

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

## **Enquête sur les ménages et l'environnement**

2009



Statistique  
Canada

Statistics  
Canada

**Canada**



### **Pour obtenir plus d'information**

Toutes les questions concernant l'ensemble de données ou son utilisation devraient être adressées à :  
Division des comptes et de la statistique de l'environnement  
Téléphone : 613-951-0297  
Télécopieur : 613-951-0634  
Courriel : [environ@statcan.gc.ca](mailto:environ@statcan.gc.ca)

### **Description, accès et utilisation**

L'Enquête sur les ménages et l'environnement (EME) de 2009 produit trois types de fichiers de microdonnées : les fichiers maîtres, les fichiers de partage et les fichiers de microdonnées à grande diffusion (FMGD).

#### **Fichiers maîtres**

Les fichiers maîtres contiennent toutes les variables et tous les enregistrements de l'enquête collectés au cours d'une période de collecte. Ces fichiers sont accessibles à Statistique Canada pour usage interne, dans les Centres de données de recherche (CDR) de Statistique Canada et peuvent aussi faire l'objet de demandes de totalisations personnalisées.

#### **Centre de données de recherche**

Le Programme des CDR permet aux chercheurs d'utiliser les données d'enquête contenues dans les fichiers maîtres dans un environnement sécuritaire situé dans plusieurs universités à travers le Canada. Les chercheurs doivent soumettre des propositions de recherche qui une fois acceptées leur donneront accès aux CDR. Pour plus de renseignements, consultez la page web suivante :  
<http://www.statcan.gc.ca/rdc-cdr/index-fra.htm>

#### **Totalisations personnalisées**

Une autre méthode d'accès aux fichiers maîtres consiste à offrir à tous les utilisateurs de faire appel au personnel de la Division des comptes et de la statistique de l'environnement pour produire des totalisations personnalisées. Ce service est offert moyennant le recouvrement des coûts. Il permet aux utilisateurs qui ne savent pas se servir de logiciels de totalisation d'obtenir des résultats personnalisés. Les résultats sont filtrés pour s'assurer qu'ils sont conformes aux normes de confidentialité et de fiabilité avant d'être diffusés. Pour plus de renseignements, communiquez avec la Division des comptes et de la statistique de l'environnement.

#### **Fichiers de partage**

Les fichiers de partage contiennent toutes les variables et tous les enregistrements des répondants de l'EME qui ont accepté de partager leurs données avec les partenaires de partage de Statistique Canada. Le fichier de partage n'est fourni qu'à ces organisations. Les identificateurs personnels sont retirés des fichiers de partage pour préserver la confidentialité des répondants. Les utilisateurs de ces fichiers doivent au préalable avoir porté serment qu'ils ne divulgueront en aucun temps toute information susceptible d'identifier un répondant à l'enquête.

#### **Fichiers de microdonnées à grande diffusion**

Les fichiers de microdonnées à grande diffusion sont élaborés à partir des fichiers maîtres suivant une technique qui vise à concilier l'impératif d'assurer la confidentialité des répondants et la nécessité de produire des données d'utilité maximale. Les FMGD doivent répondre à des normes sévères de sécurité et de confidentialité, conformément à la Loi sur la statistique avant qu'ils ne soient diffusés pour l'accès public. Pour s'assurer du respect de ces normes, chaque FMGD est soumis à un processus officiel d'examen et d'approbation par un comité formé de haut gestionnaire de Statistique Canada. Les variables les plus susceptibles de permettre l'identification d'une personne sont supprimées du fichier ou agrégées en catégories moins détaillées.

Un contrat de licence de microdonnées doit être signé avant la diffusion du (des) fichier(s). La personne responsable du fichier doit fournir son nom et coordonnées. Pour commander le FMGD communiquez avec la Division des comptes et de la statistique de l'environnement.

**L'Initiative de démocratisation des données**

Le Programme des Initiative de démocratisation des données (IDD) permet aux étudiants et chercheurs d'utiliser les fichiers de microdonnées à grande diffusion dans plusieurs universités à travers le Canada. Pour plus de renseignements, consultez la page web suivante :

<http://www.statcan.gc.ca/dli-ild/dli-idd-fra.htm>

## Table des matières

<b>1.0</b>	<b>Introduction</b> .....	<b>7</b>
<b>2.0</b>	<b>Contexte</b> .....	<b>7</b>
<b>3.0</b>	<b>Objectifs</b> .....	<b>7</b>
<b>4.0</b>	<b>Concepts et définitions</b> .....	<b>8</b>
4.1	Concepts et définitions de l'Enquête sur la santé dans les collectivités canadiennes .....	8
4.2	Concepts et définitions de l'Enquête sur les ménages et l'environnement .....	8
<b>5.0</b>	<b>Méthodologie de l'enquête</b> .....	<b>12</b>
5.1	Population visée par l'Enquête sur la santé dans les collectivités canadiennes .....	12
5.2	Plan de sondage de l'Enquête sur la santé dans les collectivités canadiennes.....	12
5.3	Taille de l'échantillon selon la province pour l'Enquête sur les ménages et l'environnement.....	13
<b>6.0</b>	<b>Collecte des données</b> .....	<b>13</b>
7.0	Traitement des données .....	14
7.1	Saisie des données.....	15
7.2	Vérification .....	15
7.3	Codage des questions ouvertes .....	15
7.4	Création de variables dérivées .....	16
7.5	Pondération.....	17
7.6	Suppression de renseignements confidentiels .....	17
<b>8.0</b>	<b>Qualité des données</b> .....	<b>17</b>
8.1	Taux de réponse .....	17
8.2	Erreurs relatives à l'enquête .....	18
8.2.1	Base de sondage .....	19
8.2.2	Collecte des données .....	19
8.2.3	Traitement des données .....	19
8.2.4	Non-réponse .....	19
8.2.5	Mesure de l'erreur d'échantillonnage.....	20
<b>9.0</b>	<b>Pondération</b> .....	<b>20</b>
9.1	Procédures de pondération pour l'Enquête sur la santé dans les collectivités canadiennes.....	20
9.2	Procédures de pondération pour l'Enquête sur les ménages et l'environnement .....	23
<b>10.0</b>	<b>Lignes directrices pour la totalisation, l'analyse et la diffusion de données</b> .....	<b>25</b>
10.1	Lignes directrices pour l'arrondissement d'estimations .....	25
10.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation .....	26
10.3	Lignes directrices pour l'analyse statistique.....	26
10.4	Lignes directrices pour la diffusion de coefficients de variation.....	27
<b>Annexe A – Estimation de la variance pour les fichiers maîtres et partagés</b> .....		<b>29</b>
<b>Annexe B – Estimation de la variance pour les fichiers de microdonnées à grande diffusion</b> .....		<b>30</b>



## 1.0 Introduction

Statistique Canada a mené l'Enquête sur les ménages et l'environnement (EME) d'octobre à novembre 2009. Ce manuel a été produit pour faciliter la manipulation du fichier de microdonnées portant sur les résultats de l'enquête.

Toutes les questions concernant l'ensemble de données ou son utilisation devraient être adressées à :

### Statistique Canada

Division des comptes et de la statistique de l'environnement

Téléphone : 613-951-0297

Télécopieur : 613-951-0634

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

## 2.0 Contexte

Supplément à l'Enquête sur la santé dans les collectivités canadiennes (ESCC), l'Enquête sur les ménages et l'environnement (EME) a été menée en octobre et novembre 2009. Cette enquête a été spécialement conçue pour répondre aux besoins de sa source de financement, l'initiative Indicateurs canadiens de durabilité de l'environnement (ICDE), fruit d'un partenariat entre Statistique Canada, Environnement Canada et Santé Canada. L'ICDE établit des rapports annuels sur la qualité de l'air et de l'eau ainsi que sur les émissions de gaz à effet de serre au Canada, à l'aide d'indicateurs permettant de repérer les domaines prioritaires pour les Canadiens et d'enregistrer les progrès accomplis.

Menée pour la première fois en 1991, l'EME a été réalisée de nouveau en 1994 et plus récemment en 2006 et 2007. L'enquête de 2009 dresse un bilan plus complet des comportements des ménages en ce qui a trait à l'environnement, tout en permettant d'effectuer des comparaisons avec certains indicateurs de l'EME de 1994 ainsi que la plupart des indicateurs issus de l'enquête de 2006 et 2007.

La population cible de l'enquête se compose des ménages au Canada à l'exclusion de ceux situés au Yukon, dans les Territoires du Nord-Ouest et au Nunavut, des ménages situés sur les réserves indiennes ou les terres publiques et ceux formés entièrement de membres à temps plein des Forces canadiennes. En outre, on a exclu les établissements et les ménages installés dans certaines régions éloignées.

## 3.0 Objectifs

L'objectif de l'Enquête sur les ménages et l'environnement (EME) est de mesurer les comportements et pratiques des ménages en matière d'environnement qui ont à voir avec la qualité de l'air, de l'eau et du sol ainsi que sur leurs contributions aux émissions de gaz à effet de serre. Plus précisément, les thèmes abordés dans le cadre de l'EME de 2009 ont été les suivants :

- la consommation d'énergie ainsi que la climatisation et le chauffage domestique;
- la consommation et la conservation de l'eau;
- les préoccupations des ménages à l'égard de la qualité de l'eau;
- l'utilisation d'engrais et de pesticides sur les pelouses et jardins;
- le compostage;
- l'environnement intérieur;
- les véhicules récréatifs et équipement muni d'un moteur;
- l'élimination des résidus domestiques dangereux;
- l'impact de la qualité de l'air et de l'eau sur les ménages;
- les décisions en matière d'achats.

## 4.0 Concepts et définitions

Ce chapitre donne un aperçu des concepts et des définitions d'intérêt pour les utilisateurs. Les concepts et les définitions utilisés dans le cadre de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) sont décrits à la section 4.1, tandis que ceux et celles propres à l'Enquête sur les ménages et l'environnement (EME) sont fournis dans la section 4.2.

### 4.1 Concepts et définitions de l'Enquête sur la santé dans les collectivités canadiennes

#### Logement

Un logement est défini comme un ensemble de pièces d'habitation distinct au niveau de la structure des pièces d'habitation d'autres logements, possédant une entrée privée avec accès sur l'extérieur ou sur un corridor ou un escalier commun à l'extérieur.

#### Type de logement

- **Maison individuelle non attenante** – structure renfermant un seul logement et qui n'est attenante à aucune autre structure, sauf un garage ou un hangar.
- **Maison double** – logement attenant à un seul autre logement dont il est séparé par un mur s'élevant du sol au toit.
- **Maison en rangée** – logements dans une rangée de trois logements ou plus séparés les uns des autres par des murs mitoyens s'élevant du sol au toit, sans autre logement au-dessus ni en dessous.
- **Duplex** – logement faisant partie d'un ensemble de deux logements superposés, non attenants à une autre structure et entourés de tous les côtés d'espaces libres.
- **Appartement dans un immeuble de moins de cinq étages** – logements situés dans les triplex, les quadruplex et les immeubles d'appartements de moins de cinq étages.
- **Appartement dans un immeuble de cinq étages ou plus** – logements distincts dans une structure résidentielle de cinq étages ou plus.
- **Maison mobile** – logement mobile conçu et construit pour être transporté sur son propre châssis et placé sur des fondations temporaires (blocs, pieux ou emplacement aménagé).

#### Ménage

Un ménage désigne toute personne ou tout groupe de personnes vivant dans un même logement. Un ménage peut se composer d'une personne vivant seule, d'un groupe de personnes non apparentées ou d'une famille.

### 4.2 Concepts et définitions de l'Enquête sur les ménages et l'environnement

#### Consommation et conservation de l'eau

« Les Canadiens sont préoccupés par les répercussions que peut avoir l'environnement sur la qualité de l'eau qu'ils boivent<sup>1</sup>. » Étant donné que les perceptions du public, comme facteur déterminant de la politique publique, peuvent être aussi persuasives que les données empiriques, il importe de comprendre comment les Canadiens perçoivent la qualité de leur approvisionnement en eau potable, ainsi que la manière dont ils se comportent en réaction, peut-être, à leurs préoccupations.

---

<sup>1</sup> *Partage des décisions en matière d'environnement : Sommaire et recommandations*. Rapport final du groupe de travail sur le Système canadien d'information pour l'environnement. Octobre 2001, Ottawa.

L'EME mesure ces comportements en considérant les achats d'eau embouteillée ou l'utilisation de filtres à eau ainsi que les motifs de ces achats (crainte de la contamination bactérienne, par exemple). Une analyse peut être effectuée pour mieux comprendre ce qui distingue les ménages dont les comportements indiquent un manque de confiance quant à la qualité de leur eau potable des ménages qui n'affichent pas ces comportements.

À cette question de l'eau potable s'ajoute l'important thème des pratiques des ménages en matière de conservation de l'eau. Le phénomène de la rareté de l'eau commence à s'imposer aux yeux de bien des Canadiens, souci qui peut venir aggraver le phénomène des changements climatiques. Maintes régions du pays ont connu des conditions de sécheresse ou de quasi-sécheresse qui ont incité les autorités municipales à se doter de règlements (restreignant la consommation d'eau, par exemple) et/ou les ménages à prendre librement des mesures de conservation.

L'EME livre une information fiable sur les pratiques des ménages (restrictions d'arrosage des pelouses, utilisation de pommeaux de douche à faible débit, de toilette à faible volume d'eau, etc.), ce qui pourra servir à juger du degré de participation populaire au mouvement de conservation de l'eau.

Les **métaux et minéraux** comprennent les substances suivantes : fer, soufre, cadmium, zinc, manganèse, plomb, mercure, arsenic.

Les **bactéries** comprennent les substances suivantes : E. coli, autres coliformes, matières fécales, giardia, cryptosporidium, parasites, protozoaire, shigella.

Les **produits chimiques et autres polluants** comprennent les substances suivantes : chlore, brome, pesticides, huile, essence, essence diesel, mazout domestique, fluorure, nitrate, tétrachloroéthane (TCE), hydrocarbures polyaromatiques (HAP), ignifuges.

Les **bassins de rétention** sont des fosses septiques qui ne sont pas dotées d'un réseau de dalots et qui doivent être vidées régulièrement.

Une **fosse septique collective** est une installation septique privée ou publique qui dessert plus d'un ménage, mais ne fait pas partie d'un réseau d'égouts municipal. Ces installations sont fréquentes dans les endroits comme les parcs de roulottes ou les quartiers où la densité de construction n'est pas suffisamment élevée pour justifier des services complets d'épuration des eaux.

Les **pommes de douche à faible débit** servent à contrôler le débit de l'eau.

Les **toilettes à faible chasse d'eau** utilisent un volume d'eau plus faible que les toilettes ordinaires. Ces toilettes utilisent habituellement 6 litres d'eau par chasse, au lieu de 12 litres.

Un **réservoir d'eau de pluie** est un contenant tel un baril qui sert à recueillir et à stocker l'eau de pluie. Le contenant est habituellement placé sous la descente d'une gouttière. L'eau recueillie est habituellement utilisée pour l'arrosage.

Une **citerne** est un réservoir artificiel qui sert à stocker les liquides. Le terme désigne en particulier un réservoir souterrain servant au stockage de l'eau de pluie.

### **Consommation d'énergie et chauffage domestique**

Le choix de la source d'énergie de chauffage domestique que fait un ménage, les réglages de la température du logement et le recours à des appareils électriques qui économisent l'énergie, voilà autant de décisions qui influent sur ce qu'apporte ce ménage aux émissions de gaz à effet de serre (GES).

Indirectement, de telles décisions nous indiquent aussi dans quelle mesure un ménage a fait

sienne l'idée et la nécessité de la conservation de l'énergie. L'EME mesure non seulement les types d'énergie consommés mais aussi les comportements qui révèlent si les ménages canadiens ont un souci de durabilité dans l'énergie qu'ils consomment.

Une **fornaise au gaz à air forcé** est une installation de chauffage qui utilise un ensemble de conduites et de sorties d'air pour faire circuler l'air chauffé par le gaz naturel en combustion.

Une **fornaise au mazout à air forcé** est une installation de chauffage qui utilise un ensemble de conduites et de sorties d'air pour faire circuler l'air chauffé par le mazout en combustion.

Une **fornaise électrique à air forcé** est une installation de chauffage qui utilise un ensemble de conduites et de sorties d'air pour faire circuler l'air chauffé par un courant électrique.

Une **fornaise à l'eau chaude à air forcé** est une installation de chauffage qui utilise un ensemble de conduites et de sorties d'air pour faire circuler l'air chauffé par l'eau chaude.

Les **radiateurs à l'eau chaude** sont des structures ou des pièces d'équipement en métal qui servent à chauffer une pièce en rayonnant de la chaleur à partir de l'eau chaude ou de la vapeur qui circule à l'intérieur.

Les **plinthes électriques** sont des appareils de chauffage fixés aux murs, près du plancher, dont les éléments sont chauffés par le passage d'un courant électrique. Le réglage des plinthes électriques s'effectue au moyen de thermostats indépendants – à raison d'un par pièce, habituellement.

Par **autre chauffage électrique**, il faut entendre la chaleur produite par un courant électrique acheminé par un appareil ou un autre moyen, à l'exception de l'air forcé ou des plinthes électriques.

Un **poêle à bois ou un foyer au bois** sont des lieux où vous pouvez faire du feu en utilisant le bois comme combustible. Le poêle ou le foyer sont raccordés à une cheminée qui permet d'évacuer à l'extérieur la fumée et l'excès de chaleur.

Un **climatiseur central** est une composante de l'installation de chauffage central de l'habitation et distribue l'air frais dans le logement au moyen de conduites, par opposition aux climatiseurs autonomes qui sont habituellement installés dans les fenêtres et qui servent à rafraîchir des endroits précis du logement.

#### **Équipement d'entretien de la pelouse**

Les moteurs à combustion interne, de petits appareils tels les tondeuses à gazon et certains véhicules récréatifs, émettent des gaz à effet de serre (GES), que l'on considère comme le principal facteur de transformation du climat. En mesurant le degré d'utilisation de tels appareils, nous établissons des données de référence que nous pourrions comparer à de futures données pour constater si cette utilisation augmente ou diminue.

Un **taille-bordure** est un appareil qui sert à couper les mauvaises herbes au moyen d'un cordon en plastique ou d'une scie circulaire à rotation rapide.

Une **souffleuse à feuilles** est un appareil qui produit un fort courant d'air servant à souffler les débris présents sur la pelouse.

#### **Utilisation d'engrais et de pesticides**

Dans l'EME de 2009, on mesure l'usage que font les ménages d'engrais et de pesticides. En

analysant ces données avec certaines caractéristiques des ménages, les décideurs pourront mieux éclairer les campagnes ciblées de sensibilisation et d'information du public.

Les **engrais chimiques** sont des produits chimiques qui sont administrés aux plantes dans le but de favoriser leur croissance. Ils sont habituellement appliqués directement dans le sol ou par vaporisation.

Les **herbicides, les insecticides et les fongicides** sont des substances ou des mélanges de substances qui visent à prévenir, détruire, repousser ou atténuer les mauvaises herbes, les insectes ou les prolifération fongiques.

### **Élimination des déchets dangereux**

Nous avons besoin de meilleures données sur les pratiques des ménages dans leur gestion des déchets dangereux. C'est une question qui ressort des contacts avec les représentants provinciaux et qui est aussi à l'ordre du jour au niveau fédéral<sup>2</sup>. Les soucis s'avivent au sujet de la quantité de substances toxiques (plomb, mercure, dioxines, etc.) que peuvent renfermer ces déchets s'intensifient et les décideurs étudient les méthodes d'atténuation des effets de ces matières sur l'environnement. L'EME détermine si les ménages éliminent les déchets dangereux de la bonne façon, s'ils les mettent tout simplement aux ordures ménagères ou s'ils restent peut-être simplement dans l'ignorance des modes d'élimination et se contentent de ranger les produits en question dans les sous-sols et les garages.

### **Pratiques de compostage**

Les statistiques sur les déchets accusent une autre lacune, puisqu'elles ne mesurent pas l'utilisation ménagère de composteurs domestiques ni le recours aux services de collecte de matières organiques. Les paliers de gouvernement et les ONG (organisations non gouvernementales) s'intéressent tous grandement à cette utilisation et aux caractéristiques socio-économiques des utilisateurs et des non-utilisateurs de composteurs domestiques et de programmes centralisés de compostage. Le compostage consiste à séparer les déchets domestiques (par ex. les restes d'aliments, les grains de café, les coquilles d'œufs, etc.) et/ou les résidus de la cour (par ex. les feuilles, les plantes ou les tontes de gazon) du reste des déchets de votre ménage. Les matériaux séparés peuvent être :

- placés dans un bac de compostage, empilés ou répandus dans le jardin;
- recueillis par la ville, village, municipalité ou une entreprise privée; ou
- apportés à un centre de collecte.

### **Qualité de l'air et de l'eau**

La qualité de l'air et des plans d'eau (lacs et rivières) à usage récréatif influe peut-être sur le comportement des Canadiens dans leurs activités quotidiennes. L'absence de réforme des comportements peut avoir des répercussions sur la qualité de vie et la santé de tous les Canadiens. L'EME établit, par exemple, dans quelle mesure les avertissements de smog ou les interdictions de baignade influencent les activités des gens.

Le **smog** est la forme la plus visible de pollution de l'air. Il consiste en une brume d'un brun jaunâtre qui se forme lorsque la chaleur et les rayons du soleil réagissent aux différents polluants dans l'air. Le smog est un problème à l'année, mais la plupart des alertes au smog se produisent d'avril à septembre, en particulier par temps chaud.

Un **avis sur la qualité de l'air** est un avertissement émis lorsque la fumée, le smog, ou la mauvaise qualité de l'air atteignent des niveaux pouvant être nocifs. Les alertes sont fondées sur les indices de qualité de l'air.

---

<sup>2</sup> *Information Technology (IT) and Telecommunication Waste in Canada*, EnviroRIS, document produit pour le Bureau national de la prévention de la pollution d'Environnement Canada. Octobre 2000, Ottawa.

## **5.0 Méthodologie de l'enquête**

L'Enquête sur les ménages et l'environnement (EME) a été menée d'octobre à novembre 2009 auprès d'un sous-échantillon des logements inclus dans l'échantillon de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) – composante annuelle 2009 entre le 1 janvier et le 30 juin 2009. Son plan de sondage est donc étroitement lié à celui de l'ESCC. Le plan de l'ESCC est décrit brièvement dans les sections 5.1 et 5.2.

### **5.1 Population visée par l'Enquête sur la santé dans les collectivités canadiennes**

Les données de l'ESCC sont recueillies auprès des personnes de 12 ans et plus vivant dans des logements privés des dix provinces et trois territoires. Sont exclus de la base de sondage les habitants des réserves indiennes et des terres de la Couronne, les personnes vivant en institution, les membres à temps plein des Forces armées canadiennes et les habitants de certaines régions éloignées. L'ESCC couvre environ 98 % de la population canadienne âgée de 12 ans et plus.

### **5.2 Plan de sondage de l'Enquête sur la santé dans les collectivités canadiennes**

Afin de produire des estimations fiables pour les 121 régions socio-sanitaires (RSS) et compte tenu du budget accordé pour l'ESCC de 2009, il a été établi que cette composante de l'enquête devrait être réalisée auprès d'un échantillon de 65 000 personnes. La stratégie de répartition de l'échantillon, qui comporte trois étapes, accorde une importance plus ou moins égale aux RSS et aux provinces. Lors des deux premières étapes, l'échantillon a été réparti entre les provinces en fonction de leur compte de population et du nombre de RSS qu'elles contiennent. À la troisième étape, chaque échantillon provincial a été réparti entre les RSS proportionnellement à la racine carrée de la population estimée de la RSS.

L'ESCC a utilisé trois bases de sondage pour sélectionner les échantillons de ménages : 50 % de l'échantillon des ménages est venu de la base aréolaire, 49 % est venu d'une base liste de numéros de téléphone et le 1 % restant est venu d'une base d'échantillonnage à composition aléatoire (CA). Pour la plupart des régions socio-sanitaires, 50 % de l'échantillon a été sélectionné à partir d'une base aréolaire et 50 % d'une base liste de numéros de téléphone.

La base aréolaire conçue pour l'Enquête sur la population active (EPA) du Canada a servi de base de sondage principale pour l'ESCC. Le plan d'échantillonnage de l'EPA est un plan d'échantillonnage en grappes stratifié à plusieurs degrés où le logement représente l'unité finale d'échantillonnage. À la première étape, on a formé des strates homogènes et sélectionné un échantillon indépendant de grappes, dans chaque strate. À la deuxième étape, on a dressé la liste des logements pour chaque grappe, puis on a sélectionné les logements, ou les ménages, d'après les listes.

Pour les besoins du plan d'échantillonnage, chaque province est répartie en trois catégories de région, à savoir les grands centres urbains, les villes et les régions rurales. Des strates géographiques ou socioéconomiques sont formées à l'intérieur de chaque grand centre urbain. Dans les strates, des grappes sont formées par regroupement. Dans certains centres urbains, des strates distinctes sont créées pour les immeubles à appartements ou les aires de diffusion (AD) pour lesquelles le revenu moyen du ménage est élevé. Dans chaque strate, on sélectionne six grappes ou immeubles résidentiels (pouvant compter de 12 à 18 appartements) par une méthode d'échantillonnage aléatoire avec probabilité proportionnelle à la taille (PPT), cette dernière correspondant au nombre de ménages. Le nombre 6 est utilisé pour l'ensemble du plan

d'échantillonnage afin de permettre le renouvellement mensuel d'un sixième de l'échantillon de l'EPA.

Les autres villes et régions rurales de chaque province sont stratifiées, en premier lieu, en fonction de données géographiques, puis selon les caractéristiques socioéconomiques. Dans la plupart des strates, on sélectionne six grappes (habituellement des AD) par la méthode PPT. Pour les strates où la densité de population est faible, on suit un plan en trois étapes en vertu duquel on sélectionne deux ou trois unités primaires d'échantillonnage (UPE), qui correspondent normalement à des groupes de AD, puis on les répartit en grappes dont six sont sélectionnées pour faire partie de l'échantillon. On obtient l'échantillon final par échantillonnage systématique des logements.

### 5.3 Taille de l'échantillon selon la province pour l'Enquête sur les ménages et l'environnement

Le tableau ci-dessous indique le nombre de logements inclus dans l'échantillon prélevé aux fins de l'EME de 2009. Sont exclus de ce tableau les logements qui avaient été non répondants à l'ESCC.

Province	Taille de l'échantillon – nombre de logements
Terre-Neuve-et-Labrador	673
Île-du-Prince-Édouard	411
Nouvelle-Écosse	780
Nouveau-Brunswick	784
Québec	4 010
Ontario	7 542
Manitoba	876
Saskatchewan	987
Alberta	1 721
Colombie-Britannique	2 216
<b>Canada</b>	<b>20 000</b>

## 6.0 Collecte des données

Une lettre d'introduction a été postée aux répondants environ une semaine avant le début de la collecte des données qui a eu lieu en octobre et novembre 2009. Les données ont été recueillies à l'aide d'interviews téléphoniques assistées par ordinateur (ITAO).

Le système ITAO contient plusieurs modules génériques qui s'adaptent rapidement à la plupart des types d'enquêtes. Un module frontal contient un ensemble de codes de réponse standards, qui s'appliquent à toutes les issues possibles des appels, ainsi que les scénarios correspondants qui sont lus par les intervieweurs. Une approche normalisée a été utilisée pour présenter l'organisme, le nom et le but de l'enquête, les clients de l'enquête, l'utilisation qui sera faite des résultats et la durée de l'interview. Nous avons expliqué aux répondants comment ils avaient été sélectionnés pour l'enquête, que leur participation à l'enquête était volontaire et que les renseignements fournis resteraient strictement confidentiels. Les intervieweurs avaient accès à des écrans d'aide grâce auxquels ils pouvaient répondre aux questions les plus fréquemment posées par les répondants.

L'application ITAO garantissait l'entrée des seules réponses valides et le bon enchaînement des questions. Des contrôles intégrés à l'application garantissaient la cohérence des réponses, repéraient et corrigeaient les valeurs aberrantes et déterminaient à qui étaient posées les différentes questions. Ainsi, à la fin du processus de collecte, les données étaient déjà passablement « épurées ».

Le gestionnaire de l'enquête a rencontré les cadres supérieurs responsables de la collecte pour discuter des problèmes et des questions avant le début de la séance de formation. Le Manuel de l'intervieweur comportait une description du contexte et les objectifs, ainsi qu'une description détaillée des concepts et des définitions propres à l'EME de 2009. Il renfermait également un glossaire et une série de questions et réponses.

Les intervieweurs ont reçu leur formation sur le contenu de l'enquête dans le cadre d'une séance de formation en classe. De plus, ils ont effectué une série d'interviews simulées pour se familiariser avec l'enquête et ses concepts, définitions et sur l'utilisation de l'application ITAO. La documentation sur les questions et réponses a été fournie aux intervieweurs afin de les aider à répondre aux questions fréquentes des répondants.

La collecte des données a été réalisée par le personnel spécialisé du bureau de Statistique Canada à Edmonton, Sturgeon Falls, Toronto, Halifax, Winnipeg et Sherbrooke. La charge de travail et les intervieweurs de chaque bureau étaient supervisés par un chargé de projet. L'ordonnanceur automatique utilisé dans le système ITAO garantissait que les cas étaient assignés au hasard aux intervieweurs et que les appels se faisaient à différents moments de la journée pendant des jours différents de la semaine pour maximiser la probabilité de contact. Un maximum de 25 appels par cas. Une fois le maximum atteint, un intervieweur principal examinait le cas et déterminait si d'autres appels seraient tentés.

La durée moyenne d'une interview a été estimée à 20 minutes. Toutefois, la durée des interviews fluctuait selon le type de répondant. Par exemple, la durée moyenne de l'interview était légèrement plus élevée pour les ménages résidant dans une maison unifamiliale par opposition à un ménage demeurant dans un immeuble à appartements.

Tous les intervieweurs travaillent sous la supervision d'un groupe d'intervieweurs principaux. Ceux-ci ont pour responsabilité de s'assurer que les intervieweurs connaissent bien les concepts et les procédures de l'enquête. Ils ont aussi pour fonction de contrôler régulièrement le travail des intervieweurs et d'examiner les documents complétés par ces derniers.

## **7.0 Traitement des données**

Le principal produit de l'Enquête sur les ménages et l'environnement (EME) est un fichier de microdonnées « épuré ». Ce chapitre présente un bref résumé des phases de traitement inhérentes à la production de ce fichier.

Le fichier de microdonnées contient des données portant sur les sujets suivants :

- HH Données démographiques sur les ménages
- DC Caractéristiques du logement
- EH Consommation d'énergie et chauffage domestique
- WA Eau
- FP Utilisation d'engrais et de pesticides
- GP Véhicules récréatifs/Équipement d'extérieur
- CP Compostage
- IE Environnement intérieur
- AQ Qualité de l'air
- HW Déchets dangereux

PD Décisions en matière d'achats  
HD Revenu

## **7.1 Saisie des données**

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

Une partie du contrôle se fait au moment de l'interview. Lorsque les renseignements introduits sont hors limites (trop faibles ou trop élevés) des valeurs attendues, ou qu'ils entrent en contradiction avec des renseignements introduits auparavant, l'intervieweur voit apparaître 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 le répondant ne connaît pas la réponse ou refuse de répondre. Pour cette raison, on soumet les données des réponses à d'autres processus de vérification après réception au bureau central.

## **7.2 Vérification**

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

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

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

Ces mesures correctives ont été suivies d'une série de vérifications pour assurer la cohérence des réponses au sein de chaque ménage.

## **7.3 Codage des questions ouvertes**

Quelques postes du questionnaire ont été remplis par l'intervieweur sous forme de questions ouvertes. Ces questions nécessitaient un codage pour inclusion dans le fichier de données de l'EME.

Le deuxième type de codage effectué concernait les questions permettant d'entrer des valeurs numériques. On vérifiait d'abord les chiffres pour éliminer les valeurs aberrantes, puis on les regroupait dans des fourchettes. Un exemple de question numérique ouvertes serait le revenu du ménages.

## 7.4 Création de variables dérivées

Un certain nombre de variables incluses dans le fichier de microdonnées ont été calculées en combinant des variables pour faciliter l'analyse des données. La liste qui suit énumère les variables dérivées de l'EME.

### Variables dérivées de l'EME

WAD04	Indication si des traitements sont utilisés pour l'eau potable dans le ménage.
WADREDUC	Indication si des dispositifs sont utilisés par le ménage pour économiser ou réduire la consommation d'eau.
EHD11	Pendant la période hivernale, à quelle température réglez-vous habituellement le chauffage lorsque vous êtes chez vous et réveillé? (Celsius)
EHD12	Pendant la période hivernale, à quelle température réglez-vous habituellement le chauffage lorsque vous dormez? (Celsius)
EHD13	Pendant la période estivale, à quelle température réglez vous habituellement votre climatiseur lorsque vous êtes chez vous et réveillé? (Celsius)
EHD14	Pendant la saison estivale, à quelle température réglez vous habituellement votre climatiseur lorsque vous dormez? (Celsius)
EHD15	Pendant la saison estivale, à quelle température réglez vous habituellement votre climatiseur lorsque vous n'êtes pas à la maison? (Celsius)
EHD12A	Réglage de la température pendant le sommeil durant la période hivernale
EHD14A	Réglage de la température pendant le sommeil durant la période estivale
GPDEQUIP	Utilisation de chasse-neige, tondeuse, coupe-bordures ou souffeuse à feuilles.
CMA	Région métropolitaine de recensement (RMR)- Code du Recensement de 2006
HHEDUCLV HHGEDUC	Plus haut niveau de scolarité atteint par l'un ou l'autre des membres du ménage.
HHAG0005	Nombre de personnes âgées de 0 à 5 ans dans le ménage (Fichiers maître et partagé)
HHAG0612	Nombre de personnes âgées de 6 à 12 ans dans le ménage (Fichiers maître et partagé)
HHAG1315	Nombre de personnes âgées de 13 à 15 ans dans le ménage (Fichiers maître et partagé)
HHAG1617	Nombre de personnes âgées de 16 à 17 ans dans le ménage (Fichiers maître et partagé)
HHAG1819	Nombre de personnes âgées de 18 à 19 ans dans le ménage (Fichiers maître et partagé)
HHAG2024	Nombre de personnes âgées de 20 à 24 ans dans le ménage (Fichiers maître et partagé)
HHAG2534	Nombre de personnes âgées de 25 à 34 ans dans le ménage (Fichiers maître et partagé)
HHAG3544	Nombre de personnes âgées de 35 à 44 ans dans le ménage (Fichiers maître et partagé)
HHAG4554	Nombre de personnes âgées de 45 à 54 ans dans le ménage (Fichiers maître et partagé)
HHAG5564	Nombre de personnes âgées de 55 à 64 ans dans le ménage (Fichiers maître et partagé)
HHAG65PL HHGAG65P	Nombre de personnes âgées de 65 ans et plus dans le ménage
HHSIZE HHGSIZE	Nombre de personnes dans le ménage
HHTYPE	Type de ménage, basé sur la composition par âge (Fichiers maître et partagé)
HDD02 HGD02	Revenu du ménage

## **7.5 Pondération**

Le principe qui sous-tend une estimation pour un échantillon probabiliste comme celui de l'EME veut que chacune des unités incluses dans l'échantillon « représente », plusieurs autres unités qui en sont exclues. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque unité incluse dans l'échantillon représente 50 unités 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 ménages qui traitent l'eau qu'ils boivent doit être estimé, cette opération est effectuée au moyen de la sélection des enregistrements se référant aux ménages inclus à l'intérieur de l'échantillon qui présentent cette caractéristique et de l'addition des poids inscrits dans ces enregistrements.

Le chapitre 9.0 renferme des détails au sujet de la méthode utilisée pour calculer ces poids.

## **7.6 Suppression de renseignements confidentiels**

Il convient de souligner que les fichiers de microdonnées « à grande diffusion » (FMGD) peuvent différer des fichiers « maîtres » de l'enquête que conserve Statistique Canada. Ces différences sont habituellement le résultat de mesures prises pour protéger l'anonymat des répondants à une enquête. Les mesures les plus courantes sont la suppression de variables du fichier, le regroupement de valeurs en des catégories plus étendues et le codage de valeurs spécifiques à la catégorie « non déclaré ». Les utilisateurs ayant besoin d'avoir accès à de l'information exclue des fichiers de microdonnées peuvent acheter des totalisations spéciales. Les estimations produites seront communiquées à l'utilisateur, sous réserve du respect des lignes directrices pour l'analyse et la diffusion dont le chapitre 10.0 de ce document fournit un aperçu.

## **8.0 Qualité des données**

### **8.1 Taux de réponse**

Le tableau qui suit renferme un résumé des taux de réponse au questionnaire de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) et à celui de l'Enquête sur les ménages et l'environnement (EME).

Province	Ménages sélectionnés de l'ESCC	Taux de réponse à l'ESCC (%)*	Ménages sélectionnés à l'EME	Ménages répondants à l'EME	Taux de réponse à l'EME (%)**
Terre-Neuve-et-Labrador	1 242	87,0	673	476	70,7
Île-du-Prince-Édouard	642	86,4	411	294	71,5
Nouvelle-Écosse	1 554	86,0	780	584	74,9
Nouveau-Brunswick	1 517	86,2	784	568	72,4
Québec	7 875	81,4	4 010	2 948	73,5
Ontario	14 692	80,2	7 542	5 692	75,5
Manitoba	2 204	84,6	876	663	75,7
Saskatchewan	2 433	84,5	987	762	77,2
Alberta	3 935	79,8	1 721	1 241	72,1
Colombie-Britannique	5 084	78,9	2 216	1 526	68,9
<b>Canada</b>	<b>41 178</b>	<b>81,5</b>	<b>20 000</b>	<b>14 754</b>	<b>73,8</b>

\* Le taux de réponse à l' ESCC est le nombre de ménages ayant répondu au questionnaire de l' ESCC exprimé sous forme de pourcentage du nombre de ménages sélectionnés pour la période entre le 1 janvier et le 30 juin 2009.

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

## 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 participants peuvent se tromper en répondant aux questions, les réponses peuvent être mal saisies sur le questionnaire et des erreurs peuvent survenir lors du traitement et de la totalisation des données. Ces erreurs sont toutes des exemples d'erreurs non dues à l'échantillonnage.

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

### **8.2.1 Base de sondage**

Comme l'EME de 2009 était une enquête supplémentaire à l'ESCC de 2009, la base de sondage de l'ESCC fut utilisée. La base de sondage de l'ESCC était à la fois constituée de la base aréolaire de l'EPA et d'une base de sondage téléphonique (incluant une composante à composition aléatoire). La couverture de l'ESCC est très bonne. Elle est estimée à environ 98 % des ménages canadiens. Il est peu probable que l'exclusion de 2 % des ménages canadiens introduise un biais significatif dans les données d'enquêtes.

Il est à noter que les interviews de l'EME ont été menées 3 mois après la collecte de l'ESCC.

### **8.2.2 Collecte des données**

La formation des intervieweurs pour l'EME consistait à lire le Manuel du surveillant, le Manuel des procédures, le Manuel de l'intervieweur, la pratique avec des cas simulés de l'EME sur ordinateur (pour IAO seulement) et une discussion avec l'intervieweur principal des questions de l'intervieweur avant le début de l'enquête. L'information fournie aux intervieweurs comprenait une description du contexte et des objectifs de l'enquête ainsi qu'un glossaire de terminologie et une série de questions et de réponses. Les intervieweurs ont recueilli l'information pour l'EME après avoir recueilli l'information de l'ESCC. Les interviews ont été menées durant les d'octobre 2009 et novembre 2009.

### **8.2.3 Traitement des données**

Le traitement des données de l'EME a été fait par étapes, y compris la vérification, le codage, le contrôle, l'estimation, 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 ponctuelle (une non-réponse à une question), une non-réponse partielle (le fait de ne pas répondre à plusieurs questions) à une non-réponse complète. S'il y a eu non-réponse complète, c'est parce que l'intervieweur a été incapable de communiquer avec le répondant, qu'aucun membre du ménage n'a pu fournir l'information demandée ou que le répondant a refusé de participer à l'enquête. Les cas de non-réponse complète ont été traités en ajustant le poids des ménages ayant 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 ponctuelle 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. Dans les cas de non-réponse ponctuelle, on n'a pas imputé de valeur; on les a codées comme « non déclaré ».

Il y a non-réponse partielle lorsqu'une interview a été commencée mais n'a pas été terminée, pour diverses raisons. Dans le cadre de l'EME, cela s'est produit dans moins

de 1 % des cas et les réponses manquantes ont été considérées comme une série de non-réponse ponctuelle et ont été codées comme « non déclaré ».

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

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

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

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

Par exemple, supposons que, d'après les résultats de l'enquête de 2006, l'on estime que 34,9 % des ménages avaient une pelouse en 2005 et utilisaient des engrais chimiques et l'on constate que l'erreur-type de cette estimation est de 0,0051. Le coefficient de variation de l'estimation est donc calculé comme suit :

$$\left( \frac{0,0051}{0,349} \right) \times 100\% = 1,46\%$$

De plus amples renseignements sur le calcul du coefficient de variation se trouvent au chapitre 10.0 ainsi que dans les annexes A et B.

## **9.0 Pondération**

Puisque l'Enquête sur les ménages et l'environnement (EME) faisait appel à un sous-échantillon de l'échantillon de l'Enquête sur la santé dans les collectivités canadiennes (ESCC), le calcul des poids pour les enregistrements des données de l'enquête est clairement lié à la procédure de pondération utilisée aux fins de l'ESCC. La procédure de pondération employée pour l'ESCC est décrite brièvement ci-dessous.

### **9.1 Procédures de pondération pour l'Enquête sur la santé dans les collectivités canadiennes**

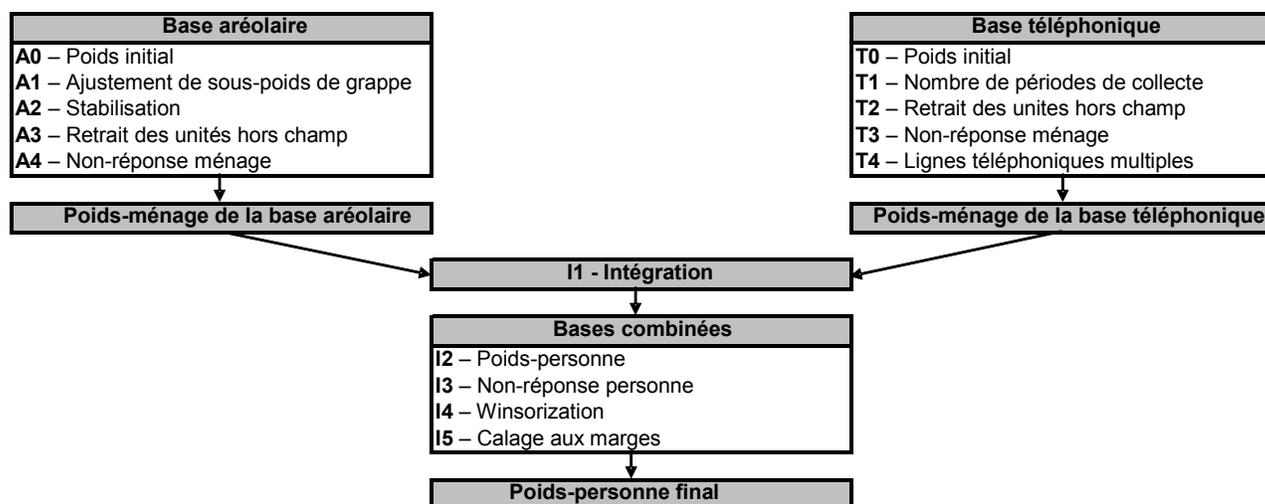
L'ESCC a recours à trois bases de sondage pour la sélection de son échantillon : une base aréolaire de logements agissant comme base principale, puis deux bases formées de numéros de téléphone utilisées pour compléter la base aréolaire. Puisque seulement quelques différences mineures distinguent les deux bases de numéros de téléphone pour la pondération, elles ont été traitées ensemble. On réfère à celles-ci comme faisant partie de la base téléphonique.

La stratégie de pondération a été développée en traitant séparément la base aréolaire et la base téléphonique. Les poids des ménages résultant de ces deux bases sont ensuite combinés en un

seul ensemble de poids des ménages lors d'une étape appelée « *intégration* ». Suite à la transformation des poids des ménages en poids de personne et à quelques autres ajustements, ce poids intégré devient le poids de personne final. Noter que, selon le besoin, une seule ou deux bases peuvent être utilisées pour la sélection de l'échantillon dans une région socio-sanitaire (RSS) donnée. La stratégie de pondération tient compte de cette particularité lors de l'étape d'intégration.

**Note:** Le sous-poids ménage de l'ESCC (après l'étape **I1 – Intégration**) correspond au poids initial de l'EME.

#### Diagramme A : Sommaire de la stratégie de pondération de l'ESCC



Le sous-poids ménage de l'ESCC (requis pour la pondération de l'EME) est disponible lorsque les étapes **A0** à **A4** (pour la base aréolaire), **T0** à **T4** (pour la base téléphonique) et **I1** (intégration) sont complétées. Chacune des étapes de pondération de l'ESCC est décrite ci-dessous.

#### Pondération de l'échantillon provenant de la base aréolaire

##### **A0 – Poids initial**

La pondération pour la base aréolaire débute avec un poids fourni par l'Enquête sur la population active (EPA). Le poids initial (**A0**) est basé sur le plan de sondage de l'EPA puisque l'échantillon de l'ESCC découle de celui de l'EPA. Le plan de sondage de l'EPA consiste en un échantillonnage de logements dans les grappes sélectionnées parmi les strates de l'EPA.

##### **A1 – Ajustement de sous-poids de grappe**

Dans les grappes où un accroissement significatif de la population est observé, une méthode de sous-échantillonnage est utilisée afin de ne pas augmenter indûment la tâche de l'interviewer. La méthode pourra être de sous-échantillonner les logements dans cette grappe, de diviser cette grappe en sous-grappes ou de redéfinir cette grappe comme étant une strate et de créer de nouvelles grappes dans cette strate. Dans tous ces cas, un ajustement de sous-poids de grappe est calculé et appliqué aux poids de l'ESCC. Cet ajustement est multiplié par le poids **A0** pour produire le poids **A1**.

##### **A2 – Stabilisation**

Dans certaines régions socio-sanitaires (RSS), l'accroissement de la taille de l'échantillon résulte en un échantillon beaucoup plus grand que nécessaire. Une stabilisation a donc été instaurée afin de ramener la taille de l'échantillon au niveau désiré. Le processus de

stabilisation consiste à sous-échantillonner des logements aléatoirement à l'intérieur de la RSS parmi les logements originalement sélectionnés dans chaque grappe. Un facteur d'ajustement représentant l'effet de la stabilisation est donc calculé afin de corriger la probabilité de sélection. Ce facteur, multiplié par le poids **A1**, produit le poids **A2**.

### **A3 – Retrait des unités hors champ**

Parmi tous les logements échantillonnés, une certaine proportion de ceux-ci est, lors de la collecte, identifiée comme étant hors du champ de l'enquête. Des logements détruits ou en construction, des logements vacants, saisonniers ou secondaires, de même que des établissements, sont tous des exemples de cas hors champ pour l'ESCC. Ces logements sont tout simplement retirés de l'échantillon, ne laissant plus que les logements faisant partie du champ de l'enquête. Les logements qui restent dans l'échantillon conservent le même poids qu'à l'étape précédente que l'on appelle maintenant poids **A3**.

### **A4 – Non-réponse ménage**

Lors de la collecte, une certaine proportion des ménages échantillonnés a inévitablement résulté en non-réponse. Le poids des ménages non-répondants est redistribué aux répondants à l'intérieur de groupes homogènes de réponse (GHR). Dans le but de créer ces GHR, la méthode du score basée sur une régression logistique est utilisée afin de déterminer la probabilité de réponse puis, ces probabilités sont utilisées afin de diviser l'échantillon par groupes ayant des propriétés de réponse similaires. L'information disponible pour les non-répondants étant limitée, le modèle de régression comprend donc des variables comme la période de collecte et des informations géographiques ainsi que des para-données ce qui inclut le nombre d'essais pour contacter le ménage, l'heure/jour des essais et si les essais ont été faits au cours de la semaine ou de la fin de semaine. Un facteur d'ajustement a donc été calculé à l'intérieur de chaque GHR de la façon suivante:

$$\frac{\text{Somme des poids } A3 \text{ pour tous les ménages}}{\text{Somme des poids } A3 \text{ pour tous les ménages répondants}}$$

Le poids **A3** des ménages répondants est multiplié par ce facteur d'ajustement pour produire le poids **A4**. Les ménages non-répondants sont éliminés du processus de pondération à partir de ce point.

## **Pondération de l'échantillon provenant de la base téléphonique**

### **T0 – Poids initial**

Le poids initial (**T0**) pour les unités sur les bases téléphoniques est défini comme l'inverse de la probabilité de sélection et il est calculé quelque peu différemment selon que l'échantillon provienne de la base à composition aléatoire (CA) ou de la base liste.

### **T1 – Nombre de périodes de collecte**

Contrairement à la base aréolaire, pour laquelle l'échantillon est sélectionné entièrement au début du processus d'échantillonnage, des échantillons sont tirés à chaque deux mois pour les bases téléphoniques. À chacun de ces échantillons correspond un poids initial faisant en sorte que chaque échantillon soit représentatif de la RSS. Toutefois, pour que l'échantillon total ne représente qu'une seule fois la population, un facteur d'ajustement doit être appliqué pour réduire les poids de chaque échantillon. Le facteur d'ajustement appliqué à chaque échantillon est égal à l'inverse du nombre d'échantillons combinés ou encore, du nombre de périodes de collecte. À partir de ce moment, l'échantillon de la base liste correspond à la moyenne des échantillons des périodes de collecte combinées. Les poids initiaux sont multipliés par ce facteur d'ajustement afin de produire le poids **T1**.

## T2 - Retrait des unités hors champ

Les numéros de téléphone associés à des entreprises, des établissements ou à d'autres logements hors du champ de l'enquête, de même que les numéros hors service sont tous des exemples de cas hors champ pour la base téléphonique. Comme pour la base aréolaire, ces cas sont simplement retirés de l'échantillon, ne laissant ainsi dans l'échantillon que les logements dans le champ de l'enquête. Ces derniers conservent le même poids qu'à l'étape précédente que l'on appelle maintenant poids **T2**.

## T3 – Non-réponse ménage

L'ajustement fait ici pour compenser l'effet de la non-réponse ménage est identique à celui appliqué pour la base aréolaire (ajustement **A4**). Par contre, les para-données utilisées diffèrent puisque des applications de collecte différentes sont utilisées pour les interviews en personne et par téléphone. C'est donc cette variable qui a été utilisée pour définir les classes d'ajustement. Le facteur d'ajustement calculé à l'intérieur de chaque classe a été obtenu de la façon suivante:

$$\frac{\text{Somme des poids T2 pour tous les ménages}}{\text{Somme des poids T2 pour tous les ménages répondants}}$$

Le poids **T2** des ménages répondants a donc été multiplié par ce facteur d'ajustement pour produire le poids **T3**. Les ménages non-répondants sont éliminés à partir de ce point.

## T4 - Lignes multiples

Le fait que certains ménages possèdent plus d'une ligne téléphonique résidentielle a un impact sur la pondération: plus le ménage a de lignes, meilleure est sa probabilité d'être sélectionné. Conséquemment, les poids doivent être ajustés pour tenir compte du nombre de lignes résidentielles que le ménage possède. Le facteur d'ajustement représente l'inverse du nombre de lignes dans le ménage et le poids **T4** est obtenu en multipliant ce facteur par le poids **T3**.

## 9.2 Procédures de pondération pour l'Enquête sur les ménages et l'environnement

Les principes qui sous-tendent le calcul des poids pour l'EME sont identiques à ceux établis aux fins de l'ESCC. Toutefois, d'autres ajustements sont apportés au sous-poids ménage de l'ESCC afin de calculer un poids final pour les différents enregistrements inclus dans le fichier de microdonnées de l'EME.

### Diagramme B : Sommaire de la stratégie de pondération de l'EME

Étapes de pondération
<b>H0</b> – Le sous-poids de l'ESCC
<b>H1</b> – Le poids initial de l'EME
<b>H2</b> – Ajustement de la non-réponse de l'EME
<b>H3</b> – Ajustement pour le fichier partagé
<b>H4</b> – Calage aux marges

### H0 – Le sous-poids de l'ESCC

Une fois que l'étape **I1** de la stratégie de pondération de l'ESCC est achevée, le sous-poids de l'ESCC est obtenu. Cette étape **I1** consiste à intégrer les poids finaux des échantillons aréolaire et téléphonique créés jusqu'à maintenant, en un seul poids en appliquant une méthode d'intégration. Pour obtenir le facteur d'ajustement d'intégration, on calcule d'abord un facteur  $\alpha$ , obtenu de la façon suivante:

$$\alpha = n_A / (n_A + n_T)$$

où  $n_A$  et  $n_T$  représentent respectivement les tailles d'échantillon des bases aréolaire et téléphonique. Le poids des unités de la base aréolaire est multiplié par ce facteur  $\alpha$ , alors que le poids des unités de la base téléphonique est multiplié par  $1 - \alpha$ . Le produit du facteur d'ajustement dérivé ici, par le poids de ménage final calculé auparavant (**A4** ou **T4** dépendant de quelle base provient l'unité), procure le poids intégré **I1**, aussi appelé le sous-poids de l'ESCC

### **H1 – Le poids initial de l'EME**

L'échantillon de l'EME est un sous-échantillon aléatoire stratifié des répondants à l'ESCC. La probabilité d'être sélectionné dans l'échantillon de l'EME est d'abord calculé. Pour chaque ménage sélectionné pour l'EME, un facteur d'ajustement est défini comme l'inverse de la probabilité de sélection. Ce facteur, multiplié par le poids **H0**, produit le poids **H1**.

### **H2 – Ajustement de la non-réponse de l'EME**

Le poids des ménages non-répondants est redistribué aux répondants à l'intérieur de groupes homogènes de réponse (GHR). Dans le but de créer ces GHR, la méthode des scores basée sur une régression logistique est utilisée afin de déterminer la probabilité de réponse puis, ces probabilités sont utilisées afin de diviser l'échantillon par groupes ayant des propriétés de réponse similaires. Un facteur d'ajustement a donc été calculé à l'intérieur de chaque GHR de la façon suivante :

$$\frac{\text{Somme des poids } H1 \text{ pour tous les ménages}}{\text{Somme des poids } H1 \text{ pour tous les ménages répondants}}$$

Le poids **H1** des ménages répondants est multiplié par ce facteur d'ajustement pour produire le poids **H2**. Les ménages non-répondants sont éliminés du processus de pondération à partir de ce point.

### **H3 – Ajustement pour le fichier partagé**

Cet ajustement concerne seulement le fichier partagé. Les ménages ayant refusé de partager leur information doivent être exclus du fichier partagé. On doit donc procéder à un ajustement des poids de manière à compenser pour les ménages qui ont décliné la question portant sur le partage des données. Un facteur d'ajustement a donc été calculé à l'intérieur de chaque province de la façon suivante :

$$\frac{\text{Somme des poids } H2 \text{ pour tous les ménages \_répondants}}{\text{Somme des poids } H2 \text{ pour tous les ménages répondants \_ et \_ partageurs}}$$

Le poids **H2** des ménages répondants et partageurs est multiplié par ce facteur d'ajustement pour produire le poids **H3**. Les ménages ayant refusé de partager leur information sont supprimés du fichier de partage à partir de ce point.

### **H4 – Calage aux marges**

La dernière étape nécessaire afin d'obtenir le poids final de l'EME est le calage aux marges (**H4**). Le calage est appliqué afin de s'assurer que la somme des poids finaux corresponde aux projections du recensement définies à l'échelle des provinces et des taille du ménage (une personne, deux personnes ou trois personnes ou plus). Le poids **H4** obtenu à cette étape est le poids final WTHM figurant dans le fichier maître des microdonnées. Le poids final pour le fichier partagé est WTHS.

## 10.0 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.

### 10.1 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 des 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 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. Les proportions et rapports doivent d'abord être calculé à partir de composantes non arrondies puis être arrondis à leur tour à la troisième décimale à l'aide de la technique d'arrondissement normale.
- d) Les sommes et les différences d'agrégats 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. Les sommes et les différences de pourcentages (ou rapports) doivent être calculées à partir de leurs composantes non arrondies correspondantes, puis être arrondies à leur tour à la décimale près (ou troisième décimale) à l'aide de la technique d'arrondissement normale.
- e) Dans les cas, où, en raison de limitations d'ordre techniques 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 le sont en réalité.

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

Le plan d'échantillonnage utilisé pour l' EME 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 des poids de l'enquête par ces progiciels.

## **10.3 Lignes directrices pour l'analyse statistique**

L' EME 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.

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 ménages du Québec. Les étapes à suivre pour remettre à l'échelle les poids sont les suivantes :

1. sélectionner tous les ménages du fichier qui ont déclaré PROV = 24, Québec;
2. calculer le poids MOYEN pour ces enregistrements en additionnant les poids originaux des ménages établis à partir du fichier de microdonnées pour ces enregistrements puis diviser cette somme par le nombre de ménages ayant déclaré PROV = 24;
3. pour chacun de ces enregistrements, calculer un poids REMIS À L'ÉCHELLE égal au poids original du ménage divisé par le poids MOYEN;
4. effectuer l'analyse portant sur ces ménages en utilisant le poids REMIS À L'ÉCHELLE.

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

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

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

Avant de diffuser et/ou de publier toutes estimations établies à partir de l'EME2009, les utilisateurs devraient premièrement déterminer le niveau de qualité de cette estimation. Les niveaux de qualité sont *acceptables*, *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 (CV) 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'échantillons 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**

#### **Catégorie 1 – Acceptable**

Les estimations présentent de faibles coefficients de variation, de l'ordre de 0,0 à 16,5 %.  
Aucune restriction de diffusion : les données sont suffisamment exactes, si bien qu'un avertissement particulier aux utilisateurs ou d'autres restrictions ne sont pas nécessaires.

#### **Catégorie 2 – Médiocre**

Les estimations présentent des coefficients de variation élevés, de l'ordre de 16,6 à 33,3 %.  
Diffusion avec mise en garde : les données sont potentiellement utiles pour certaines fins, mais devraient être accompagnées d'un avertissement aux utilisateurs concernant leur exactitude.

Ces estimations devraient être signalées par la lettre E (ou un quelconque identificateur similaire).

#### **Catégorie 3 – Inacceptable**

Les estimations présentent des coefficients de variation très élevés, supérieurs à 33,3 %.  
Diffusion déconseillée : les données contiennent un niveau d'erreur élevé au point qu'elles ne

devraient pas être diffusées dans la plupart des circonstances afin d'éviter d'induire les utilisateurs en erreur. Si les utilisateurs insistent pour inclure les données de la catégorie 3 dans un produit non normalisé, même après avoir été informés de leur exactitude, les données devraient être accompagnées d'un avis de non-responsabilité. L'utilisateur devrait reconnaître les mises en garde reçues et s'engager à ne pas diffuser, présenter ni déclarer les données, directement ou indirectement, sans cet avis de non-responsabilité.

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 :

« 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. »

## **Annexe A – Estimation de la variance pour les fichiers maîtres et partagés**

Pour juger de la qualité d'une estimation et en calculer le CV, on doit d'abord établir l'écart-type. On a aussi besoin de l'écart-type d'estimations pour les intervalles de confiance. Dans l'EME, on utilise des plans d'échantillonnage et d'étalonnage à plusieurs degrés et, par conséquent, il n'y a pas de formule simple d'estimation de variance qu'on puisse appliquer. Il faut donc procéder par approximation en utilisant la méthode bootstrap. Avec l'aide des poids bootstrap et du Bootvar dont il sera question à la sous-section suivante, les CVs et autres mesures de variance peuvent être calculées avec précision.

### **Méthode bootstrap pour estimer la variance**

On prélève indépendamment sur chaque strate un échantillon aléatoire simple de  $(n - 1)$  des unités  $n$  de l'échantillon. À noter que, comme la sélection s'opère avec remise, une unité peut être choisie plusieurs fois. On reprend l'opération  $R$  fois pour ainsi obtenir  $R$  échantillons bootstrap. On calcule une valeur moyenne initiale de pondération bootstrap à partir des  $R$  échantillons pour chaque unité échantillonnée de la strate. Toute l'opération (prélèvement d'échantillons aléatoires simples et repondération de chaque strate) a lieu à  $B$  reprises –  $B$  étant une valeur élevée –, ce qui donne  $B$  valeurs initiales de pondération bootstrap.

On repondère alors comme dans la pondération courante par correction de non-réponse, étalonnage, etc. On obtient  $B$  valeurs de pondération bootstrap moyenne finale pour chaque unité de l'échantillon. On rapporte la variation des  $B$  estimations possibles correspondant aux  $B$  valeurs bootstrap à la variance de l'estimateur par pondération courante; la variance peut être estimée par ce moyen.

### **Progiciels statistiques pour estimer la variance**

#### **Bootvar**

Veillez prendre note que les poids bootstrap sont fournis et devraient être utilisés pour calculer la variance. BOOTVAR est un programme macro qui peut être utilisé pour calculer la variance en utilisant les poids bootstrap. Celui-ci est disponible en format SAS ou SPSS. Il est formé de macro instructions d'estimation de variance pour les totaux, les rapports et les différences entre rapports, ainsi que de régression linéaire et logistique.

Le Bootvar peut être téléchargé à partir du site Internet des Centres de données de recherche (CDR) de Statistique Canada. Il faut accepter la licence d'adhésion automatique Bootvar avant de pouvoir lire les fichiers.

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)

#### **Autres progiciels**

Il existe d'autres progiciels commerciaux pouvant être utilisés pour estimer la variance; Stata 9 ou 10, SUDAAN et WasVar.

Ces méthodes peuvent être adaptées pour l'EME d'après un article d'Owen Phillips intitulé *Comment utiliser les poids bootstrap avec Wes Var et SUDAAN* (n° 12-002-X20040027032 au catalogue) paru dans « *Le Bulletin technique et d'information des Centres de données de recherche (Index chronologique)* », automne 2004, vol.1 no 2, Statistique Canada, n° 12-002-XIF au catalogue.

## Annexe B – Estimation de la variance pour les fichiers de microdonnées à grande diffusion

### 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.

### 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 ménages 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 ménages possédant une caractéristique donnée

Le coefficient de variation d'une proportion estimée ou d'un pourcentage estimé dépend à la fois de la taille de la proportion ou du pourcentage et de la taille du total sur lequel la proportion ou le pourcentage repose. Les proportions estimées ou les pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque la proportion ou que le pourcentage repose sur un sous-groupe de la population. La proportion, par exemple, d'anciens fumeurs qui ont cessé pour des problèmes actuels de santé est plus fiable que le nombre estimé d'anciens fumeurs qui ont cessé pour des problèmes actuels de santé. (Remarquez que dans les tables la valeur des coefficients de variation diminue lorsqu'on les lit de gauche à droite.)

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

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

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

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

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

où  $\hat{X}_1$  est l'estimation 1,  $\hat{X}_2$  est l'estimation 2 et  $\alpha_1$  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 fumeurs et le numérateur, le nombre de fumeurs quotidiens.

Dans le cas où le numérateur n'est pas un sous-ensemble du dénominateur, comme dans l'exemple du rapport du nombre de fumeurs quotidiens comparativement au nombre de non-fumeurs, 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.

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

Les exemples ci-dessous utilisent des données du fichier annuel de l'Enquête de surveillance de l'usage du tabac au Canada (ESUTC) 2002 et sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter. Veuillez noter que les données utilisées dans ces exemples diffèrent des véritables résultats de l'enquête et tiennent seulement lieu de guide.

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

Supposons qu'un utilisateur estime que durant la période de référence, 5 414 335 de personnes étaient des fumeurs actuels (DVSST1 = 1) au Canada. 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 – Tous les âges du fichier des personnes.

Enquête de surveillance de l'usage du tabac au Canada, 2002 - février à décembre - Fichier des personnes														
Tables de variabilité d'échantillonnage approximative pour le Canada - Tous les âges														
NUMÉRATEUR DU POURCENTAGE ( '000)	POURCENTAGE ESTIMÉ													
	0,1%	1,0%	2,0%	5,0%	10,0%	15,0%	20,0%	25,0%	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%
1	197,2	196,3	195,3	192,3	187,1	181,9	176,4	170,8	165,0	159,0	152,8	139,5	108,0	62,4
2	139,4	138,8	138,1	135,9	132,3	128,6	124,8	120,8	116,7	112,5	108,0	98,6	76,4	44,1
3	113,8	113,3	112,7	111,0	108,0	105,0	101,9	98,6	95,3	91,8	88,2	80,5	62,4	36,0
4	98,6	98,1	97,6	96,1	93,6	90,9	88,2	85,4	82,5	79,5	76,4	69,7	54,0	31,2
5	88,2	87,8	87,3	86,0	83,7	81,3	78,9	76,4	73,8	71,1	68,3	62,4	48,3	27,9
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
75	*****	22,7	22,5	22,2	21,6	21,0	20,4	19,7	19,1	18,4	17,6	16,1	12,5	7,2
80	*****	21,9	21,8	21,5	20,9	20,3	19,7	19,1	18,5	17,8	17,1	15,6	12,1	7,0
85	*****	21,3	21,2	20,9	20,3	19,7	19,1	18,5	17,9	17,2	16,6	15,1	11,7	6,8
90	*****	20,7	20,6	20,3	19,7	19,2	18,6	18,0	17,4	16,8	16,1	14,7	11,4	6,6
95	*****	20,1	20,0	19,7	19,2	18,7	18,1	17,5	16,9	16,3	15,7	14,3	11,1	6,4
100	*****	19,6	19,5	19,2	18,7	18,2	17,6	17,1	16,5	15,9	15,3	13,9	10,8	6,2
125	*****	17,6	17,5	17,2	16,7	16,3	15,8	15,3	14,8	14,2	13,7	12,5	9,7	5,6
150	*****	16,0	15,9	15,7	15,3	14,8	14,4	13,9	13,5	13,0	12,5	11,4	8,8	5,1
200	*****	13,9	13,8	13,6	13,2	12,9	12,5	12,1	11,7	11,2	10,8	9,9	7,6	4,4
250	*****	12,4	12,4	12,2	11,8	11,5	11,2	10,8	10,4	10,1	9,7	8,8	6,8	3,9
300	*****	*****	11,3	11,1	10,8	10,5	10,2	9,9	9,5	9,2	8,8	8,1	6,2	3,6
350	*****	*****	10,4	10,3	10,0	9,7	9,4	9,1	8,8	8,5	8,2	7,5	5,8	3,3
400	*****	*****	9,8	9,6	9,4	9,1	8,8	8,5	8,3	8,0	7,6	7,0	5,4	3,1
450	*****	*****	9,2	9,1	8,8	8,6	8,3	8,1	7,8	7,5	7,2	6,6	5,1	2,9
500	*****	*****	8,7	8,6	8,4	8,1	7,9	7,6	7,4	7,1	6,8	6,2	4,8	2,8
750	*****	*****	*****	7,0	6,8	6,6	6,4	6,2	6,0	5,8	5,6	5,1	3,9	2,3
1000	*****	*****	*****	6,1	5,9	5,8	5,6	5,4	5,2	5,0	4,8	4,4	3,4	2,0
1500	*****	*****	*****	*****	4,8	4,7	4,6	4,4	4,3	4,1	3,9	3,6	2,8	1,6
2000	*****	*****	*****	*****	4,2	4,1	3,9	3,8	3,7	3,6	3,4	3,1	2,4	1,4
3000	*****	*****	*****	*****	*****	3,3	3,2	3,1	3,0	2,9	2,8	2,5	2,0	1,1
4000	*****	*****	*****	*****	*****	*****	2,8	2,7	2,6	2,5	2,4	2,2	1,7	1,0
5000	*****	*****	*****	*****	*****	*****	2,5	2,4	2,3	2,2	2,2	2,0	1,5	0,9
6000	*****	*****	*****	*****	*****	*****	*****	2,2	2,1	2,1	2,0	1,8	1,4	0,8
7000	*****	*****	*****	*****	*****	*****	*****	*****	2,0	1,9	1,8	1,7	1,3	0,7
8000	*****	*****	*****	*****	*****	*****	*****	*****	*****	1,8	1,7	1,6	1,2	0,7
9000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1,6	1,5	1,1	0,7
10000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1,5	1,4	1,1	0,6
12500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1,2	1,0	0,6
15000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	0,9	0,5

NOTA : Pour utiliser ces tables correctement, veuillez vous référer à la documentation reliée aux microdonnées.

- 2) L'agrégat estimé (5 414 335) 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 5 000 000.
- 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 2,5 %.
- 4) Le coefficient de variation approximatif de l'estimation est donc 2,5 %. Le résultat selon lequel il y avait 5 414 335 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 8.1) de fumeurs actuels durant la période de référence, peut être publié sans réserve.

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

Supposons qu'un utilisateur estime à  $2\,865\,929 / 12\,436\,728 = 23,0\%$  la proportion d'hommes qui fument actuellement au Canada durant la période de référence. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Reportez-vous à la table des coefficients de variation pour le CANADA du fichier des personnes (voir ci-dessus). Le tableau au niveau du CANADA devrait être utilisé parce qu'il est le plus petit tableau qui contient le domaine de l'estimation, tous les hommes au Canada.
- 2) Parce que l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c'est-à-dire les hommes), il faut utiliser à la fois le pourcentage (23,0 %) et la portion numérateur du pourcentage (2 865 929) pour déterminer le coefficient de variation.
- 3) Le numérateur, 2 865 929, 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 3 000 000. 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 25,0 %.
- 4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 3,1 %, est le coefficient de variation à employer.
- 5) Le coefficient de variation approximatif de l'estimation est donc 3,1 %. Le résultat selon lequel 23,0 % des hommes fument actuellement, peut être publié sans réserve.

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

Supposons qu'un utilisateur estime à  $2\,548\,406 / 12\,814\,359 = 19,9\%$  la proportion de femmes qui fument actuellement au Canada et à  $2\,865\,929 / 12\,436\,728 = 23,0\%$  la proportion d'hommes qui fument actuellement au Canada. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

- 1) L'utilisation de la table des coefficients de variation pour le CANADA du fichier des personnes (voir ci-dessus) de la même façon que celle décrite dans l'exemple 2, donne un CV de l'estimation pour les femmes de 3,2 % et un CV de l'estimation pour les hommes de 3,1 %.

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

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

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

C'est-à-dire que l'erreur-type de la différence  $\hat{d} = 0,230 - 0,199 = 0,031$  est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0,230)(0,031)]^2 + [(0,199)(0,032)]^2} \\ &= \sqrt{(0,00005) + (0,00004)} \\ &= 0,009\end{aligned}$$

- 3) Le coefficient de variation de  $\hat{d}$  est donné par  $\sigma_{\hat{d}} / \hat{d} = 0,009 / 0,031 = 0,290$ .
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 29,0 %. La différence entre les estimations est considérée médiocre 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 E (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les utilisateurs subséquents en garde contre les hauts taux d'erreur associés à l'estimation.

#### Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime à 237 261 le nombre de femmes âgées de 15 à 19 ans qui fument actuellement et à 220 511 le nombre d'hommes âgés de 15 à 19 ans qui fument actuellement. L'utilisateur est intéressé à comparer l'estimation des femmes à celle des hommes sous la forme d'un rapport. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation d'un rapport, où le numérateur de l'estimation ( $\hat{X}_1$ ) est le nombre de femmes âgées de 15 à 19 ans qui fument actuellement. Le dénominateur de l'estimation ( $\hat{X}_2$ ) est le nombre d'hommes âgés de 15 à 19 ans qui fument actuellement.
- 2) Reportez-vous à la table des coefficients de variation pour le CANADA – 15 à 19 ans du fichier des personnes.

**Enquête de surveillance de l'usage du tabac au Canada, 2002 - février à décembre - Fichier des personnes**

**Tables de variabilité d'échantillonnage approximative pour le Canada - 15 à 19 ans**

NUMÉRATEUR DU POURCENTAGE ( '000)	POURCENTAGE ESTIMÉ													
	0,1%	1,0%	2,0%	5,0%	10,0%	15,0%	20,0%	25,0%	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%
1	95,8	95,3	94,9	93,4	90,9	88,3	85,7	83,0	80,2	77,3	74,2	67,8	52,5	30,3
2	67,7	67,4	67,1	66,0	64,3	62,5	60,6	58,7	56,7	54,6	52,5	47,9	37,1	21,4
3	*****	55,0	54,8	53,9	52,5	51,0	49,5	47,9	46,3	44,6	42,9	39,1	30,3	17,5
4	*****	47,7	47,4	46,7	45,5	44,2	42,9	41,5	40,1	38,6	37,1	33,9	26,2	15,2
5	*****	42,6	42,4	41,8	40,7	39,5	38,3	37,1	35,9	34,6	33,2	30,3	23,5	13,6
6	*****	38,9	38,7	38,1	37,1	36,1	35,0	33,9	32,7	31,5	30,3	27,7	21,4	12,4
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
95	*****	*****	*****	9,6	9,3	9,1	8,8	8,5	8,2	7,9	7,6	7,0	5,4	3,1
100	*****	*****	*****	9,3	9,1	8,8	8,6	8,3	8,0	7,7	7,4	6,8	5,2	3,0
125	*****	*****	*****	*****	8,1	7,9	7,7	7,4	7,2	6,9	6,6	6,1	4,7	2,7
150	*****	*****	*****	*****	7,4	7,2	7,0	6,8	6,5	6,3	6,1	5,5	4,3	2,5
200	*****	*****	*****	*****	6,4	6,2	6,1	5,9	5,7	5,5	5,2	4,8	3,7	2,1
250	*****	*****	*****	*****	*****	5,6	5,4	5,2	5,1	4,9	4,7	4,3	3,3	1,9
300	*****	*****	*****	*****	*****	5,1	4,9	4,8	4,6	4,5	4,3	3,9	3,0	1,7
350	*****	*****	*****	*****	*****	*****	4,6	4,4	4,3	4,1	4,0	3,6	2,8	1,6
400	*****	*****	*****	*****	*****	*****	4,3	4,1	4,0	3,9	3,7	3,4	2,6	1,5
450	*****	*****	*****	*****	*****	*****	*****	3,9	3,8	3,6	3,5	3,2	2,5	1,4
500	*****	*****	*****	*****	*****	*****	*****	3,7	3,6	3,5	3,3	3,0	2,3	1,4
750	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,7	2,5	1,9	1,1
1000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,1	1,7	1,0
1500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	0,8

**NOTA : Pour utiliser ces tables correctement, veuillez vous référer à la documentation reliée aux microdonnées.**

- 3) Le numérateur de cette estimation de rapport est 237 261. Le chiffre qui s'en rapproche le plus est 250 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 5,6 %.
- 4) Le dénominateur de cette estimation de rapport est 220 511. Le chiffre qui s'en rapproche le plus est 200 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 6,4 %.
- 5) Le coefficient de variation approximatif de l'estimation du rapport est donc donné par la règle 4, qui est :

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

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

C'est-à-dire que :

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0,056)^2 + (0,064)^2} \\ &= \sqrt{0,003136 + 0,004096} \\ &= 0,085\end{aligned}$$

- 6) Le rapport obtenu entre les femmes et les hommes âgés de 15 à 19 ans qui fument actuellement est 237 261 / 220 511, c'est-à-dire 1,08 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 8.1). Le coefficient de variation de cette estimation est 8,5 %, ce qui fait qu'on peut la diffuser sans réserve.

### 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.

**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 des hommes qui fument actuellement (d'après l'exemple 2) serait calculé comme suit :

$$\hat{X} = 23,0 \% \text{ (ou exprimé sous forme de proportion } 0,230)$$

$$t = 2$$

$\alpha_{\hat{x}}$  = 3,1 % (0,031 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.

$$IC_{\hat{x}} = \{0,230 - (2) (0,230) (0,031), 0,230 + (2) (0,230) (0,031)\}$$

$$IC_{\hat{x}} = \{0,230 - 0,014, 0,230 + 0,014\}$$

$$IC_{\hat{x}} = \{0,216, 0,244\}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 21,6 % et 24,4 % des hommes fument actuellement.

**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.

**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 d'hommes qui fument actuellement et la proportion de femmes qui fument actuellement. D'après l'exemple 3, il s'est avéré que l'erreur-type de la différence entre ces deux estimations était 0,009. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_d} = \frac{0,230 - 0,199}{0,009} = \frac{0,031}{0,009} = 3,44$$

Puisque  $t = 3,44$  est plus grand que 2, il faut en conclure qu'il existe une différence significative entre les deux estimations au niveau de signification de 0,05.