GUIDE DE L'UTILISATEUR DES MICRODONNÉES DE L'ENQUÊTE SUR L'UTILISATION DE L'ÉNERGIE PAR LES MÉNAGES

FÉVRIER 1993

TABLE DES MATIÈRES

1. INTRODUCTION

2. CONTEXTE

3. OBJECTIFS

4. CONCEPTS ET DÉFINITIONS

5. MÉTHODOLOGIE DE L'ENQUÊTE

    5.1 Champ de l'enquête
    5.2 Plan d'échantillonnage

        5.2.1 Stratification primaire
        5.2.2 Types de secteurs
        5.2.3 Stratification secondaire
        5.2.4 Délimitation et sélection des grappes
        5.2.5 Sélection des logements
        5.2.6 Sélection des personnes

    5.3 Taille de l'échantillon
    5.4 Renouvellement de l'échantillon
    5.5 Modifications apportées au plan de l'EPA pour l'enquête supplémentaire
    5.6 Taille de l'échantillon par province pour l'enquête supplémentaire

6. COLLECTE DES DONNÉES

    6.1 L'interview de l'EPA
    6.2 Surveillance et contrôle
    6.3 Modification de la collecte des données pour l'EUEM
    6.4 Non-réponse à l'EPA
    6.5 Non-réponse à l'Enquête sur l'utilisation de l'énergie par les ménages

7. TRAITEMENT DES DONNÉES

    7.1 Saisie des données
    7.2 Vérification
    7.3 Création de variables calculées
    7.4 Pondération
    7.5 Suppression des renseignements confidentiels

8. QUALITÉ DES DONNÉES

    8.1 Taux de réponse
    8.2 Erreurs relatives à l'enquête

9. LIGNES DIRECTRICES POUR LA PUBLICATION ET LA DIFFUSION

9.1 Organisation du fichier de microdonnées
9.2 Lignes directrices pour l'arrondissement
9.3 Lignes directrices pour la pondération de l'échantillon en vue de la totalisation

9.3.1 Définitions des genres d'estimations de type nominal ou quantitatives
9.3.2 Totalisation d'estimations de type nominal
9.3.3 Totalisation d'estimations quantitatives

9.4 Lignes directrices pour l'analyse statistique
9.5 Lignes directrices pour la diffusion en fonction de la variabilité d'échantillonnage

10. TABLES DE VARIABILITÉ D'ÉCHANTILLONNAGE APPROXIMATIVE

10.1 Comment utiliser les tables de c.v. pour les estimations de type nominal
10.2 Exemples d'utilisation des tables de c.v. pour des estimations de type nominal
10.3 Comment utiliser les tables de c.v. pour calculer des limites de confiance
10.4 Exemples d'utilisation de tables de c.v. pour obtenir des limites de confiance
10.5 Comment utiliser les tables de c.v. pour effectuer un test t
10.6 Exemple d'utilisation des tables de c.v. pour effectuer un test t
10.7 Coefficients de variation d'estimations quantitatives
10.8 Seuils pour la diffusion des estimations relatives à l'EUEM
10.9 Tables de c.v.

11. PONDÉRATION

11.1 Pondération de l'EPA
11.2 Pondération pour l'Enquête sur l'utilisation de l'énergie par les ménages

12. CLICHÉ D'ARTICLE ET VARIABLES UNIQUES

13. QUESTIONNAIRES ET FEUILLES DES CODES

1. INTRODUCTION

L'Enquête de 1993 sur l'utilisation de l'énergie par les ménages (EUEM) a été effectuée par Statistique Canada pour le compte de Ressources naturelles Canada et avec le concours des provinces de la Nouvelle-Écosse, du Nouveau-Brunswick, de l'Ontario, du Manitoba et de la Saskatchewan, ainsi que SaskPower. Ce manuel a pour but de faciliter l'utilisation des résultats de l'enquête réunis dans un fichier de microdonnées. Il est important de bien connaître le manuel avant de produire des estimations provenant de ce fichier.

Prière de vous adresser aux personnes suivantes pour toutes les questions concernant les ensembles de données :

Elizabeth Majewski
Groupe des enquêtes spéciales
Statistique Canada
5-A6, Immeuble Jean-Talon
Parc Tunney
Ottawa, Ontario
K1A 0T6
Téléphone : (613) 951-4584
Télécopieur : (613) 951-0562

Yves Bélanger
Groupe des enquêtes spéciales
Statistique Canada
5-D6, Immeuble Jean-Talon
Parc Tunney
Ottawa, Ontario
K1A 0T6
Téléphone : (613) 951-9477
Télécopieur : (613) 951-0562

André Bourbeau
Direction de l'efficacité énergétique et des énergies
de remplacement
Ressources naturelles Canada
15e étage, 580 rue Booth
Ottawa (Ontario)
K1A 0E4
Téléphone : (613) 995-7307
Télécopieur : (613) 947-4120

2. CONTEXTE

Jusqu'à ce que l'Enquête sur l'utilisation de l'énergie par les ménages soit entreprise, l'Enquête sur l'équipement ménager réalisée chaque année par Statistique Canada constituait la seule source de données nationales sur le stock d'appareils qui consomment de l'énergie. Cependant, celle-ci n'est pas assez détaillée et n'englobe pas les caractéristiques de la maison dont Ressources naturelles Canada a besoin pour alimenter la Base nationale de données sur l'utilisation de l'énergie qui a été créée récemment. Statistique Canada a été chargée de mettre sur pied l'EUEM afin d'enrichir le volet de cette base de données traitant du secteur résidentiel.

L'enquête a été réalisée à titre de supplément à l'Enquête sur la population active de février 1993. La population visée englobait toutes les unités d'habitation au Canada (à l'exception du Yukon et des Territoires du Nord-Ouest) qui étaient occupées à titre de résidence principale, soit par leur propriétaire ou un locataire.

3. OBJECTIFS

L'Enquête sur l'utilisation de l'énergie par les ménages avait pour but de fournir des renseignements détaillés sur :

- les caractéristiques de la maison qui influent sur son étanchéité (enveloppe thermique),
- l'équipement de chauffage des locaux et les combustibles utilisés,
- l'équipement de climatisation,
- l'équipement de chauffage de l'eau,
- le stock et les caractéristiques des gros appareils électroménagers consommateurs d'énergie,
- l'utilisation des appareils, le chauffage et la climatisation.

En outre, les résultats de l'enquête seront complétés par deux ensembles de données. Le premier consistera en un fichier renfermant la consommation moyenne d'énergie des appareils électroménagers utilisés par les ménages de l'échantillon qui ont donné des renseignements sur la marque et le modèle. Le deuxième ensemble consistera en un fichier renfermant la consommation annuelle en électricité, en gaz naturel et en mazout d'un sous-ensemble des ménages enquêtés qui payent la facture de leurs services publics, qui ont autorisé Statistique Canada à obtenir cette information de leurs fournisseurs et pour lesquels on a réussi à coupler les données avec les fichiers des services publics.

4. CONCEPTS ET DÉFINITIONS

Logement

Ensemble distinct de pièces d'habitation ayant une entrée privée donnant sur l'extérieur ou sur un corridor ou escalier commun à l'intérieur.

Genre de logement

Maison individuelle - une maison ne comprenant qu'un logement et non rattachée à un autre immeuble ou à une autre construction.

Maison jumelée - un de deux logements réunis côte à côte.

Maison en rangée - logement dans une rangée d'au moins trois logements réunis côte à côte, sans aucun autre logement au-dessus ou en dessous.

Duplex - deux logements superposés qui ne sont joints à aucun autre logement ou construction.

Appartement dans un immeuble de moins de cinq étages - logement dans un triplex ou quadruplex ou dans un autre immeuble de moins de cinq étages.

Appartement dans un immeuble de cinq étages ou plus.

Maison mobile - logement individuel conçu et construit pour être transporté sur son propre châssis.

Ménage

Personne ou groupe de personnes occupant un logement. Le nombre de ménages est égal au nombre de logements occupés.

Situation familiale

Une famille est «un groupe de deux personnes ou plus qui vivent ensemble dans le même logement et qui sont apparentées par le sang, le mariage ou l'adoption». Une personne qui vit seule ou n'est apparentée à aucun occupant du logement où elle demeure est classée comme «personne seule». On établit qui est le «chef de famille» en demandant à l'enquêté qui est, selon lui, la personne qui joue ce rôle. Ce concept n'a aucune connotation économique; il ne s'agit que d'un procédé statistique.

Région urbaine - région rurale

Statistique Canada considère comme une région urbaine toute région dont la population atteint au moins 1,000 habitants et dont la densité démographique est d'au moins 400 habitants au kilomètre carré, d'après le dernier recensement. Tout territoire situé à l'extérieur des régions urbaines est considéré comme région rurale.

Situation vis-à-vis de l'activité

Situation du répondant sur le marché du travail; un membre de la population civile hors institution active âgé de 15 ans et plus est soit **occupé**, soit en **chômage**, soit **inactif**.

Personnes occupées

Les personnes occupées sont celles qui, au cours de la semaine de référence :

a) ont fait un travail[1]

b) avaient un emploi mais n'étaient pas au travail pour l'une des raisons suivantes :
   - maladie ou invalidité (de l'enquêté)
   - obligations personnelles ou familiales
   - mauvais temps
   - conflit de travail
   - vacances
   - autre raison non précisée ci-dessus (à l'exception des personnes mises à pied et de celles qui devaient commencer à travailler à une date ultérieure déterminée).

[1] On entend par «travail» tout travail fait contre rémunération ou en vue d'un bénéfice, c'est-à-dire le travail fait contre rémunération pour un employeur ou à son propre compte. Il peut également s'agir d'un travail familial non rémunéré, c'est-à-dire d'un travail non rémunéré qui contribue directement à l'exploitation d'une ferme, d'une entreprise ou d'un bureau professionnel, que possède ou exploite un membre apparenté du ménage. Ce travail peut consister par exemple à tenir la comptabilité, à vendre des produits et à faire le service aux tables. Le travail ménager et l'entretien d'une maison ne constituent pas un travail familial non rémunéré.

Chômeurs

Les chômeurs sont les personnes qui, au cours de la semaine de référence :

a) étaient sans emploi, avaient activement cherché du travail au cours des quatre dernières semaines (y compris la semaine de référence) et étaient prêtes à travailler[2];

b) n'avaient pas activement cherché de travail au cours des quatre dernières semaines, mais avaient été mises à pied[3] et étaient prêtes à travailler;

c) n'avaient pas activement cherché de travail au cours des quatre dernières semaines, mais devaient commencer un nouvel emploi dans quatre semaines ou moins à compter de la semaine de référence, et étaient prêtes à travailler.

Inactifs

Il s'agit de la partie de la population civile hors institutions âgée de 15 ans et plus qui était ni occupée ni en chômage pendant la semaine de référence.

---

[2] On considère comme prêtes à travailler les personnes de ce groupe qui:

(i) étudiaient à plein temps tout en cherchant du travail à temps partiel et qui répondaient au critère (ii) ci-dessous. (Les étudiants à plein temps qui cherchaient du travail à plein temps ne sont pas considérés comme ayant été prêts à travailler pendant la semaine de référence.)

(ii) ont déclaré qu'il n'y avait aucune raison pour laquelle elles ne pouvaient prendre un emploi pendant la semaine de référence ou ne pouvaient en prendre un en raison de «maladie ou invalidité», «d'obligations personnelles ou familiales» ou «parce qu'elles avaient déjà un emploi».

[3] On considère comme mises à pied les personnes qui s'attendent à réintégrer l'emploi qu'elles occupaient avant d'être mises à pied.

5. MÉTHODOLOGIE DE L'ENQUÊTE

Comme l'Enquête sur l'utilisation de l'énergie par les ménages a été menée en février 1993 auprès d'un sous-échantillon des logements de l'Enquête sur la population active (EPA), son plan d'échantillonnage est étroitement lié à celui de l'EPA. Le plan d'échantillonnage de l'EPA est décrit brièvement aux sections 5.1 à 5.4[4]. La section 5.5 décrit les différences entre le plan de l'Enquête sur l'utilisation de l'énergie et le plan de base de l'EPA en février 1993.

5.1 Champ de l'enquête

L'EPA est une enquête-ménages mensuelle dont l'échantillon est représentatif de la population civile hors institutions des dix provinces du Canada qui est âgée de 15 ans ou plus. Sont spécifiquement exclus du champ de l'enquête les résidents du Yukon et des Territoires du Nord-Ouest, les personnes vivant sur les réserves indiennes, les membres à temps plein des Forces armées canadiennes et les pensionnaires d'institutions. Ces groupes représentent en tout environ 2% de la population âgée de 15 ans ou plus.

5.2 Plan d'échantillonnage

L'échantillon de l'EPA est basé sur un plan à plusieurs degrés stratifié qui fait appel à un échantillonnage probabiliste à toutes les étapes. Les principes du plan sont les mêmes pour toutes les provinces. On trouvera, dans le présent document, un schéma récapitulant les degrés du plan.

5.2.1 Stratification primaire

Les provinces sont d'abord stratifiées en régions économiques, qui sont des régions géographiques de structure économique plus ou moins homogène constituées en vertu d'ententes fédérale-provinciales et qui sont relativement stables dans le temps.

Ces régions économiques sont considérées comme des strates primaires et on procède à une stratification plus poussée à l'intérieur de chacune d'entre elles (voir la sous-section 5.2.3).

---

[4] On trouvera une description détaillée du plan d'échantillonnage de l'EPA dans la publication de Statistique Canada intitulée **Méthodologie de l'enquête sur la population active du Canada**, 1984-1990 (n° 71-526 au catalogue).

5.2.2 Types de secteurs

Les régions économiques sont subdivisées en 3 catégories : les unités autoreprésentatives (UAR), les unités non autoreprésentatives (UNAR) et les secteurs spéciaux. Les UAR sont généralement des régions urbaines dont la population (selon le recensement de 1981) excède 15,000 personnes ou dont les caractéristiques uniques relatives à la population active exigent qu'elles soient constituées en UAR. Dans la plupart des cas, les limites des UAR correspondent aux délimitations établies pour le recensement.

Toutes les UAR de chaque région économique font partie de l'enquête et, comme leur nom l'indique, chacune est représentée par son propre échantillon.

Les UNAR sont les régions à l'extérieur des UAR. Elles sont principalement constituées de petits centres urbains et de régions rurales. Chaque région économique renferme une UNAR qui est représentée par son propre échantillon.

Une petite proportion (environ 1%) de la population de l'EPA demeure dans des institutions (par exemple, le personnel résident des hôpitaux ou des écoles ou les résidents permanents des hôtels ou des motels), dans des bases militaires (le personnel civil seulement), ou dans des régions éloignées des provinces, que les intervieweurs de l'EPA ne peuvent atteindre facilement. À des fins administratives, ce segment de la population est échantillonné séparément à l'aide de la base de sondage des secteurs spéciaux. Ce segment de l'échantillon est sélectionné à l'échelle provinciale indépendamment de la stratification utilisée pour les UAR et les UNAR.

5.2.3 Stratification secondaire

Les UAR sont ensuite délimitées individuellement en strates du plan, qui reflètent des régions de statut socio-économique semblable selon le recensement de 1981. Le degré de stratification (c'est-à-dire le nombre de strates) dépend de la taille de l'UAR.

Dans les régions économiques où la population des UNAR constitue une proportion importante de la population de la région économique, l'UNAR est ensuite délimitée en strates urbaines et rurales distinctes. À l'intérieur de chacune de ces strates, on procède à une nouvelle stratification pour refléter les différences en ce qui a trait à un certain nombre de caractéristiques de la population active.

Dans les secteurs spéciaux, les strates sont constituées à l'échelle provinciale. Les strates reflètent les principaux types de groupes spéciaux de la population pour lesquels il faut utiliser des procédures administratives spéciales d'échantillonnage. Il s'agit des établissements militaires, des institutions et des régions éloignées.

5.2.4 Délimitation et sélection des grappes

Dans chaque strate secondaire se trouvant dans les UAR, un certain nombre de groupes de logements géographiquement contigus, ou grappes, sont constitués à partir d'une combinaison de comptes du recensement de 1981 et d'un dénombrement sur le terrain. Ces grappes coïncident généralement avec des îlots urbains ou des côtés d'îlot. La sélection d'un échantillon de grappes (en général, 6 ou 12 grappes) à partir de chacune de ces strates secondaires représente le premier degré de l'échantillonnage dans les UAR.

Dans chacune des strates secondaires des UNAR, un certain nombre de grandes régions géographiques sont délimitées de façon à ce que chacune reflète, en fonction d'un certain nombre de caractéristiques socio-économiques, la composition de la strate où elle est située. Deux ou quatre de ces régions, appelées unités primaires d'échantillonnage (UPÉ) sont sélectionnés dans l'échantillon à partir de chaque strate secondaire. Dans chaque UPÉ sélectionnée, un certain nombre de plus petits groupes de logements géographiquement contigus, ou grappes, sont ensuite constitués d'après des caractéristiques physiques bien définies identifiables tant sur les cartes que sur le terrain.

Dans les secteurs spéciaux, les secteurs de dénombrement du recensement (régions géographiques parcourues par un seul recenseur) représentent le premier degré de sélection. Dans les secteurs de dénombrement sélectionnés, des groupes de logements géographiquement contigus, ou grappes, sont constitués au besoin et la sélection d'un échantillon parmi ces grappes représente le deuxième degré de l'échantillonnage.

5.2.5 Sélection des logements

Dans les trois types de secteurs (UAR, UNAR et secteurs spéciaux), les intervieweurs commencent par visiter les grappes sélectionnées, puis on dresse une liste de tous les logements privés dans chaque grappe. Un échantillon de 6 logements (en moyenne) est ensuite sélectionné à partir de cette liste, ce qui représente le dernier degré de l'échantillonnage.

Dans les 17 UAR les plus importantes, on sélectionne un échantillon d'appartements dans les grands immeubles d'appartements à partir d'un registre distinct fondé sur les renseignements fournis par la SCHL. Cette opération a pour but d'assurer une meilleure représentation des locataires d'appartements dans l'échantillon et de minimiser l'effet de croissance dans les grappes en raison de la construction de nouveaux immeubles d'appartements.

5.2.6 Sélection des personnes

On obtient des données démographiques pour toutes les personnes ayant pour domicile habituel le logement sélectionné. Les données de l'EPA sont recueillies pour tous les membres civils du ménage âgés de 15 ans ou plus.

5.3 Taille de l'échantillon

La taille de l'échantillon de personnes admissibles à l'EPA est calculée pour que l'on atteigne la précision statistique désirée en ce qui a trait à diverses caractéristiques de la population active aux niveaux provincial et infraprovincial, afin de répondre aux besoins des administrations fédérale, provinciales et municipales ainsi que d'une foule d'autres utilisateurs des données.

L'échantillon mensuel de l'EPA comprend approximativement 73,000 logements. Après avoir exclu les logements vacants, démolis ou convertis à des usages non résidentiels et ceux où ne vivent que des personnes non admissibles ainsi que les logements en construction et les logements saisonniers, il reste environ 63,000 logements occupés par une ou plusieurs personnes admissibles. Dans ces logements, l'EPA permet de recueillir des renseignements sur environ 122,000 civils âgés de 15 ans ou plus.

5.4 Renouvellement de l'échantillon

L'EPA emploie un plan par panel où l'ensemble de l'échantillon mensuel des logements est constitué de 6 panels, ou groupes de renouvellement, à peu près de la même taille, que l'on considère représentatifs de toute la population de l'EPA. Tous les logements d'un groupe de renouvellement demeurent dans l'échantillon de l'EPA pendant 6 mois consécutifs. Ils sont ensuite remplacés (supprimés de l'échantillon par renouvellement) par un nouveau panel de logements sélectionné à partir des mêmes grappes ou d'autres grappes semblables.

Ce procédé de renouvellement a été adopté afin de s'assurer que l'échantillon des logements reflète constamment les changements au parc immobilier actuel et de minimiser les problèmes de non-réponse ou de fardeau du répondant qui se produiraient si les ménages devaient demeurer plus de 6 mois dans l'échantillon. Il a également pour avantage, sur le plan statistique, de fournir une base de sondage commune en vue de réaliser des comparaisons à court terme, d'un mois à l'autre, des caractéristiques de l'EPA.

Grâce à la méthode des groupes de renouvellement, on peut facilement réaliser des enquêtes supplémentaires au moyen du plan d'échantillonnage de l'EPA mais en n'utilisant pas l'échantillon complet de l'enquête.

5.5 Modifications apportées au plan d'échantillonnage de l'EPA pour l'enquête supplémentaire

L'Enquête sur l'utilisation de l'énergie par les ménages a utilisé un sous-ensemble des six groupes de renouvellement faisant partie de l'échantillon de l'EPA en février 1993. Selon la province, le questionnaire d'enquête a été envoyé à un nombre variable de groupes de renouvellement, allant de la moitié à deux groupes complets. En outre, dans cinq provinces, on a également utilisé un ou deux groupes de renouvellement qui avaient déjà participé à l'EPA (la dernière fois en décembre 1992 ou en janvier 1993) et qui avaient été remplacés. Le tableau suivant présente sommairement le nombre de groupes de renouvellement utilisés dans chaque province.

PROVINCE	Nombre de groupes de renouvellement de l'EPA de février 1993	Nombre de groupes de renouvellement d'une EPA antérieure	Nombre total de groupes de renouvellement
Terre-Neuve	2	0	2
Île-du-Prince-Édouard	2	2	4
Nouvelle-Écosse	2	1	3
Nouveau-Brunswick	2	1	3
Québec	1/2	0	1/2
Ontario	7/10	0	7/10
Manitoba	2	1	3
Saskatchewan	2	1	3
Alberta	1	0	1
Colombie-Britannique	1	0	1

Dans le cas de l'EUEM, la couverture de l'EPA a été modifiée afin d'exclure tous les ménages collectifs, par exemple les institutions et les hôtels.

5.6 Taille de l'échantillon par province pour l'enquête supplémentaire

Le tableau suivant présente le nombre de ménages admissibles à l'EUEM. Ces ménages faisaient partie de la population cible et des groupes de renouvellement compris dans l'échantillon, et ils avaient répondu au questionnaire de l'EPA.

PROVINCE	TAILLE DE L'ÉCHANTILLON
Terre-Neuve	1,013
Île-du-Prince-Édouard	947
Nouvelle-Écosse	2,165
Nouveau-Brunswick	1,878
Québec	914
Ontario	1,987
Manitoba	2,065
Saskatchewan	2,389
Alberta	913
Colombie-Britannique	911
CANADA	**15,182**

Les écarts entre les échantillons provinciaux sont dus au fait que plusieurs provinces ont choisi d'augmenter la taille de leur échantillon et en ont assumé les coûts supplémentaires. Ce sont la Nouvelle-Écosse, le Nouveau-Brunswick, l'Ontario, le Manitoba et la Saskatchewan.

6. COLLECTE DES DONNÉES

La collecte de données de l'EPA s'effectue chaque mois durant la semaine qui suit la semaine de référence de l'EPA, habituellement la troisième semaine du mois.

6.1 L'interview de l'EPA

Les intervieweurs de Statistique Canada, qui sont des employés à temps partiel engagés et formés pour mener l'EPA, se mettent en rapport avec chacun des logements de l'échantillon pour obtenir les renseignements souhaités sur la population active. Chaque intervieweur rejoint environ 70 logements par mois.

On effectue une visite dans les nouveaux logements de l'échantillon. L'intervieweur commence par recueillir des données socio-démographiques pour chaque membre du ménage, puis les renseignements sur la population active pour tous les membres admissibles. Si le logement est doté d'un téléphone, les interviews suivantes se feront par téléphone moyennant autorisation. Par conséquent, environ 85% de tous les logements sont interrogés par téléphone. Lors des interviews mensuelles subséquentes, l'intervieweur confirme les données socio-démographiques recueillies le premier mois et recueille les données sur la population active pour le mois en cours.

Dans tous les logements, on obtient des données sur tous les membres du ménage auprès d'un membre du ménage bien informé (habituellement la personne qui se trouve à la maison au moment de l'appel de l'intervieweur). Les renseignements recueillis par personne interposée, qui constituent environ 55% de ces données, permettent d'éviter les coûts élevés et le gaspillage de temps que représenteraient les visites ou appels à effectuer par la suite pour obtenir des renseignements auprès de chaque répondant.

À la fin de l'interview mensuelle de l'EPA, les intervieweurs présentent l'enquête supplémentaire à mener, le cas échéant, auprès d'une partie ou de la totalité des membres du ménage ce mois-là.

Si au cours de la période de six mois pendant lequel un logement fait normalement partie de l'échantillon, un ménage entier déménage et qu'un autre le remplace, on recueille des données sur le nouveau ménage durant le reste de la période de six mois.

6.2 Surveillance et contrôle

Tous les intervieweurs de l'EPA travaillent sous la surveillance d'intervieweurs principaux. Ces derniers veillent à ce que les intervieweurs connaissent bien les concepts et les méthodes de l'EPA et de ses nombreuses enquêtes supplémentaires, exercent un contrôle périodique de leurs intervieweurs et examinent les documents qu'ils ont remplis. À leur tour, ils travaillent sous la supervision des gestionnaires du programme de l'EPA, qui se trouvent dans les 8 bureaux régionaux de Statistique Canada.

6.3 Modification de la collecte des données pour l'EUEM

La période de collecte a duré deux semaines afin de favoriser un meilleur taux de réponse à l'enquête supplémentaire. Environ une semaine avant la date de collecte, les ménages ont reçu par la poste un guide les aidant à se préparer à l'interview. Toujours dans le but d'obtenir de meilleurs renseignements, on a proposé à certains ménages de recevoir la visite de l'intervieweur au lieu de répondre comme d'habitude au téléphone. On a supposé que deux catégories de ménage profiteraient le plus de l'aide d'un intervieweur : les ménages locataires dont les membres avaient tous 65 ans ou plus ainsi que ceux où le niveau de scolarité le plus élevé était la 8^{e} année.

L'intervieweur a fait remplir le questionnaire de l'EUEM par le membre du ménage qui, selon le répondant de l'EPA, était le plus au courant de l'équipement du ménage, des appareils électroménagers et de la maison ou de l'appartement.

6.4 Non-réponse à l'EPA

On recommande aux intervieweurs de faire tout leur possible pour mener une interview de l'EPA auprès des membres des ménages admissibles. Pour obtenir la collaboration des personnes qui avaient tout d'abord refusé de participer à l'EPA, le bureau régional envoie une lettre à l'adresse du logement en insistant sur l'importance de l'enquête et demandant la collaboration du ménage. Cette lettre est suivie d'un second appel (ou d'une visite) de l'intervieweur. Si le moment de l'appel (ou de la visite) ne convient pas, l'intervieweur prend rendez-vous pour rappeler à un moment plus opportun. Les logements échantillonnés ne doivent en aucun cas être remplacés par d'autres pour cause de non-réponse.

Chaque mois, même après avoir tout mis en oeuvre pour obtenir des interviews, il reste un petit nombre de ménages non répondants. Si on avait obtenu des renseignements de la part de ces ménages pour l'EPA du mois précédent, ces renseignements sont reportés et servent de données de l'EPA pour le mois en cours. On ne recueille pas de renseignements pour l'enquête supplémentaire auprès de ces ménages.

6.5 Non-réponse à l'Enquête sur l'utilisation de l'énergie par les ménages

Pour les ménages qui ont participé à l'EPA, l'étape suivante consistait à faire remplir le questionnaire de l'EUEM. Il y avait en tout 15,182 ménages admissibles à l'enquête supplémentaire; de ce nombre, 10,982 ont complété une entrevue pour l'EUEM. Il s'agit d'un taux de réponse de 72.3%. On trouvera plus de renseignements sur les taux de réponse dans le chapitre 8 (Qualité des données).

7. TRAITEMENT DES DONNÉES

Le principal résultat de l'enquête est un fichier de microdonnées «épuré». Ce chapitre décrit brièvement les étapes du traitement intervenant dans la réalisation de ce fichier.

7.1 Saisie des données

La saisie des données de l'Enquête sur l'utilisation de l'énergie a été effectuée à l'aide de mini-ordinateurs situés dans le Bureau central de Statistique Canada. Au cours de ce processus, tous les documents contenant au moins un poste rempli par un intervieweur ont été saisis et une version non vérifiée de l'enregistrement informatique a été transmise électroniquement au Groupe des enquêtes spéciales pour traitement ultérieur. Au total, 15,277 documents relatifs à l'enquête ont été saisis et transmis.[5]

7.2 Vérification

Les données saisies comportent des erreurs faites par les intervieweurs ou par les préposés à la saisie. La vérification a pour but de corriger ces erreurs pour que chaque enregistrement soit cohérent.

La vérification a été effectuée de manière «descendante», ce qui signifie que les valeurs manquantes ou improbables ont été corrigées en suivant le déroulement du questionnaire. Cependant, les renseignements sur la marque et le modèle des électroménagers et de l'équipement ont été conservés, peu importe les réponses données aux questions antérieures.

Le questionnaire se divisait en deux sections, l'une s'adressant à ceux qui habitent une maison et l'autre à ceux qui habitent un appartement. Dans certains enregistrements, les répondants se sont trompés de section. La vérification a alors déplacé les réponses à la bonne section.

Les vérifications ont également converti les réponses données selon le système métrique en unités impériales et a modifié le codage selon les fourchettes appropriées.

Les enregistrements ne comportant aucune réponse, ou très peu, aux questions fondamentales ont été supprimés.

---

[5] On notera que le nombre de questionnaires saisis (15,277) est supérieur au nombre de ménages admissibles (15,182). La différence est due à certains ménages qui n'avaient pas participé à l'EPA. La presque totalité de ces ménages n'ont pas participé non plus à l'EUEM, même si un questionnaire a été transmis au Bureau central.

7.3 Création de variables calculées

On a calculé un certain nombre de données élémentaires du fichier de microdonnées en combinant des postes du questionnaire pour faciliter l'analyse des données. Par exemple, la variable Q85ADV représente le combustible utilisé par le système de chauffage principal. Elle est calculée à partir des variables Q78 (thermopompes), Q84A (plinthes chauffantes et chauffage radiant) et Q85A (combustible utilisé par des appareils de chauffage autres que les thermopompes, les plinthes électriques et le chauffage radiant).

La variable Q183INDR représente un autre exemple. Elle représente le nombre total d'ampoules à incandescence qui sont utilisées à l'intérieur. Elle correspond tout simplement à la somme des variables Q183A à Q183I.

7.4 Pondération

Le principe sur lequel repose l'estimation pour un échantillon probabiliste comme celui de l'EPA veut que chacune des personnes faisant partie de l'échantillon «représente», à part elle-même, plusieurs autres personnes qui n'en font pas partie. Par exemple, dans un échantillon aléatoire simple de 2% de la population, chaque personne dans l'échantillon représente 50 personnes dans la population.

La phase de pondération est l'étape où l'on calcule ce nombre (poids) pour chaque enregistrement et où on l'enregistre dans le fichier des microdonnées pour chacun d'entre eux. Il faut utiliser ce poids pour faire des estimations utiles à partir du fichier de microdonnées. Par exemple, si l'on veut estimer le nombre de maisons individuelles dotées d'un grand réfrigérateur, on choisit les enregistrements se rapportant aux ménages de l'échantillon qui possèdent cette caractéristique et on fait la somme des poids inscrits sur ces enregistrements.

On trouvera au chapitre 11 des précisions sur la méthode utilisée pour calculer ces poids.

7.5 Suppression des renseignements confidentiels

Il convient de mentionner que les fichiers de microdonnées à grande diffusion décrits ci-dessus diffèrent à bien des égards des fichiers «principaux» de l'enquête conservés par Statistique Canada. Ces écarts sont dus à des mesures prises pour protéger l'anonymat des répondants à l'enquête. Les utilisateurs qui veulent consulter des renseignements exclus des fichiers de microdonnées peuvent acheter des totalisations personnalisées. Les estimations produites seront communiquées à l'utilisateur, à condition qu'elles soient conformes aux lignes directrices pour l'analyse et la diffusion décrites au chapitre 9 de ce document.

Identificateurs géographiques

Le fichier principal de l'enquête comprend des identificateurs géographiques explicites pour la province, la région économique et la région métropolitaine de recensement. Les fichiers de microdonnées à grande diffusion identifient la province ainsi que le caractère urbain ou rural de la région. En outre, les régions urbaines sont divisées selon la taille.

Noms des fournisseurs d'énergie

Les répondants ont donné le nom de leurs fournisseurs d'énergie, ainsi que l'autorisation de communiquer avec ces derniers. Cette information était destinée uniquement à Statistique Canada et a été supprimée.

Situation vis-à-vis de l'activité

Les catégories «chômeur» et «inactif» ont été fusionnées.

Loyer

Les loyers de plus de $999 ont été fusionnés.

8. QUALITÉ DES DONNÉES

8.1 Taux de réponse

Le tableau ci-dessous résume les taux de réponse à l'Enquête sur l'utilisation de l'énergie par les ménages.

PROVINCE	Nombre de ménages répondants	Taux de réponse (%)
Terre-Neuve	837	82.6
Île-du-Prince-Édouard	732	77.3
Nouvelle-Écosse	1,518	70.1
Nouveau-Brunswick	1,235	65.8
Québec	720	78.8
Ontario	1,407	70.8
Manitoba	1,476	71.4
Saskatchewan	1,704	71.3
Alberta	674	73.8
Colombie-Britannique	679	74.5
CANADA	**10,982**	**72.3**

Nota : Le taux de réponse correspond au nombre de ménages qui ont répondu à l'EUEM par rapport au nombre de ménages qui ont répondu à l'EPA et qui étaient admissibles à l'EUEM. Le taux de réponse global pour l'EPA de février 1993 était de 94.5%.

8.2 Erreurs relatives à l'enquête

L'enquête permet de produire des estimations fondées sur les données portant sur un échantillon et recueillies auprès de ses membres. On aurait pu obtenir des estimations légèrement différentes si l'on avait procédé à un recensement complet en faisant appel au même questionnaire et aux mêmes intervieweurs, superviseurs, méthodes de traitement, etc., que pour cette enquête. L'écart entre les estimations découlant de l'échantillon et celles tirées d'un dénombrement complet effectué dans des conditions semblables est appelé erreur d'échantillonnage des estimations.

Des erreurs qui ne se rapportent pas à l'échantillonnage peuvent se produire à presque toutes les phases d'une enquête. Il se peut que les intervieweurs comprennent mal les instructions, que les répondants fassent des erreurs en répondant aux questions, que des réponses soient mal inscrites sur le questionnaire et que des erreurs se produisent 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 tirées des données recueillies dans le cadre de l'enquête. Toutefois, les erreurs qui se produisent systématiquement contribueront à des biais dans les estimations de l'enquête. On a consacré beaucoup de temps et d'efforts à réduire les erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été appliquées à chaque étape du cycle de collecte et de traitement des données afin de surveiller la qualité des données. Ces mesures comprenaient l'emploi d'intervieweurs très expérimentés, une formation poussée des intervieweurs pour ce qui est des procédures d'enquête et du questionnaire, l'observation des intervieweurs afin de trouver les problèmes liés à la conception du questionnaire ou à la mauvaise interprétation des instructions, des procédures afin de s'assurer que les erreurs de saisie des données étaient minimisées et des vérifications de la qualité du codage et du contrôle afin de vérifier la logique du traitement.

L'effet de la non-réponse sur les résultats de l'enquête constitue une source importante d'erreurs non dues à l'échantillonnage dans les enquêtes. L'importance de la non-réponse varie d'une non-réponse partielle (le fait de ne pas répondre à une ou à plusieurs questions) à une non-réponse totale. Il y a non-réponse totale quand l'intervieweur n'a pu communiquer avec le répondant, lorsqu'aucun membre du ménage ne pouvait fournir les renseignements ou quand le répondant a refusé de participer à l'enquête. On a traité les cas de non-réponse totale en ajustant le poids des ménages qui ont répondu à l'enquête afin de compenser ceux qui n'ont pas répondu.

Dans la majorité des cas, il y a non-réponse partielle à l'enquête quand le répondant n'a pas compris ou a mal interprété une question, quand il a refusé de répondre à une question, quand il ne pouvait se rappeler le renseignement demandé ou qu'il n'a pu fournir de renseignements de remplacement.

Comme les estimations d'une enquête par sondage comportent inévitablement une erreur d'échantillonnage, de bonnes méthodes statistiques exigent que les chercheurs fournissent aux utilisateurs une certaine indication de la grandeur de cette erreur d'échantillonnage. Ce chapitre de la documentation décrit les mesures de l'erreur d'échantillonnage utilisées habituellement par Statistique Canada et dont le Bureau conseille vivement l'emploi aux utilisateurs qui font des estimations à partir de ce fichier de microdonnées.

La mesure de l'importance éventuelle des erreurs d'échantillonnage est fondée sur l'erreur-type des estimations découlant des résultats de l'enquête.

Cependant, en raison de la diversité des estimations que l'on peut tirer 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 (c.v.) 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 qu'à partir des résultats de l'enquête, on estime que 78.5% des foyers canadiens possèdent un four micro-ondes et on constate que cette estimation a une erreur-type de 0.006. Le coefficient de variation de cette estimation est calculé comme suit :

$$\left(\frac{0.006}{0.785} \right) \; x \; 100 \; \% \; = \; 0.8 \; \%$$

9. LIGNES DIRECTRICES POUR LA PUBLICATION ET LA DIFFUSION

Ce chapitre décrit les lignes directrices que doivent suivre les utilisateurs qui publient ou diffusent autrement des données tirées des bandes de microdonnées de l'enquête. Ces lignes directrices devraient permettre aux utilisateurs d'arriver aux mêmes chiffres que Statistique Canada, tout en étant en mesure d'obtenir, conformément à ces lignes directrices, des chiffres qui n'ont pas été publiés.

9.1 Organisation du fichier de microdonnées

Le fichier renferme les données provenant de trois sources : l'EPA, l'Enquête sur les loyers réalisée de concert avec l'EPA et l'Enquête supplémentaire de l'EPA sur l'utilisation de l'énergie par les ménages.

La partie EUEM du fichier correspond au questionnaire : les questions 13 à 77 et 164 à 195 concernent tous les logements, les questions 78 à 142 et 161 à 163, aux logements qualifiés de maisons et les questions 143 à 160, aux logements considérés comme des appartements.

Pour acheminer un logement à la section «maison» ou à la section «appartement» du questionnaire, on a utilisé la classification établie pour l'EPA quant au genre de logement et au mode d'occupation (propriétaire-locataire).

Les genres de logement suivants ont répondu aux questions destinées aux maisons :
- maison individuelle,
- maison jumelée,
- maison en rangée,
- **duplex, s'il appartenait à un membre du ménage**,
- maison mobile.

Les genres de logement suivants ont répondu aux questions destinées aux appartements :
- **duplex, s'il était loué**,
- appartement dans un immeuble de moins de cinq étages,
- appartement dans un immeuble de cinq étages ou plus.

Les utilisateurs doivent se souvenir qu'il faut combiner des données sur les maisons et celles sur les appartements pour mesurer certaines caractéristiques des logements. Par exemple, si l'on veut estimer le pourcentage de logements qui utilisent les plinthes électriques comme chauffage principal, il faut combiner les réponses des questions 84 et 144.

D'autre part, pour estimer le pourcentage de duplex dotés d'un foyer au bois, il faut combiner le nombre de duplex qui ont répondu par l'affirmative à la question 91 ou à la question 148, puis diviser la somme par le nombre total de duplex.

9.2 Lignes directrices pour l'arrondissement

Afin que les estimations destinées à la publication ou à toute autre forme de diffusion qui sont tirées de ces bandes de microdonnées correspondent à celles de Statistique Canada, nous conseillons vivement aux utilisateurs de se conformer aux lignes directrices ci-après en ce qui concerne l'arrondissement de ces estimations :

a) Les estimations dans le corps d'un tableau statistique doivent être arrondies à la centaine près au moyen de la technique d'arrondissement classique. Dans cette technique, si le premier ou le seul chiffre à supprimer se situe entre 0 à 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente de 1 la valeur du dernier chiffre à conserver. Par exemple, lorsqu'on cherche à arrondir à la centaine près de la façon classique, si les deux derniers chiffres se situent entre 00 et 49, il faut les remplacer par 00 et laisser le chiffre précédent (le chiffre des centaines) tel quel. Si les deux derniers chiffres se situent entre 50 et 99, on les remplace par 00 et on augmente de 1 le chiffre précédent.

b) Les totaux partiels marginaux et les totaux marginaux des tableaux statistiques doivent être calculés à partir de leurs composantes correspondantes non arrondies, puis arrondis à leur tour à la centaine près à l'aide de la technique d'arrondissement classique.

c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir d'éléments non arrondis (c'est-à-dire les numérateurs et (ou) les dénominateurs), puis arrondis à une décimale au moyen de la technique d'arrondissement classique. Lorsqu'on cherche à arrondir à un seul chiffre par cette technique, 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, on augmente de 1 le dernier chiffre à conserver.

d) Les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la centaine près (ou à la décimale près) en employant la technique d'arrondissement classique.

e) Si, en raison de limitations d'ordre technique ou de toute autre nature, on utilise une autre technique d'arrondissement que l'arrondissement classique, de sorte que les estimations à publier ou à diffuser sous une forme quelconque diffèrent des estimations correspondantes publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces divergences dans le ou les documents à publier ou à diffuser.

f) Des estimations non arrondies ne doivent en aucun cas être publiées ou diffusées par les utilisateurs. Des estimations qui ne sont pas arrondies laissent supposer qu'elles sont beaucoup plus précises qu'elles ne le sont en réalité.

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

Le plan d'échantillonnage utilisé pour l'Enquête sur l'utilisation de l'énergie par les ménages n'était pas autopondéré. Pour produire des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent employer le poids approprié. Sinon, les estimations calculées à partir des bandes de microdonnées ne peuvent être considérées comme représentatives de la population observée et ne correspondront pas à celles de Statistique Canada.

Les utilisateurs doivent également noter qu'en raison de la manière dont ils traitent le champ du poids, certains progiciels ne permettent peut-être pas d'obtenir des estimations correspondant exactement à celles de Statistique Canada.

9.3.1 Définitions des genres d'estimations : de type nominal ou quantitatives

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

(1) Estimations de type nominal

Les estimations de type nominal sont des estimations du nombre ou du pourcentage de personnes dans la population visée par l'enquête qui possèdent certaines caractéristiques ou qui font partie d'une catégorie définie. Le nombre de maisons à un étage ou le pourcentage de maisons possédant un four à micro-ondes sont des exemples d'estimations de ce genre. On peut aussi utiliser l'expression «estimation d'un agrégat» pour parler d'une estimation du nombre de personnes qui possèdent une caractéristique donnée.

Exemples de questions de type nominal :

Q : Avez-vous un foyer au bois dans votre maison?
R : Oui / Non

Q : Quelle proportion du sous-sol est-elle chauffée? (N'inscrivez qu'une seule réponse)
R : Tout le sous-sol / Plus de la moitié / Environ la moitié / Moins de la moitié / Ne sait pas

(2) Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes ou d'autres mesures de tendance centrale de quantités basées sur certains ou sur tous les membres de la population visée par l'enquête. Elles comprennent aussi explicitement des estimations de la forme $\bar{X}/\bar{Y}$ où $\bar{X}$ est une estimation de la quantité totale pour la population visée par l'enquête et $\bar{Y}$ est une estimation du nombre de personnes dans la population visée par l'enquête qui contribuent à cette quantité totale.

Un exemple d'une estimation quantitative est le nombre moyen de fenêtres à vitrage triple pour les maisons qui en possèdent. Le numérateur correspond à une estimation du nombre total de fenêtres à vitrage triple et le dénominateur, au nombre de maisons dotées de fenêtres à vitrage triple.

Exemples de questions de type quantitatif :

Q : Combien (portes en bois) ... avec des contre-portes?
R : |_|

Q : Combien d'ampoules ordinaires (à incandescence) avez-vous dans votre ... cuisine?
R : |_|

9.3.2 Totalisation d'estimations de type nominal

On peut obtenir des estimations du nombre de personnes possédant une caractéristique particulière à partir du fichier de microdonnées en additionnant les poids finals de tous les enregistrements possédant la ou les caractéristiques qui nous intéressent. On obtient les proportions et les rapports en (a) additionnant les poids finals des enregistrements possédant la caractéristique qui nous intéresse, pour le numérateur, (b) en additionnant les poids finals des enregistrements possédant la caractéristique qui nous intéresse, pour le dénominateur, puis (c) en divisant l'estimation du numérateur par celle du dénominateur.

9.3.3 Totalisation d'estimations quantitatives

On peut obtenir des estimations de quantités à partir du fichier de microdonnées en multipliant la valeur de la variable qui nous intéresse par le poids final de chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements qui nous intéressent. Par exemple, pour obtenir une estimation du nombre total de fenêtres à vitrage triple utilisées dans les maisons individuelles, on multiplie la somme des valeurs inscrites pour les variables Q128A1 (puits de lumière), Q129A1 (fenêtres panoramiques) et Q129A2 (autres dimensions) - en prenant bien soin de ne pas tenir compte des 9 ou 99 correspondant à «non déclaré» - par le poids final de l'enregistrement. Ensuite on additionne ce résultat par tous les enregistrements ayant LOGEMENT=1 (maison individuelle).

Pour obtenir une moyenne pondérée, on calcule le numérateur comme une estimation quantitative et le dénominateur comme une estimation de type nominal. Par exemple, pour estimer le nombre moyen de fenêtres à vitrage triple utilisées dans les maisons individuelles, a) on estime le nombre total de fenêtres à vitrage triple de la manière décrite plus haut, b) on estime le nombre de maisons de cette catégorie en faisant la somme des poids finals de tous les enregistrements ayant LOGEMENT=1, puis c) on divise l'estimation a) par l'estimation b).

Notons que dans ce cas la moyenne est calculée pour la totalité des maisons individuelles, qu'elles soient ou non dotées de fenêtres à vitrage triple. Il est possible d'obtenir une estimation plus utile en limitant le dénominateur au nombre de maisons individuelles qui sont dotées de fenêtres à vitrage triple. Autrement dit, à b), on fait la somme des poids des enregistrements ayant LOGEMENT=1 et la somme des variables Q128A1, Q129A1 et Q129A2 supérieure à zéro.

9.4 Lignes directrices pour l'analyse statistique

L'Enquête sur l'utilisation de l'énergie par les ménages est basée sur un plan complexe, comportant une stratification et de multiples degrés de sélection, ainsi que des probabilités inégales de sélection des répondants. L'utilisation des données provenant d'enquêtes aussi complexes pose des difficultés aux analystes parce que le plan d'enquête et les probabilités de sélection influent sur les méthodes d'estimation et de calcul de la variance qui doivent être utilisées.

Bien que de nombreuses méthodes d'analyse qui font partie des progiciels statistiques permettent d'utiliser des poids, le poids diffère, sous l'angle de la définition ou du sens, de celui qui convient à une enquête par sondage. Par conséquent, si les estimations faites au moyen de ces progiciels sont exactes dans bien des cas, les variances calculées n'ont pratiquement aucune signification.

Dans le cas de nombreuses techniques d'analyse (par exemple, la régression linéaire, la régression logistique, l'analyse de variance), il existe un moyen de rendre l'application des progiciels standard plus significative. Si l'on transforme les poids des enregistrements de manière à ce que le poids moyen soit un (1), les résultats produits par les progiciels standard seront plus raisonnables et tiendront compte des probabilités inégales de sélection, bien qu'ils continueront à ne faire aucun cas de la stratification et de la structure en grappes du plan d'échantillonnage. On peut effectuer cette transformation en divisant chaque poids par le poids moyen global avant d'entreprendre l'analyse.

Afin de fournir un moyen d'évaluer la qualité des estimations totalisées, Statistique Canada a produit un ensemble de tables de variabilité d'échantillonnage approximative pour l'EUEM. On peut employer ces tables pour obtenir des coefficients de variation approximatifs pour les estimations et les proportions de type nominal. Voir le chapitre 10 pour plus de détails.

9.5 Lignes directrices pour la diffusion en fonction de la variabilité d'échantillonnage

Avant de diffuser et (ou) de publier des estimations tirées de ces bandes de microdonnées, les utilisateurs doivent commencer par établir le nombre d'enquêtés qui ont fourni les données entrant dans le calcul de l'estimation. Si ce nombre est inférieur à 30, l'estimation pondérée ne doit pas être diffusée quelle que soit la valeur de son coefficient de variation. Pour les estimations pondérées basées sur des échantillons composés de 30 enquêtés ou plus, les utilisateurs doivent calculer le coefficient de variation de l'estimation et suivre les lignes directrices ci-dessous.

Lignes directrices relatives à la variabilité d'échantillonnage

Type d'estimation	c.v. (en %)	Lignes directrices
1. Sans réserve	0.0 - 16.5	On peut envisager une diffusion générale non restreinte des estimations. Aucune annotation particulière n'est nécessaire.
2. Sous réserve	16.6 - 25.0	On peut envisager une diffusion générale non restreinte des estimations, en y joignant une mise en garde aux utilisateurs suivants quant à la variabilité d'échantillonnage élevée liée aux estimations. Les estimations de ce genre doivent être identifiées par la lettre R (ou d'une autre manière semblable).
3. Confidentielle	25.1 - 33.3	On ne peut envisager une diffusion générale non restreinte des estimations que si la variabilité d'échantillonnage est obtenue au moyen d'une méthode de calcul exact de la variance. À défaut de variances exactes, il faut supprimer ces estimations des tableaux statistiques et les remplacer par des tirets (---).
4. Diffusion interdite	33.4 ou plus	Les estimations ne peuvent être diffusées sous aucune forme OU dans aucun cas. Dans les tableaux statistiques, ces estimations doivent être supprimées et remplacées par des tirets (---).

Nota : Ces lignes directrices relatives à la variabilité d'échantillonnage doivent être appliquées à des estimations pondérées arrondies.

10. TABLES DE VARIABILITÉ D'ÉCHANTILLONNAGE APPROXIMATIVE

Tout d'abord, afin de permettre aux utilisateurs d'avoir facilement accès à des coefficients de variation qui s'appliqueraient à une gamme étendue d'estimations de type nominal réalisées à partir de cette bande de microdonnées, on a préparé un ensemble de «tables à consulter», appelées Tables de variabilité d'échantillonnage approximative.

Les coefficients de variation (c.v.) dans ces tables sont obtenus au moyen de la formule de la variance utilisée pour l'échantillonnage aléatoire simple en y incorporant un facteur reflétant la structure en grappes à plusieurs degrés du plan d'échantillonnage. Pour obtenir ce facteur, appelé effet du plan, on a commencé par calculer les effets du plan pour une gamme étendue de caractéristiques, puis on a choisi parmi les nombres ainsi obtenus une valeur modérée à employer dans les tables à consulter qui serait ensuite appliquée à l'ensemble des caractéristiques.

Voici la table des effets du plan, des tailles de l'échantillon et des chiffres de population par province employés pour produire les tables de variabilité d'échantillonnage approximative.

PROVINCE	EFFETS DU PLAN	TAILLE DE L'ÉCHANTILLON	POPULATION CIBLE
Terre-Neuve	3.0	837	186,070
Île-du-Prince-Édouard	2.1	732	45,736
Nouvelle-Écosse	3.0	1,518	336,080
Nouveau-Brunswick	2.5	1,235	260,915
Québec	2.4	720	2,710,836
Ontario	2.7	1,407	3,810,478
Manitoba	2.6	1,476	402,524
Saskatchewan	2.3	1,704	368,270
Alberta	2.2	674	934,816
Colombie-Britannique	3.0	679	1,303,492
Provinces Atlantiques	3.0	4,322	828,801
Provinces des Prairies	2.6	3,854	1,705,610
CANADA	**6.0**	**10,982**	**10,359,217**

Il convient de mentionner que tous les coefficients de variation dans les tables de variabilité d'échantillonnage approximative sont approximatifs et ne doivent donc pas être considérés comme des valeurs officielles.

On peut se procurer, contre recouvrement des coûts, des estimations de la variance réelle de variables particulières auprès de Statistique Canada. L'utilisation d'estimations de la variance réelle peut permettre aux utilisateurs de diffuser des estimations que l'on ne pourrait diffuser autrement, c'est-à-dire des estimations avec des coefficients de variation tels qu'elles sont «confidentielles» ou à «diffusion interdite».

N'oubliez pas : Si le nombre d'observations sur lesquelles une estimation est basée est inférieur à 30, il ne faut pas diffuser l'estimation pondérée quelle que soit la valeur de son coefficient de variation.

10.1 Comment utiliser les tables de c.v. pour les estimations de type nominal

Les règles suivantes doivent permettre à l'utilisateur de calculer à partir des tables de variabilité d'échantillonnage les coefficients de variation approximatifs d'estimations relatives au nombre, à la proportion ou au pourcentage de personnes dans la population enquêtée possédant une caractéristique donnée ainsi que des rapports et des écarts entre ces estimations.

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

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

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

Le coefficient de variation d'une proportion ou d'un pourcentage estimé dépend à la fois de la grandeur de cette proportion ou de ce pourcentage et de la grandeur du total sur lequel est fondé cette proportion ou ce pourcentage. Les proportions ou les pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque cette proportion ou ce pourcentage est fondé sur un sous-ensemble de la population. Par exemple, la proportion de maisons individuelles ayant un foyer au bois est plus fiable que le nombre estimé de maisons individuelles ayant un foyer au bois. (À noter que dans les tables, la valeur des coefficients de variation décroît de gauche à droite sur une même ligne.)

Lorsque la proportion ou le pourcentage est fondé sur la population totale de la région géographique à laquelle la table s'applique (P=X/N), le coefficient de variation de la proportion ou du pourcentage (P) est égal à celui du numérateur (X) de la proportion ou du pourcentage. Dans ce cas, on peut appliquer la règle 1.

Lorsque la proportion ou le pourcentage est fondé sur un sous-ensemble de la population totale (par exemple, un genre de logement particulier), il faut se reporter à la proportion ou au pourcentage (en haut de la table) et au numérateur de la proportion ou du pourcentage (dans la colonne de gauche de la table). Le coefficient de variation se trouve à l'intersection de la ligne et de la colonne appropriées.

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

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

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1 \ \sigma_1)^2 \ + \ (\hat{X}_2 \ \sigma_2)^2}$$

où $\hat{X}_1$ représente l'estimation 1, $\hat{X}_2$ l'estimation 2, et α_1 et α_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 donne un résultat exact en ce qui a trait à la différence entre des caractéristiques distinctes et non corrélées, mais elle ne donne que des résultats approximatifs dans les autres cas.

Règle 4 : Estimations de rapports

Si le numérateur est un sous-ensemble du dénominateur, il faut convertir le rapport en pourcentage et appliquer la règle 2 (on pourrait citer comme exemple d'une telle situation le cas où le dénominateur représente le nombre de maisons individuelles et le numérateur, le nombre de maisons individuelles ayant un foyer au bois).

Si le numérateur n'est pas un sous-ensemble du dénominateur, (par exemple, le rapport entre le nombre de maisons ayant un foyer au **bois** et le nombre de maisons ayant un foyer au **gaz**), l'écart-type du rapport entre les estimations est approximativement égal à la racine carrée de la somme des carrés de chaque coefficient de variation pris séparément multipliée par R. L'erreur-type d'un rapport ($\hat{R} = \hat{X}_1 / \hat{X}_2$) est donc :

$$\sigma_{\hat{R}} \ = \ \hat{R} \ \sqrt{\sigma_1^2 \ + \ \sigma_2^2}$$

où α_1 et α_2 sont les coefficients de variation de $\hat{X}_1$ et $\hat{X}_2$ respectivement.

Le coefficient de variation de $\hat{R}$ est donné par $\sigma_{\hat{R}}/\hat{R}$. La formule tend à surestimer l'erreur si $\hat{X}_1$ et $\hat{X}_2$ sont corrélées positivement et à sous-estimer l'erreur si $\hat{X}_1$ et $\hat{X}_2$ sont corrélées négativement.

Règle 5 : Estimations des différences entre des rapports

Dans ce cas, on combine les règles 3 et 4. On commence par calculer les coefficients de variation des deux rapports au moyen de la règle 4, puis le coefficient de variation de leur différence au moyen de la règle 3.

10.2 Exemples d'utilisation des tables de c.v. pour des estimations de type nominal

Les exemples «concrets» ci-dessous sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter.

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

Supposons qu'un utilisateur estime que 2,321,823 ménages au Canada utilisent des ampoules halogènes à l'intérieur et à l'extérieur de la maison. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

1) Reportez-vous à la table de c.v. pour le CANADA.

2) L'agrégat estimé (2,321,823) ne se trouve pas dans la colonne de gauche («Numérateur du pourcentage»), il faut donc utiliser le nombre qui s'en rapproche le plus, c'est-à-dire 2,000,000.

3) Le coefficient de variation pour un agrégat estimé est la première inscription, autre que des astérisques, sur cette ligne, c'est-à-dire 4.8%.

4) Le coefficient de variation approximatif de l'estimation est donc 4.8%. L'estimation que 2,321,823 ménages au Canada utilisent des ampoules halogènes peut être publiée sans mise en garde.

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

Supposons que l'utilisateur estime que 2,849,308/5,823,176 = 48.9% des maisons individuelles au Canada ont un étage. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

1) Reportez-vous à la table de c.v. pour le CANADA.

2) Parce que l'estimation est un pourcentage basé sur un sous-ensemble de la population totale (c.-à-d. les maisons individuelles), il faut utiliser à la fois le pourcentage (48.9%) et la partie numérateur du pourcentage (2,849,308) pour déterminer le coefficient de variation.

3) Le numérateur 2,849,308 ne figure pas dans la colonne de gauche («Numérateur du pourcentage»). Il faut donc utiliser le nombre qui s'en rapproche le plus, soit 3,000,000. De même, l'estimation du pourcentage ne figure pas parmi les en-têtes de colonnes, il faut donc utiliser le nombre qui s'en rapproche le plus, 50.0%.

4) Le nombre qui se trouve à l'intersection de la ligne 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%. L'estimation que 48.9% des maisons individuelles ont un étage peut être publiée sans mise en garde.

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

Supposons qu'un utilisateur estime que dans 3,830,339/7,261,564 = 52.7% des maisons, on baisse la température la nuit, tandis que 782,367/3,097,653 = 25.3% des appartements font de même. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

1) À l'aide de la table de c.v. pour le CANADA utilisée de la même façon que dans l'exemple 2, on trouve que le c.v. de l'estimation pour les maisons est de 2.7% et que le c.v. de l'estimation pour les appartements est de 7.5%.

2) Selon la règle 3, l'erreur-type pour 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 α_1 et α_2 sont les coefficients de variation de $\hat{X}_1$ et $\hat{X}_2$ respectivement.

C'est-à-dire que l'erreur-type de la différence $\hat{d} = (.527 - .253) = .274$ est :

$$\sigma_{\hat{d}} = \sqrt{[(0.527)(0.027)]^2 + [(0.253)(0.075)]^2}$$

$$= \sqrt{(0.0002025) + (0.0003601)}$$

$$= 0.024$$

3) Le coefficient de variation de $\hat{d}$ est donné par $\sigma_{\hat{d}}/\hat{d} = .024/.274 = .088$.

4) Le coefficient de variation approximatif de la différence entre les estimations est donc 8.8%. Cette estimation peut être publiée sans mise en garde.

Exemple 4 : Estimations de rapports

Supposons que l'utilisateur estime qu'il y a 2,053,991 maisons individuelles avec un foyer au bois et qu'il y a 316,530 maisons individuelles avec un foyer au gaz. L'utilisateur veut comparer ces deux estimations sous forme d'un rapport. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

1) Tout d'abord, cette estimation est une estimation de rapport, où le numérateur de l'estimation (= $\bar{X}_1$) est le nombre de maisons individuelles avec un foyer au bois. Le dénominateur de l'estimation (= $\bar{X}_2$) est le nombre de maisons individuelles avec un foyer au gaz.

2) Reportez-vous à la table de c.v. pour le CANADA.

3) Le numérateur de cette estimation de rapport est 2,053,991. Le chiffre qui se rapproche le plus de ce nombre est 2,000,000. On détermine le coefficient de variation pour cette estimation en trouvant la première inscription, autre que des astérisques, sur cette ligne, soit, 4.8%.

4) Le dénominateur de cette estimation de rapport est 316,530. Le chiffre qui se rapproche le plus de ce nombre est 300,000. On détermine le coefficient de variation de cette estimation en trouvant la première inscription, autre que des astérisques, sur cette ligne, soit, 13.4%.

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 α_1 et α_2 sont les coefficients de variation de $\bar{X}_1$ et $\bar{X}_2$ respectivement.

Donc

$$\alpha_{\hat{R}} = \sqrt{(0.048)^2 + (0.134)^2}$$

$$= 0.142$$

Le rapport obtenu entre le nombre de maisons individuelles avec un foyer au bois et le nombre de maisons individuelles avec un foyer au gaz est 2,053,991/316,550, soit 6.49:1. Le coefficient de variation de cette estimation est 14.2%. L'estimation peut donc être diffusée sans mise en garde.

10.3 Comment utiliser les tables de c.v. pour calculer des limites de confiance

Bien que les coefficients de variation soient largement utilisés, l'intervalle de confiance d'une estimation donne une mesure intuitive plus significative de l'erreur d'échantillonnage. Un intervalle de confiance est une façon d'énoncer la probabilité que la valeur vraie de la population se situe dans une plage de valeurs données. Par exemple, un intervalle de confiance à 95% peut être décrit comme suit :

Si l'échantillonnage de la population se répète à l'infini, chacun des échantillons donnant un nouvel intervalle de confiance pour une estimation, l'intervalle contiendra alors la valeur vraie de la population dans 95% des cas.

Une fois déterminée l'erreur-type d'une estimation, on peut calculer des intervalles de confiance pour les estimations en partant de l'hypothèse qu'en procédant à un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique de la population sont réparties selon une distribution normale 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. On appelle ces différents degrés de confiance des niveaux de confiance.

Les intervalles de confiance pour une estimation, $\bar{X}$, sont généralement exprimés sous forme de deux nombres, l'un étant inférieur à l'estimation et l'autre supérieur à celle-ci, sous la forme ($\bar{X}$-k, $\bar{X}$+k), où k varie selon le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

On peut calculer directement les intervalles de confiance pour une estimation à partir des tables de variabilité d'échantillonnage approximative, en commençant par trouver, dans la table appropriée, le coefficient de variation de l'estimation $\bar{X}$, puis en utilisant la formule ci-dessous pour obtenir l'intervalle de confiance IC correspondant :

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

où $\alpha_{\bar{X}}$ est le coefficient de variation trouvé pour $\bar{X}$

t = 1 si l'on désire un intervalle de confiance à 68%
t = 1.6 si l'on désire un intervalle de confiance à 90%
t = 2 si l'on désire un intervalle de confiance à 95%
t = 3 si l'on désire un intervalle de confiance à 99%.

Nota : Les lignes directrices qui s'appliquent à la diffusion des estimations s'appliquent aussi aux intervalles de confiance. Par exemple, si l'estimation ne peut être diffusée, alors l'intervalle de confiance ne peut, lui non plus, être diffusé.

10.4 Exemples d'utilisation de tables de c.v. pour obtenir des limites de confiance

Voici comment on calculerait un intervalle de confiance à 95% pour la proportion estimée de maisons individuelles ayant un étage (d'après l'exemple 2 de la section 10.2).

X = 48.9% (ou, sous forme de proportion, .489)

t = 2

α_X = 3.1% (.031 sous forme de proportion) est le coefficient de variation de cette estimation tel que déterminé à partir des tables.

IC_x = {.489 - (2) (.489) (.031), .489 + (2) (.489) (.031)}

IC_x = {.489 - .030, .489 + .030}

IC_x = {.459, .519}

On peut dire, avec une probabilité de 95%, qu'entre 45.9% et 51.9% des maisons individuelles ont un étage.

10.5 Comment utiliser les tables de c.v. pour effectuer un test t

On peut aussi utiliser les erreurs-types pour effectuer des tests d'hypothèses, une technique qui permet de faire la distinction entre les paramètres d'une population à l'aide d'estimations basées sur un échantillon. Ces estimations peuvent être des nombres, des moyennes, des pourcentages, des rapports, etc. Les tests peuvent être effectués à divers niveaux de signification, où un niveau de signification est la probabilité de conclure que les caractéristiques sont différentes quand, en fait, elles sont identiques.

Soit $\mathbf{X_1}$ et $\mathbf{X_2}$ des estimations basées sur un échantillon pour deux caractéristiques qui nous intéressent. Posons que l'erreur-type de la différence $\bar{X}_1 - \bar{X}_2$ est $\sigma_{\hat{d}}$

Si $t = \dfrac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ est compris entre -2 et 2, alors on ne peut tirer de conclusion à propos de la différence entre les caractéristiques au niveau de signification de 5%. Toutefois, si ce rapport est inférieur à -2 ou supérieur à +2, la différence observée est significative au niveau de 0.05. Autrement dit, les caractéristiques sont significatives.

10.6 Exemple d'utilisation des tables de c.v. pour effectuer un test t

Supposons que nous désirons tester, au niveau de signification de 5%, l'hypothèse qu'il n'y a pas de différence entre la proportion de maisons où la température est baissée la nuit et la proportion d'appartements où l'on fait de même. Dans l'exemple 3 de la section 10.2, nous avons trouvé que l'erreur-type de la différence entre ces deux estimations était .024. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.527 - 0.253}{0.024} = \frac{0.274}{0.024} = 11.42.$$

Puisque t = 11.42 est supérieur à 2, on doit conclure qu'il y a une différence significative entre les deux estimations au niveau de signification de 0.05.

10.7 Coefficients de variation d'estimations quantitatives

Pour calculer l'erreur d'échantillonnage d'estimations quantitatives, il faudrait élaborer des tables spéciales, ce qui n'a pas été fait puisque les variables de l'Enquête sur l'utilisation de l'énergie par les ménages sont surtout de type nominal.

Néanmoins, le coefficient de variation d'un total quantitatif est généralement supérieur à celui de l'estimation de type nominal correspondante (c'est-à-dire l'estimation du nombre de personnes contribuant à l'estimation quantitative). Si l'estimation de type nominal correspondante ne peut être diffusée, il en sera de même pour l'estimation quantitative. Par exemple, le coefficient de variation du nombre total de foyers au bois serait supérieur au coefficient de variation de la proportion correspondante de ménages ayant un foyer au bois. Par conséquent, si le coefficient de variation de la proportion ne peut être diffusé, ce sera aussi le cas pour le coefficient de variation de l'estimation quantitative correspondante.

On peut calculer, au besoin, les coefficients de variation d'estimations de ce genre pour une estimation particulière au moyen d'une technique appelée pseudo-réplication, qui consiste à diviser les enregistrements des fichiers de microdonnées en sous-groupes (ou sous-échantillons) et à calculer la variabilité de l'estimation d'un sous-échantillon à l'autre. Les utilisateurs désireux de calculer le coefficient de variation d'estimations quantitatives peuvent demander conseil à Statistique Canada en ce qui concerne la manière de répartir les enregistrements en sous-échantillons appropriés et les formules à utiliser pour ces calculs.

10.8 Seuils pour la diffusion des estimations relatives à l'EUEM

La taille minimum des estimations aux niveaux provincial, régional et national est indiquée dans le tableau ci-dessous. Les estimations correspondant à des tailles inférieures aux tailles minimales figurant dans la colonne «Confidentielles» ne peuvent être diffusées dans aucun cas.

PROVINCE	PUBLIABLE	PUBLIABLE SOUS RÉSERVE	CONFIDENTIELLES
Terre-Neuve	21,500	10,000	6,000
Île-du-Prince-Édouard	4,500	2,000	1,000
Nouvelle-Écosse	22,500	10,500	6,000
Nouveau Brunswick	18,000	8,000	4,500
Québec	295,500	137,500	79,000
Ontario	251,000	113,500	65,000
Manitoba	24,500	11,000	6,500
Saskatchewan	17,500	8,000	4,500
Alberta	100,000	46,500	26,500
Colombie-Britannique	182,000	86,000	50,000
Provinces Atlantiques	20,500	9,000	5,000
Provinces Prairies	41,000	18,000	10,500
CANADA	**204,000**	**90,000**	**51,000**

10.9 Tables de c.v.

Les tables de variabilité d'échantillonnage approximative pour chaque province et pour le Canada figurent sur les pages suivantes.

11. PONDÉRATION

Étant donné que l'EUEM a utilisé un sous-échantillon de l'EPA, le calcul des poids pour les enregistrements de l'enquête est évidemment lié à la procédure de pondération employée pour l'EPA. Voici une brève description de la procédure de pondération de l'EPA.

11.1 Pondération de l'EPA

Dans l'EPA, le poids final attribué à chaque enregistrement est le produit des facteurs suivants : le poids de base, le sous-poids de grappe, le facteur de compensation de la non-réponse, le facteur rural-urbain et le facteur d'ajustement du rapport âge-sexe-province. Voici une description de chacun de ces facteurs.

Poids de base

Dans un échantillon probabiliste, les poids à utiliser pour obtenir des estimations non biaisées de la population sont déterminés par le plan d'échantillonnage. Chaque enregistrement doit être pondéré par l'inverse de la probabilité de choisir la personne à laquelle se rapporte l'enregistrement. Dans le cas d'un échantillon aléatoire simple de 2%, cette probabilité serait de .02 pour chaque personne et les données dans chaque enregistrement doivent être pondérées par 1/.02=50. Comme on interroge toutes les personnes admissibles d'un logement (directement ou par personne interposée), cette probabilité est essentiellement la même que celle qui s'applique à la sélection du logement.

Sous-poids de grappe

La délimitation des grappes se fait de manière telle que le nombre de logements dans l'échantillon augmente très légèrement lorsque le parc immobilier s'accroît modérément. On peut tolérer une croissance importante dans une grappe isolée avant que l'échantillon supplémentaire ne pose des difficultés pour la collecte des données sur le terrain. Cependant, s'il y a une croissance pour plusieurs grappes dans la tâche d'un intervieweur, son effet cumulatif peut engendrer une surcharge de travail. Dans les grappes où il s'est produit une croissance importante, on procède à un sous-échantillonnage afin que le travail confié aux intervieweurs demeure raisonnable. Le sous-poids de grappe représente l'inverse du rapport de sous-échantillonnage dans les grappes où il y a eu sous-échantillonnage.

Non-réponse

En dépit des mécanismes de contrôle stricts employés pour l'EPA, un certain taux de non-réponse est inévitable malgré les multiples tentatives des intervieweurs. Le taux de non-réponse à l'EPA est d'environ 5%. Pour certains types de non-réponses (ménage temporairement absent, refus), si l'on a recueilli des renseignements sur le ménage lors d'une interview réalisée au cours d'un mois précédent, ces données sont reportées et utilisées pour le mois en cours.

Dans d'autres cas, on compense la non-réponse en augmentant proportionnellement le poids des ménages répondants. On multiplie le poids de chaque enregistrement de réponse par le rapport entre le nombre de ménages qui auraient dû être interviewés et le nombre de ménages qui l'ont effectivement été. Cette correction se fait séparément pour des régions géographiques qu'on appelle unités de compensation. Elle repose sur l'hypothèse que les ménages qui ont été interviewés représentent les caractéristiques de ceux qui auraient dû l'être. Les estimations seront légèrement biaisées dans la mesure où cette hypothèse s'avère fausse.

Facteur rural-urbain

Dans les UNAR où la population rurale et urbaine n'est pas suffisante pour pouvoir constituer des strates rurales et urbaines bien délimitées, chaque unité primaire d'échantillonnage (UPÉ) est composée d'une partie rurale et d'une partie urbaine. Le recensement de 1981 fournit des renseignements sur la population totale des secteurs rural et urbain de chaque UPÉ ainsi que sur chacune des régions économiques (RÉ) où il n'existe pas de stratification clairement établie en secteur urbain et secteur rural. Une comparaison par RÉ avec les chiffres ruraux ou urbains effectifs du recensement de 1981 indique si ces secteurs respectifs sont sur ou sous-représentés par les UPÉ sélectionnées. On divise le rapport des chiffres ruraux-urbains effectifs par les estimations correspondantes. Ces deux facteurs sont calculés pour chacune des régions économiques (RÉ) concernées au moment de la sélection des UPÉ et on les introduit dans chacun des enregistrements tirés de l'échantillon pour le secteur approprié (rural ou urbain) de l'UNAR. Les modifications à ces facteurs sont incorporées au moment du renouvellement des UPÉ.

Ajustements au niveau infra-provincial et au rapport âge-sexe-province

En appliquant les facteurs de pondération que nous venons de décrire (sous-poids), on peut calculer une estimation valable pour toutes les caractéristiques sur lesquelles des renseignements sont recueillis par l'EPA. On produit, en particulier, des estimations du nombre total de personnes de 15 ans et plus résidant dans les régions économiques provinciales et dans les 24 grandes régions métropolitaines ainsi que de groupes d'âge-sexe désignés dans chacune des dix provinces.

Des estimations distinctes sont disponibles, chaque mois, pour divers groupes d'âge-sexe par province. Il s'agit de projections démographiques fondées sur les données les plus récentes du recensement, sur les derniers enregistrements des naissances et des décès et sur les dernières estimations de la migration. Dans la dernière étape, on utilise ces renseignements auxiliaires pour transformer le sous-poids en poids final. Pour ce faire, on a recours à un modèle de régression linéaire. Ce modèle est structuré de sorte que la somme des poids finals produits soit égale aux projections du recensement pour les variables auxiliaires, nommément divers groupes d'âge-sexe, diverses régions économiques et diverses régions métropolitaines de recensement.

Cette procédure de pondération permet d'assurer la cohérence avec des chiffres externes du recensement, elle assure aussi que l'on attribue le même poids à chaque membre de la famille économique.

11.2 Pondération pour l'Enquête sur l'utilisation de l'énergie par les ménages

Les principes régissant le calcul des poids pour l'EUEM et pour l'EPA sont les mêmes. Toutefois, deux ajustements sont apportés aux poids de l'EPA afin de calculer un poids final pour chacun des enregistrements dans le fichier de microdonnées de l'EUEM.

1) Un ajustement pour tenir compte de la non-réponse additionnelle à l'enquête supplémentaire, c.-à-d. la non-réponse à l'EUEM dans le cas des ménages qui ont effectivement répondu à l'EPA ou dont les données de l'EPA du mois précédent ont été reportées.

2) Un ajustement pour tenir compte des projections indépendantes par genre de logement au niveau de la province, une fois que l'ajustement ci-dessus a été apporté. Ce nouvel ajustement corrige également pour l'utilisation d'un sous-échantillon de l'échantillon complet de l'EPA.

On tient compte de l'ajustement (1) en multipliant le sous-poids de l'EPA pour chaque enregistrement correspondant à un répondant à l'EUEM par :

$$\frac{\textit{la somme des sous-poids de l'EPA pour les personnes admissibles à l'EUEM}}{\textit{la somme des sous-poids de l'EPA pour les personnes ayant répondu à l'EUEM}}$$

afin d'obtenir un sous-poids corrigé pour la non-réponse de l'EUEM (POIDS1).

On calcule l'ajustement (2) en multipliant POIDS1 pour chaque ménage répondant à l'EUEM par :

$$\frac{\textit{nombre de logements dans la province-genre i}}{\textit{la somme des POIDS1 pour les logements ayant répondu dans la province-genre i}}$$

En pratique, les genres de logement ont été combinés de la manière suivante en vue de l'ajustement (2) : maisons individuelles (genre=1), appartements (genre=5 ou 6) et autres (genre=2, 3, 4 ou 9); à Terre-Neuve cependant, seules deux catégories ont été utilisées : maisons individuelles et autres (y compris les appartements).

Le poids résultant est le poids final qui est enregistré dans le fichier de microdonnées de l'Enquête sur l'utilisation de l'énergie par les ménages.

12. CLICHÉ D'ARTICLE ET VARIABLES UNIQUES

Les utilisateurs doivent savoir que les questions ne s'appliquent pas toutes au même nombre de répondants. Le questionnaire va aider à déterminer l'enchaînement exact des questions.

13. QUESTIONNAIRES ET FEUILLES DES CODES

1) Dossier du ménage (formulaire 03) et Feuilles des codes
2) Questionnaire de l'Enquête sur la population active (formulaire 05) et Feuille des codes
3) Questionnaire sur le loyer (formulaire 04)
4) Questionnaire de l'Enquête de 1993 sur l'utilisation de l'énergie par les ménages (formulaire 08)