant. 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 calculés pour le fichier principal.
- 2) Parce que l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c'est-à-dire d'enfants de 0 à 17 ans), il faut utiliser à la fois le pourcentage (62,1 %) et la portion numérateur du pourcentage (4 174 026) pour déterminer le coefficient de variation.
- 3) Le numérateur, 4 174 026, 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 4 000 000. 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 70,0 %.

- 4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 1,3 %, est le coefficient de variation à employer.
- 5) Le taux d'imputation pour FP\_Q01P est de 1,9 %. Le facteur d'inflation pour l'imputation est d'environ 1,01, d'où le CV corrigé =  $1,01 \times 1,3 \% = 1,3 \%$ . (Le taux d'imputation est suffisamment faible pour que le CV ne change pas.)
- 6) Le coefficient de variation approximatif de l'estimation est donc 1,3 %. La constatation que 62,1 % d'enfants ont une PMR1 qui espère que l'enfant poursuivra des études postsecondaires et qui a fait des économies pour payer l'éducation postsecondaire de l'enfant peut être publiée sans restriction.

**Exemple 2B : Estimations de proportions ou de pourcentages des activités de formation liée à l'emploi possédant une caractéristique donnée**

Supposons que l'utilisateur estime que  $1\,162\,711 / 15\,462\,178 = 7,5 \%$  des cours, ateliers ou séminaires en rapport avec l'emploi pris par les personnes de 18 à 64 ans entre juillet 2007 et juin 2008 ont été pris entièrement par enseignement à distance. 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 calculés pour le fichier du cours.
- 2) Parce que l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c'est-à-dire des cours, ateliers ou séminaires en rapport avec l'emploi pris par les personnes de 18 à 64 ans entre juillet 2007 et juin 2008), il faut utiliser à la fois le pourcentage (7,5 %) et la portion numérateur du pourcentage (1 162 711) pour déterminer le coefficient de variation.
- 3) Le numérateur, 1 162 711, 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 1 000 000. 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 10,0 %.
- 4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 13,0 %, est le coefficient de variation à employer.
- 5) La variable CN\_Q04 n'ayant pas été imputée, le CV ne doit pas être corrigé.
- 6) Le coefficient de variation approximatif de l'estimation est donc 13,0 %. La constatation que 7,5 % des cours, ateliers ou séminaires en rapport avec l'emploi ont été pris entièrement par enseignement à distance peut être publiée sans aucune restriction.

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

Supposons que l'utilisateur estime que  $2\,890\,532 / 4\,593\,340 = 62,9 \%$  de jeunes enfants de 0 à 12 ans ont une PMR1 qui espère qu'ils poursuivront des études postsecondaires et qui a fait (éventuellement avec son (sa) conjoint(e)) des économies pour payer l'éducation postsecondaire de l'enfant. Supposons que l'utilisateur estime aussi que la proportion équivalente d'enfants de 13 à 17 ans est égale à  $1\,283\,494 / 2\,127\,809 = 60,3 \%$ . Comment l'utilisateur détermine-t-il le coefficient de variation de la différence

entre ces deux estimations?

- 1) En utilisant la table des coefficients de variation du CANADA calculés pour le fichier principal de la même façon que celle décrite dans l'exemple 2A, on obtient un CV de l'estimation pour les jeunes enfants de 1,5 % et un CV de l'estimation pour les enfants plus âgés de 2,1 %.
- 2) Les taux d'imputation pour FP\_Q01P sont de 2,45 % pour les jeunes enfants et de 0,74 % pour les enfants plus âgés. Les taux d'imputation sont suffisamment faibles pour que les CV corrigés ne changent pas.
- 3) 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 (les jeunes enfants),  $\hat{X}_2$  est l'estimation 2 (les enfants plus âgés) 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  $\hat{d} = 0,629 - 0,603 = 0,026$  est :

$$\begin{aligned} \sigma_{\hat{d}} &= \sqrt{[(0,629)(0,015)]^2 + [(0,603)(0,021)]^2} \\ &= \sqrt{(0,000089) + (0,000160)} \\ &= 0,016 \end{aligned}$$

- 4) Le coefficient de variation de  $\hat{d}$  est donné par  $\sigma_{\hat{d}} / \hat{d} = 0,016 / 0,026 = 0.615$ .
- 5) Le coefficient de variation approximatif de la différence entre les estimations est donc 61,5 %. La différence entre les estimations est considérée inacceptable et Statistique Canada recommande de ne pas publier cette estimation. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre 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 que l'utilisateur estime que 1 060 647 jeunes femmes de 18 à 24 ans ont fait au moins des études postsecondaires partielles, tandis que 1 004 179 jeunes hommes ont fait au moins des études postsecondaires partielles. L'utilisateur souhaite comparer l'estimation pour les femmes et pour les hommes sous 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 jeunes femmes qui ont fait au moins des études postsecondaires partielles. Le dénominateur de l'estimation ( $\hat{X}_2$ ) est le nombre de jeunes hommes qui ont fait au moins des études postsecondaires partielles.

- 2) Reportez-vous à la table des coefficients de variation pour le CANADA calculés pour le fichier principal.
- 3) Le numérateur de cette estimation de rapport est 1 060 647. 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 4,6 %.
- 4) Le dénominateur de cette estimation de rapport est 1 004 179. 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 4,6 %.
- 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 :

$$\begin{aligned} \alpha_{\hat{R}} &= \sqrt{(0,046)^2 + (0,046)^2} \\ &= \sqrt{0,00212 + 0,00212} \\ &= 0,0651 \end{aligned}$$

- 6) Le taux d'imputation pour EA\_Q01 pour les jeunes est de 11,1 %. Le facteur d'inflation pour tenir compte de l'imputation est égal à environ 1,06, d'où le CV corrigé = 1,06 x 6,5 % = 6,9 %.
- 7) Le rapport obtenu entre les jeunes femmes et les jeunes hommes qui ont fait au moins des études secondaires partielles est 1 060 647 / 1 004 179, c'est-à-dire 1,06 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1). Le coefficient de variation de cette estimation est 6,9 %, ce qui fait qu'on peut la diffuser sans réserve.

#### Exemple 5 : Estimations de différences de rapports

Supposons que l'utilisateur estime que le rapport du nombre de jeunes femmes au nombre de jeunes hommes qui ont fait au moins des études secondaires partielles est de 1,11 pour l'Ontario, mais qu'il est de 0,97 pour la Colombie-Britannique. L'utilisateur est intéressé à comparer les deux rapports pour voir s'il y a une différence statistique entre ceux-ci. Comment l'utilisateur détermine-t-il le coefficient de variation de cette différence?

- 1) Tout d'abord calculez le coefficient de variation approximatif pour le rapport de l'Ontario ( $\hat{R}_1$ ) et le rapport de la Colombie-Britannique ( $\hat{R}_2$ ) tel qu'il est décrit dans l'exemple 4. Le CV approximatif pour le rapport de l'Ontario est 10,8 % et 20,1 % pour celui de la Colombie-Britannique.
- 2) Les taux d'imputation pour EA\_Q01 sont de 13,1 % pour l'Ontario et de 13,7 % pour

la Colombie-Britannique. Les CV corrigés sont de 11,6 % pour l'Ontario et de 21,7 % pour la Colombie-Britannique.

- 3) 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  $\hat{d} = 1,11 - 0,97 = 0,14$  est :

$$\begin{aligned} \sigma_{\hat{d}} &= \sqrt{[(1,11)(0,116)]^2 + [(0,97)(0,217)]^2} \\ &= \sqrt{(0,0166) + (0,0443)} \\ &= 0,247 \end{aligned}$$

- 4) Le coefficient de variation de  $\hat{d}$  est donné par  $\sigma_{\hat{d}} / \hat{d} = 0,247 / 0,14 = 1,76$ .
- 5) Le coefficient de variation approximatif de la différence entre les estimations est donc 176 %. 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 d'enfants de 0 à 17 ans qui ont une PMR1 qui espère que l'enfant poursuivra des études postsecondaires et qui a fait (éventuellement avec son (sa) conjoint(e)) des économies pour payer l'éducation postsecondaire de l'enfant (d'après l'exemple 2A à la section 10.1.1) serait calculé comme suit :

$$\hat{X} = 62,1 \% \text{ (ou exprimé sous forme de proportion } 0,621)$$

$$t = 2$$

$$\alpha_{\hat{x}} = 1,3 \% \text{ (0,013 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.}$$

$$IC_{\hat{x}} = \{0,621 - (2)(0,621)(0,013), 0,621 + (2)(0,621)(0,013)\}$$

$$IC_{\hat{x}} = \{0,621 - 0,016, 0,621 + 0,016\}$$

$$IC_{\hat{x}} = \{0,605, 0,637\}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 60,5 % et 63,7 % d'enfants de 0 à 17 ans ont une PMR1 qui espère que l'enfant poursuivra des études postsecondaires et qui a fait (éventuellement avec son (sa) conjoint(e)) des économies

pour payer l'éducation postsecondaire de l'enfant.

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

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

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

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

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

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

Supposons que l'utilisateur désire tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y a pas de différence entre la proportion de jeunes enfants de 0 à 12 ans et d'enfants plus âgés de 13 à 17 ans qui ont une PMR1 qui espère qu'ils poursuivront des études postsecondaires et qui a fait (éventuellement avec son (sa) conjoint(e)) des économies pour payer l'éducation postsecondaire de l'enfant. 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 0,016. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0,629 - 0,603}{0,016} = \frac{0,026}{0,016} = 1,63$$

Puisque  $t = 1,63$  est inférieur à 2, il faut en conclure qu'il n'existe pas de différence significative entre les deux estimations au niveau de signification de 0,05.

### 10.4 Coefficients de variation pour des estimations quantitatives

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 de l'EASEF 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 de la valeur totale des REEE épargnés pour les enfants

serait supérieur au coefficient de variation de la proportion correspondante d'enfants ayant des REEE. 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 l'annexe G pour les tables de coefficients de variation reliées au fichier principal.

Consulter l'annexe H pour les tables de coefficients de variation reliées au fichier du cours.



## 11.0 Pondération

Le présent chapitre décrit les étapes de pondération suivies pour calculer les poids finaux pour l'Enquête sur l'accès et le soutien à l'éducation et à la formation (EASEF). Les enregistrements ont été classés de la façon suivante :

Catégorie		Description	Nombre
Chevauchement	OVL P	Chevauchement avec d'autres enquêtes – pas envoyés à la collecte des données	3 503
Non résolu	UNR1	Pas de données de présélection et pas de liste des membres du ménage; code de résultat non résolu	576
	UNR2	Pas de données de présélection; code de résultat dans le champ de l'enquête	12 127
	UNR3	Données de présélection pour déterminer si dans le champ de l'enquête, mais pas de liste des membres du ménage	1 528
Hors du champ de l'enquête	OOS1	Code de résultat hors du champ de l'enquête, sauf les cas dans OOS2	15 225
	OOS2	Téléphones mobiles ou logements secondaires avec données de présélection	3 069
	OOS3	Liste des membres du ménage établie, ménage hors du champ de l'enquête (tous les membres de 65 ans et plus)	2 204
Dans le champ de l'enquête	SHR	Partageurs : répondants qui ont autorisé le partage de leurs données	29 755
	NSHR	Non-partageurs : répondants qui n'ont pas autorisé le partage de leurs données	1 754
	NR	Non-répondants : liste des membres du ménage établie, mais aucune donnée pour le répondant sélectionné	5 320
<b>Total taille initiale de l'échantillon</b>			<b>75 061</b>

Les étapes de pondération suivantes ont été exécutées :

### 1. Calcul des poids de sondage

Un poids de sondage,  $W_1$ , a été attribué à chacun des 75 061 enregistrements de l'échantillon initial.

Pour la partie Recensement de l'échantillon, les poids de sondage étaient égaux au poids 2B du recensement multiplié par l'inverse de la probabilité de sélection. Pour la strate Administrative, les poids de sondage étaient égaux à l'inverse de la probabilité de sélection.

### 2. Correction pour le chevauchement avec d'autres enquêtes (OVL P)

Une correction de la pondération a été effectuée pour tenir compte des 3 503 numéros de téléphone qui n'ont pas été envoyés sur le terrain pour la collecte des données parce qu'ils concordaient avec des numéros de téléphone déjà utilisés pour l'interview dans d'autres enquêtes. Des classes de pondération ont été créées en se fondant sur la strate et sur les données de recensement pour certains enregistrements. La correction suivante a été calculée dans chaque classe de pondération :

$$W_2 = \left( \frac{\sum W_1 \text{ pour les enregistrements envoyés sur le terrain} + \sum W_1 \text{ pour les enregistrements dans OVL P}}{\sum W_1 \text{ pour les enregistrements envoyés sur le terrain}} \right) \times W_1$$

### 3. Correction pour les unités non résolues pour lesquelles il n'existe pas de données (UNR1)

Une correction de la pondération a été effectuée pour tenir compte des 576 enregistrements qui ont été envoyés sur le terrain pour la collecte des données, mais pour lesquels aucune donnée n'a été reçue afin de déterminer s'ils étaient dans le champ de l'enquête. La correction qui suit a été calculée pour chaque strate :

$$W_3 = \left( \frac{\sum W_2 \text{ pour les enregistrements envoyés sur le terrain}}{\sum W_2 \text{ pour les enregistrements classés SHR, NSHR, NR, OOS1, OOS2, OOS3, UNR2, UNR3}} \right) \times W_2$$

#### 4. Correction pour les unités non résolues sans données de présélection (UNR2)

Une correction de la pondération a été effectuée pour tenir compte des 12 127 enregistrements ayant un code de résultat dans le champ de l'enquête, mais pour lesquels il n'existait aucune donnée de présélection ni aucune liste des membres du ménage. La propension à répondre a été modélisée en utilisant les données du recensement (pour la partie Recensement de l'échantillon) et des parodonnées. Des classes de pondération ont été créées en se fondant sur la propension prédite à répondre. La correction qui suit a été calculée dans chaque classe de pondération :

$$W_4 = \left( \frac{\sum W_3 \text{ pour les enregistrements classés SHR, NSHR, NR, OOS2, OOS3, UNR3} + \sum W_3 \text{ pour les enregistrements classés UNR2}}{\sum W_3 \text{ pour les enregistrements classés SHR, NSHR, NR, OOS2, OOS3, UNR3}} \right) \times W_3$$

#### 5. Correction pour les unités non résolues sans liste des membres du ménage (UNR3)

Une correction de la pondération a été effectuée pour tenir compte des 1 528 enregistrements possédant un code de résultat dans le champ de l'enquête pour lesquels il existait des données de présélection, mais pas de liste des membres du ménage. La propension à répondre a été modélisée en utilisant les données du recensement (pour la partie Recensement de l'échantillon) et des parodonnées. Des classes de pondération ont été créées en se basant sur la propension prédite à répondre. La correction qui suit a été calculée dans chaque classe de pondération :

$$W_5 = \left( \frac{\sum W_4 \text{ pour les enregistrements classés SHR, NSHR, NR, OOS3} + \sum W_4 \text{ pour les enregistrements classés UNR3}}{\sum W_4 \text{ pour les enregistrements classés SHR, NSHR, NR, OOS3}} \right) \times W_4$$

#### 6. Calcul des poids au niveau de la personne

Les poids au niveau de la personne ont été calculés pour les cas dans le champ de l'enquête pour lesquels avaient été recueillies les données de présélection et la liste des membres du ménage et pour lesquels un membre du ménage avait été sélectionné. Les poids obtenus à l'étape précédente ont été multipliés par l'inverse de la probabilité de sélection pour la personne sélectionnée. Une limite a été imposée à la correction afin d'éviter les poids extrêmes. Les poids au niveau de la personne ont été désignés  $W_6$ .

#### 7. Correction pour la non-réponse d'une personne (NR)

Une correction de la pondération a été effectuée pour tenir compte des 5 320 cas dans le champ de l'enquête pour lesquels le membre du ménage sélectionné d'après la liste des membres n'a pas fourni suffisamment de données pour être considéré comme un répondant. La propension à répondre a été modélisée en utilisant les données du recensement (pour la partie Recensement de l'échantillon) et les données provenant des listes de membres du ménage de l'EASEF. Des classes de pondération ont été créées en se basant sur la propension prédite à répondre. La correction suivante a été calculée dans chaque classe de pondération :

$$W_7 = \left( \frac{\sum W_6 \text{ pour les enregistrements classés SHR, NSHR} + \sum W_6 \text{ pour les enregistrements classés NR}}{\sum W_6 \text{ pour les enregistrements classés SHR, NSHR}} \right) \times W_6$$

#### 8. Calage pour le fichier maître (SHR et NSHR)

Les poids pour les 31 509 répondants figurant dans le fichier maître (partageurs et non-partageurs) ont été calés de manière que la somme des poids soit égale aux chiffres de population de juillet 2008, au niveau de la province selon le groupe d'âge et le sexe. Le calage a également été utilisé pour réduire la

sous-représentation des jeunes de 18 à 24 ans ne vivant pas avec leurs parents. Les poids finaux (au niveau de la personne) pour le fichier maître sont appelés WTPM.

### **9. Calage pour le fichier de partage des données (SHR)**

Les poids calculés pour les 29 755 répondants figurant dans le fichier partage des données (répondants qui ont autorisé le partage de leurs données) ont été calés sur les mêmes chiffres de population au niveau de la province selon le groupe d'âge et le sexe que pour le fichier maître. En outre, les poids ont été calés de manière que le fichier de partage des données produise certaines estimations identiques à celles calculées d'après le fichier maître. Les poids finaux (au niveau de la personne) pour le fichier de partage des données sont appelés WTPS.

### **10. Calcul des poids au niveau du cours pour le fichier maître de données au niveau du cours**

Les poids au niveau du cours ont été calculés pour tous les répondants qui ont déclaré avoir suivi une formation pour des raisons d'emploi ou de carrière à la question CR1\_Q02. Pour ceux ayant déclaré dix cours ou moins à la question EC\_Q03, les poids pour le cours ont été calculés de la façon suivante :

$$\text{WTCM} = \text{WTPM} \times \text{Nombre de cours en rapport avec l'emploi}$$

Pour les répondants ayant déclaré plus de dix cours à la question EC\_Q03, les poids du cours ont été calculés de la façon suivante :

$$\text{WTCM} = \text{WTPM} \times (\text{EC\_Q03} / 10) \times \text{Nombre de cours en rapport avec l'emploi}$$

Les poids au niveau du cours doivent être utilisés avec le fichier de données au niveau du cours qui contient les données pour le cours sélectionné aléatoirement. Les poids pour le fichier maître de données sur le cours sont appelés WTCM.

### **11. Calcul des poids au niveau du cours pour le fichier de partage des données sur le cours**

Les poids au niveau du cours ont été calculés pour tous les répondants ayant déclaré avoir suivi une formation pour des raisons d'emploi ou de carrière à la question CR1\_Q02. Pour ceux ayant déclaré dix cours ou moins à la question EC\_Q03, les poids du cours ont été calculés de la façon suivante :

$$\text{WTCS} = \text{WTPS} \times \text{Nombre de cours en rapport avec l'emploi}$$

Pour les répondants ayant déclaré plus de dix cours à la question EC\_Q03, les poids du cours ont été calculés de la façon suivante :

$$\text{WTCS} = \text{WTPS} \times (\text{EC\_Q03} / 10) \times \text{Nombre de cours en rapport avec l'emploi}$$

Les poids au niveau du cours doivent être utilisés avec le fichier de données au niveau du cours, qui contient les données pour le cours sélectionné aléatoirement. Les poids pour le fichier de données de partage sur le cours sont appelés WTCS.



## **12.0 Questionnaire**

Le questionnaire de l'Enquête sur l'accès et le soutien à l'éducation et à la formation (EASEF) a servi à recueillir des données en 2008. Le fichier EASEF2008\_QuestF.pdf renferme le questionnaire français.



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

Consulter le fichier EASEF2008\_Principal\_LvCds.pdf pour le cliché d'enregistrement à valeurs univariées pour le fichier principal.

Consulter le fichier EASEF2008\_Cours\_LvCds.pdf pour le cliché d'enregistrement à valeurs univariées pour le fichier du cours.