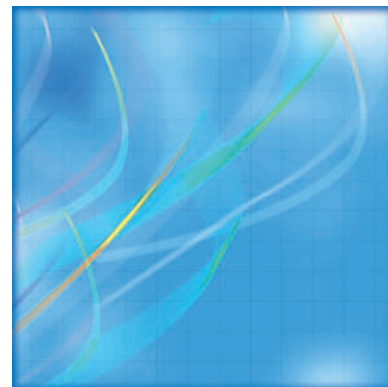


N° 92-568-X au catalogue

Rapport technique du Recensement de 2006 : Échantillonnage et pondération



Année de recensement 2006

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca. Vous pouvez également communiquer avec nous par courriel à infostats@statcan.gc.ca ou par téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

Centre de contact national de Statistique Canada

Numéros sans frais (Canada et États-Unis) :

Service de renseignements	1-800-263-1136
Service national d'appareils de télécommunications pour les malentendants	1-800-363-7629
Télécopieur	1-877-287-4369

Appels locaux ou internationaux :

Service de renseignements	1-613-951-8116
Télécopieur	1-613-951-0581

Programme des services de dépôt

Service de renseignements	1-800-635-7943
Télécopieur	1-800-565-7757

Comment accéder à ce produit

Le produit n° 92-568-X au catalogue est disponible gratuitement sous format électronique. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.gc.ca et de choisir la rubrique « Publications ».

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « À propos de nous » > « Offrir des services aux Canadiens ».

Rapport technique du Recensement de 2006 : Échantillonnage et pondération

Année de recensement 2006

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2009

Tous droits réservés. Le contenu de la présente publication électronique peut être reproduit en tout ou en partie, et par quelque moyen que ce soit, sans autre permission de Statistique Canada, sous réserve que la reproduction soit effectuée uniquement à des fins d'étude privée, de recherche, de critique, de compte rendu ou en vue d'en préparer un résumé destiné aux journaux et/ou à des fins non commerciales. Statistique Canada doit être cité comme suit : Source (ou « Adapté de », s'il y a lieu) : Statistique Canada, année de publication, nom du produit, numéro au catalogue, volume et numéro, période de référence et page(s). Autrement, il est interdit de reproduire le contenu de la présente publication, ou de l'emmagasiner dans un système d'extraction, ou de le transmettre sous quelque forme ou par quelque moyen que ce soit, reproduction électronique, mécanique, photographique, pour quelque fin que ce soit, sans l'autorisation écrite préalable des Services d'octroi de licences, Division des services à la clientèle, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Août 2009

N° 92-568-X au catalogue

ISBN 978-1-100-90963-9

Périodicité : hors-série

Ottawa

This publication is also available in English.

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Échantillonnage et pondération

Table des matières

	Page
Introduction.....	8
1 Collecte des données du recensement	10
1.1 Généralités	10
1.2 Méthodes de collecte.....	11
2 Traitement des données du recensement.....	12
2.1 Introduction.....	12
2.2 Réception et enregistrement	12
2.3 Imagerie et saisie des données des questionnaires papier	13
2.4 Contrôles de couverture	13
2.5 Suivi des contrôles d'intégralité et des rejets au contrôle	13
2.6 Codage	14
2.7 Classification et ajustements pour la non-réponse des logements inoccupés et des logements non répondants.....	14
2.8 Contrôle et imputation	15
2.9 Pondération	15
3 Échantillonnage dans les recensements canadiens.....	17
3.1 Historique de l'échantillonnage au recensement canadien.....	17
3.2 Plan d'échantillonnage utilisé lors du Recensement de 2006.....	18
4 Estimation basée sur l'échantillon du recensement.....	20
4.1 Considérations d'ordre opérationnel	19
4.2 Considérations d'ordre théorique	20
4.3 Élaboration d'une méthode d'estimation pour l'échantillon du recensement.....	21
4.4 Estimateur de régression pseudo-optimale en deux étapes.....	22
4.5 Traitement à deux passages	26
4.6 Différences entre les chiffres de population et les estimations pondérées finales	27
4.7 Les différents univers	27
5 Programme d'évaluation de l'échantillonnage et de la pondération.....	29
5.1 Biais d'échantillonnage.....	29
5.2 Évaluation des méthodes de pondération	29
5.3 Cohérence entre les estimations-échantillon et les chiffres de population	30
5.4 Variance d'échantillonnage	29

Table des matières (suite)

	Page
6 Biais d'échantillonnage.....	31
7 Évaluation des méthodes de pondération.....	39
7.1 Formation des régions de pondération (RP).....	39
7.2 Évaluation de la méthode de pondération.....	43
7.2.1 Répartition des poids.....	43
7.2.2 Écarts entre les chiffres de population et les estimations-échantillon	47
7.2.3 Retranchement des contraintes	58
8 Cohérence entre les estimations-échantillon et les chiffres de population.....	63
8.1 Aires de diffusion	64
8.2 Régions de pondération	64
8.3 Subdivisions de recensement	64
8.4 Secteurs de recensement	65
8.5 Divisions de recensement	65
9 Variance d'échantillonnage.....	73
10 Conclusion.....	86
Annexe A Glossaire	87
Annexe B Contraintes au niveau des RP et des AD appliquées aux poids des recensements de 2006 et de 2001.....	89
Annexe C Statistiques utilisées dans l'étude du biais d'échantillonnage.....	90
Bibliographie.....	90
Centre de contact national de Statistique Canada.....	91

Table des matières (suite)

Liste des tableaux

	Page
6.1 Différences population/estimation fondées sur les poids initiaux, recensements de 2006 et de 2001	33
7.1.1 Répartition des régions de pondération selon leur taille	40
7.1.2 Nombre de subdivisions de recensement et de secteurs de recensement qui respectent les limites des régions de pondération, Recensement de 2006	41
7.2.2.1 Comparaison des écarts population/estimation des recensements de 2001 et de 2006, pour le Canada	51
7.2.2.2 Comparaison des écarts population/estimation entre le premier passage et le deuxième passage fondés sur les poids finaux, pour le Canada, Recensement de 2006	55
7.2.2.3 Comparaison des univers – Chiffres de population et estimations, Recensement de 2006	57
7.2.3.1 Fréquence à laquelle des contraintes au niveau des régions de pondération ont été retranchées en 2001 et en 2006 lors de l’ajustement des poids finaux	61
7.2.3.2 Fréquence à laquelle des contraintes au niveau des régions de pondération et des aires de diffusion ont été retranchées en 2001 et en 2006 lors de l’ajustement des poids finaux – Statistiques sommaires	62
8.2.1 Centiles des écarts population/estimation pour les régions de pondération	68
8.5.1 Centiles des écarts population/estimation pour les divisions de recensement	71
9.1 Estimations non ajustées des erreurs types des estimations-échantillon	77
9.2 Facteurs d’ajustement de l’erreur type au niveau national ou provincial et des régions de pondération	78

Table des matières (suite)

Liste des graphiques

	Page
6.1 Statistiques Z des différences population/estimation fondées sur les poids initiaux, pour le Canada, recensements de 2006 et de 2001	36
6.2 Statistiques Z selon les régions en 2006	36
6.3 Statistiques Z selon les régions en 2001	37
7.2.1.1 Comparaison des poids finaux associés aux ménages des recensements de 2006 et de 2001	43
7.2.1.2 Comparaison des poids initiaux et des poids stratifiés « <i>a posteriori</i> », Recensement de 2006.....	44
7.2.1.3 Comparaison des poids stratifiés « <i>a posteriori</i> » et des poids obtenus à la première étape, Recensement de 2006	45
7.2.1.4 Comparaison des poids obtenus à la première étape et des poids finaux, Recensement de 2006	46
7.2.2.1 Comparaison des écarts entre les poids initiaux avec et sans imputation de ménages au complet	51
7.2.2.2 Écarts population/estimation fondés sur les poids finaux.....	53
7.2.2.2-A Écarts population/estimation fondés sur les poids finaux (agrandissement)	54
7.2.2.3 Comparaison des écarts population/estimation entre le premier passage et le deuxième passage fondés sur les poids finaux, pour le Canada, Recensement de 2006.....	55
7.2.2.4 Comparaison des écarts population/estimation, différences entre le premier passage et le deuxième passage, recensements de 2006 et de 2001	56
8.1.1 Centiles des écarts population/estimation pour les aires de diffusion de 2006 et de 2001 (groupes d'âge).....	66
8.1.2 Centiles des écarts population/estimation pour les aires de diffusion de 2006 et de 2001 (autres caractéristiques)	67
8.3.1 Centiles des écarts population/estimation pour les subdivisions de recensement (groupes d'âge).....	68
8.3.2 Centiles des écarts population/estimation pour les subdivisions de recensement (autres caractéristiques)	69
8.4.1 Centiles des écarts population/estimation pour les secteurs de recensement	70

Introduction

Le Recensement de 2006 a nécessité la participation de l'ensemble de la population canadienne, soit plus de 31 millions de personnes réparties sur un territoire couvrant 9 millions de kilomètres carrés. Même si des normes de qualité rigoureuses régissent la collecte et le traitement des données, il est impossible de supprimer toutes les erreurs. Afin d'aider les utilisateurs à évaluer l'utilité des données du recensement en fonction de leurs besoins, la série des Rapports techniques du Recensement de 2006 explique le cadre théorique et les définitions ayant servi à mener le recensement, de même que les méthodes de collecte des données et les méthodes de traitement employées. De plus, elle couvre les principales sources d'erreur, y compris, dans la mesure du possible, la taille de ces erreurs, ainsi que les circonstances inhabituelles pouvant limiter l'utilité ou l'interprétation des données du recensement. À l'aide de ces renseignements, les utilisateurs peuvent évaluer les risques entourant des conclusions ou des décisions fondées sur les données du recensement.

Le présent *Rapport technique du Recensement de 2006* porte sur la méthode d'échantillonnage et de pondération utilisée au Recensement de 2006 et sur l'incidence qu'elle a eue sur les résultats. Comme certaines données sont recueillies au moyen d'un échantillon et sont pondérées à l'échelle de la population totale, un biais et des divergences peuvent être observés dans les estimations définitives. Le présent rapport identifie ces différences observées et en explique les causes probables. Ce rapport a été préparé par Wesley Benjamin, Darryl Janes et Mike Bankier, avec le concours du personnel de deux divisions de Statistique Canada : la Division des méthodes d'enquêtes sociales et la Division des opérations du recensement.

De nos jours, l'échantillonnage constitue une technique largement utilisée dans maints domaines. Nous pouvons juger de la qualité d'un produit du marché à l'aide d'un échantillon avant d'en faire l'achat; nous formons une opinion des gens à partir d'échantillons de leur comportement; nos impressions sur tel pays ou telle ville sont basées sur les courts voyages que nous y avons faits. Ce sont là des exemples d'échantillonnage où l'on tire des conclusions sur le « tout » à partir de connaissances portant sur une « partie ».

D'une façon plus scientifique, l'échantillonnage est utilisé par les comptables, par exemple, lorsqu'ils effectuent la vérification des états financiers; dans l'industrie, pour vérifier la qualité d'articles produits en série, et par les initiateurs de sondages d'opinion et d'enquêtes pour recueillir des renseignements sur les opinions ou les caractéristiques d'une population donnée. On a généralement recours à l'échantillonnage pour réduire les coûts ou pour obtenir rapidement des données, ou les deux. Dans certains cas, le mesurage peut détruire le produit à l'étude (p. ex., l'évaluation de la durée de vie des ampoules électriques) et par conséquent, l'échantillonnage s'avère essentiel. L'un des inconvénients de l'échantillonnage est que les résultats obtenus à partir d'un échantillon peuvent ne pas être aussi précis que ceux recueillis auprès de la population totale. Toutefois, l'échantillonnage est souvent avantageux quand la perte de précision (qui peut être minime si l'échantillon est suffisamment grand) est acceptable en fonction de l'utilisation prévue des résultats.

Pour le Recensement de la population de 2006, on a appliqué les méthodes d'échantillonnage de plusieurs façons. On y a eu recours pour s'assurer que la qualité du travail des agents recenseurs dans le cadre de la collecte des questionnaires respectait certaines normes; on y a eu recours pour le contrôle qualitatif du codage des réponses au cours du dépouillement; on y a eu recours pour estimer le sous-dénombrement et le surdénombrement; on y a eu recours pour évaluer la qualité des données du recensement. Toutefois, la principale application de l'échantillonnage au recensement a été effectuée pendant le dénombrement sur le terrain où toutes les données, excepté les données de base, ont été recueillies auprès d'un échantillon de ménages. Le présent rapport décrit cette dernière utilisation de l'échantillonnage et présente une évaluation de l'incidence de cette méthode sur la qualité des données du recensement.

Les [chapitres 1 et 2](#) décrivent les procédures utilisées pour la collecte et le traitement des données. Le [chapitre 3](#) fait l'historique de l'application des méthodes d'échantillonnage aux recensements canadiens et décrit celles qui ont été utilisées au Recensement de 2006. Le [chapitre 4](#) décrit les méthodes utilisées pour pondérer les données-échantillon en vue d'obtenir des estimations à l'échelle de la population totale et explique les considérations d'ordre pratique et théorique qui régissent le choix de ces méthodes. On présente au [chapitre 5](#) un aperçu des études effectuées dans le cadre du programme d'évaluation des méthodes d'échantillonnage et de pondération utilisées au Recensement de 2006. Les [chapitres 6, 7, 8 et 9](#) donnent les résultats de ces études. Le [chapitre 10](#) présente certaines conclusions sur les méthodes de pondération utilisées en 2006.

Les utilisateurs trouveront des renseignements supplémentaires sur les concepts, les variables et la géographie du recensement, de même qu'un aperçu des différentes étapes du recensement, dans le [Dictionnaire du Recensement de 2006 \(produit n° 92-566-XWF au catalogue\)](#).

1 Collecte des données du recensement

1.1 Généralités

L'étape de la collecte des données du processus du Recensement de 2006 visait à faire en sorte que chacun des 13,5 millions de logements du Canada ait reçu un questionnaire du recensement. Le recensement visait l'ensemble de la population du Canada, qui comprend les citoyens canadiens (de naissance ou par naturalisation), les immigrants reçus et les résidents non permanents, ainsi que les membres de leur famille vivant sous le même toit. Les résidents non permanents sont les personnes qui vivent au Canada et qui sont titulaires d'un permis de travail ou d'études, ou qui revendiquent le statut de réfugié, ainsi que les membres de leur famille qui vivent avec eux.

Le recensement dénombrait également les citoyens canadiens et les immigrants reçus qui étaient temporairement à l'extérieur du pays le jour du recensement, y compris les fonctionnaires fédéraux et provinciaux en poste à l'extérieur du Canada, le personnel des ambassades canadiennes à l'étranger, les membres des Forces canadiennes en poste à l'étranger et les Canadiens qui se trouvaient à bord de navires marchands.

Au recensement du Canada, divers questionnaires et formules sont utilisés pour recueillir les données. Les formules suivantes sont mentionnées dans le présent rapport.

La formule 1 est le Registre des visites (RV) qui sert à dresser la liste des logements privés et collectifs occupés et inoccupés ainsi que des exploitants et exploitations agricoles qui se trouvent dans l'unité de collecte. Le RV sert de liste d'adresses aux fins du contrôle et des opérations sur le terrain dans le cadre de la collecte des données.

La formule 2A est le questionnaire abrégé de base, qui est distribué à quatre ménages sur cinq. La formule 2B est le questionnaire complet qui renferme, outre les questions figurant sur la formule 2A, des questions permettant de recueillir des données sur divers sujets. Le questionnaire 2B est remis à un ménage sur cinq. Chaque ménage qui reçoit un questionnaire du recensement 2A ou 2B est invité à y dénombrer tous les membres du ménage qui sont inclus dans la population du recensement et à répondre aux questions pour eux.

La formule 2C sert principalement à dénombrer les Canadiens en poste à l'étranger, notamment les fonctionnaires (fédéraux et provinciaux) et les membres de leur famille, ainsi que les membres des Forces canadiennes et les membres de leur famille. La formule 2C comporte les mêmes questions que le questionnaire 2B, sans toutefois compter les questions relatives au logement.

La formule 2D contient les mêmes questions que la formule 2B, mais elle sert à dénombrer la population des régions du Nord et de la plupart des réserves indiennes, des établissements indiens, des « *Indian government districts* » et des terres réservées. Dans les secteurs de recensement par interview, elle sert aussi à dénombrer les résidents habituels d'une colonie huttérite.

La formule 3 est un questionnaire de recensement individuel qui sert à dénombrer les personnes dans les logements collectifs (chaque résident d'un logement collectif doit remplir sa propre formule 3). Elle peut également être utilisée pour dénombrer des résidents habituels d'un ménage privé qui veulent remplir leur propre questionnaire du recensement plutôt qu'être inscrits sur un questionnaire 2A ou 2B. La formule 3A est la version abrégée du questionnaire, tandis que la formule 3B est la version intégrale.

1.2 Méthodes de collecte

Pour assurer la meilleure couverture possible, le Canada a été divisé en petits secteurs géographiques appelés unités de collecte (UC). Au Recensement de 2006, on comptait environ 50 000 unités de collecte.

Environ 98 % des ménages ont été recensés selon la méthode de l'autodénombrement. À compter du 2 mai, Postes Canada a livré un questionnaire du recensement à environ 70 % des ménages, les 30 % restants ayant obtenu leur questionnaire d'un agent recenseur. Les responsables du ménage ont été invités à remplir le questionnaire pour tous les membres du ménage, puis à le retourner par Internet ou dans l'enveloppe-réponse jaune port payé au plus tard le jour du recensement, le 16 mai.

Par ailleurs, environ 2 % des ménages ont été dénombrés par interview. Un agent recenseur a rendu visite au ménage et a rempli lui-même par interview sur place un questionnaire pour le ménage. Cette méthode était normalement utilisée dans les régions éloignées et nordiques du pays, de même que dans la plupart des réserves indiennes. Elle est en outre utilisée dans les grandes régions urbaines, où les répondants sont de passage.

Pour la première fois, le Recensement de 2006 a offert à tous les ménages du Canada la possibilité de remplir leur questionnaire par Internet. Chaque questionnaire papier était doté d'un code d'accès Internet unique indiqué sur la page couverture, ainsi que de l'adresse du site Internet du Recensement de 2006 (www.recensement2006.ca). Les répondants avaient besoin de ce code d'accès pour remplir leur questionnaire en ligne. Lorsque le questionnaire était rempli et retourné par Internet, on entrait directement les renseignements dans le système du Centre de traitement des données et on les vérifiait pour en assurer l'intégralité. Environ 18 % de la population a répondu au questionnaire par Internet.

Certains ménages ont été dénombrés par l'entremise de l'Assistance téléphonique du recensement (ATR), un service multilingue offert gratuitement à tous les répondants à l'échelle du pays. L'Assistance téléphonique du recensement a permis de recueillir les données du recensement au moyen d'un système d'interview téléphonique assistée par ordinateur (ITAO). Les agents recenseurs ont communiqué avec chacun des ménages dont on n'avait pas reçu de questionnaire dans un délai acceptable afin de recueillir leurs renseignements. Les agents recenseurs ont également eu recours à la méthode des ITAO pour communiquer avec les ménages en vue d'obtenir les réponses manquantes de leur questionnaire.

2 Traitement des données du recensement

2.1 Introduction

Cette étape du processus du recensement consiste à assurer le dépouillement de tous les questionnaires remplis, depuis la saisie des données jusqu'à la constitution d'une base de données du recensement exacte et complète. Il s'agit ici de l'enregistrement des questionnaires, de la saisie des données, de l'imagerie des questionnaires, du contrôle, de la correction des erreurs, du codage, de l'imputation et de la pondération. Dans le reste du présent chapitre, chaque opération de traitement des données est passée en revue.

Les processus automatisés, mis en œuvre en vue du Recensement de 2006, ont dû être surveillés de manière à ce que toutes les résidences canadiennes soient dénombrées une seule fois. Le Système de contrôle principal a été conçu pour contrôler et pour surveiller l'enchaînement des opérations. Le Système de contrôle principal était doté d'une liste principale de tous les logements du Canada (chaque logement était associé à un code d'identification unique, et environ les deux tiers des logements étaient également associés à une adresse). Ce système était mis à jour quotidiennement au moyen des renseignements sur la situation de chaque logement dans l'enchaînement des opérations du recensement (c.-à-d. livré, reçu, traité, etc.). Les rapports étaient produits et mis à la disposition des chefs du recensement en ligne, afin de que les opérations se déroulent de façon efficace et efficiente.

2.2 Réception et enregistrement

Les répondants qui avaient rempli un questionnaire papier le renvoyaient par la poste à un centre de traitement des données centralisé. Postes Canada enregistrait leur réception automatiquement en lisant le code à barres sur la page couverture du questionnaire, dans la partie transparente de l'enveloppe-réponse. Les enveloppes étaient ensuite transportées au Centre de traitement des données, accompagnées d'un disque compact renfermant la liste de tous les identificateurs pour les questionnaires inscrits.

Les réponses reçues par Internet ou par l'entremise d'une interview téléphonique de l'Assistance téléphonique du recensement étaient reçues directement par le Centre de traitement des données, et leur réception était enregistrée automatiquement.

L'enregistrement de chaque questionnaire retourné était signalé dans le Système de contrôle principal à Statistique Canada. Environ 10 jours après le jour du recensement, une liste de tous les logements pour lesquels un questionnaire n'avait pas été reçu était produite par le Système de contrôle principal, avant d'être transmise aux opérations sur le terrain en vue d'un suivi. Les mises à jour de l'enregistrement étaient envoyées quotidiennement aux opérations sur le terrain pour éviter de devoir effectuer des suivis auprès des ménages qui avaient rempli leur questionnaire ultérieurement, que ce soit par téléphone ou par Internet.

2.3 Imagerie et saisie des données des questionnaires papier

Le Recensement de 2006 était le premier recensement du Canada à saisir les données au moyen de technologies de saisie automatisée plutôt que l'entrée des données à la main. Le processus de l'imagerie comportait cinq étapes :

- **Préparation des documents** : les questionnaires retournés par la poste étaient sortis des enveloppes, et le matériel superflu, comme les trombones et les agrafes, était enlevé en vue de la numérisation. Les formules présentées sous forme de livret étaient divisées en feuilles individuelles en supprimant le dos.
- **Numérisation** : la numérisation, au moyen de 18 numériseurs à haute vitesse, convertissait les images imprimées en images numériques.
- **Processus automatisé d'assurance de la qualité des images** : le système automatisé vérifiait la qualité de la numérisation. Les images qui n'étaient pas conformes à ce processus étaient signalées pour la renumérisation ou la saisie à partir du document sur papier.
- **Saisie de données automatisée** : les technologies de reconnaissance optique de marques et de reconnaissance optique de caractères étaient utilisées pour extraire des images des données des répondants. Lorsque les systèmes ne pouvaient pas reconnaître l'écriture avec une précision suffisante, les données étaient corrigées par un opérateur.
- **Vérification à la sortie** : une fois les questionnaires traités conformément aux étapes ci-dessus, les questionnaires papier étaient retirés du système. La vérification à la sortie est un processus d'assurance de la qualité qui a pour objet de veiller à ce que les images et les données saisies soient de qualité suffisante et que les questionnaires papier ne soient plus nécessaires pour le traitement subséquent. Les questionnaires qui avaient été marqués comme contenant des erreurs étaient retirés au contrôle et traités de nouveau au besoin.

2.4 Contrôles de couverture

À cette étape, plusieurs contrôles automatisés étaient effectués aux données des répondants. Ces contrôles avaient pour objet de détecter les cas où des personnes invalides pourraient avoir été créées, soit en raison d'une erreur du répondant ou d'une erreur de saisie des données. Par exemple, mentionnons les données entrées par erreur dans une colonne de personnes vides, les données barrées ou rayées saisies par erreur et les données fournies pour la même personne plus d'une fois, habituellement en raison de la réception de formules en double (p. ex., un homme a répondu au questionnaire par Internet, tandis que sa femme l'a rempli sur papier et l'a retourné par la poste). Les contrôles étaient également conçus pour détecter l'absence éventuelle de résidents habituels, lorsque les données ne sont pas fournies pour chaque membre du ménage énuméré au début du questionnaire.

Les données des questionnaires qui échouaient aux contrôles étaient transmises aux commis du traitement en vue d'une vérification. Un système interactif permettait aux commis d'examiner les données saisies et de les comparer avec l'image, si cette dernière était disponible (les questionnaires en ligne n'ont pas d'image). Les rejets au contrôle étaient résolus au moyen d'une suppression manuelle des personnes invalides ou en double et de l'ajout des personnes manquantes (c.-à-d. la création de dossiers personnels vierges), au besoin et lorsque nécessaire.

2.5 Suivi des contrôles d'intégralité et des rejets au contrôle

Après les contrôles de couverture, une autre série de contrôles automatisés était exécutée pour détecter les cas qui comportaient trop de réponses manquantes ou qui portaient à croire que les données n'avaient peut-être pas été fournies pour tous les résidents habituels du ménage. Les ménages rejetés à ces contrôles étaient transmis à l'Assistance téléphonique du recensement en vue d'un suivi. Un intervieweur téléphonait au répondant afin de résoudre les éventuels problèmes de couverture et d'obtenir les renseignements manquants au moyen d'un logiciel d'interview téléphonique assistée par ordinateur. Les données étaient alors renvoyées au Centre de traitement des données pour être réintégrées dans le système en vue d'un traitement subséquent.

2.6 Codage

Les questionnaires complets (2B, 2C, 2D et 3B) contenaient des questions dont les réponses pouvaient être cochées à partir d'une liste, ainsi que des questions qui nécessitaient une réponse écrite de la part du répondant dans les cases fournies à cette fin. Ces réponses écrites ont fait l'objet d'un codage automatisé pour associer à chacune un code numérique, au moyen des fichiers de référence, des ensembles de codes et des classifications types de Statistique Canada. Des fichiers de référence pour le processus de couplage automatisé ont été créés au moyen des réponses obtenues aux recensements précédents. Des codeurs et des spécialistes du domaine ayant reçu une formation particulière réglaient les cas où il était impossible d'assigner un code automatiquement. Voici les variables auxquelles s'appliquait le codage : Lien avec la Personne 1, Lieu de naissance, Citoyenneté, Langues non officielles, Langue parlée à la maison, Langue maternelle, Origine ethnique, Groupe de population, Bande indienne/Première nation, Lieu de résidence 1 an auparavant, Lieu de résidence 5 ans auparavant, Principal domaine d'études, Lieu des études, Lieu de naissance des parents, Langue de travail, Industrie, Profession et Lieu de travail.

Environ 37 millions de réponses écrites ont été codées à partir des questionnaires complets de 2006. En moyenne, environ 82 % d'entre elles ont été codées automatiquement.

Une fois que toutes les réponses associées à une variable avaient été codées, les données étaient acheminées à l'étape du contrôle et imputation.

2.7 Classification et ajustements pour la non-réponse des logements inoccupés et des logements non répondants

L'Enquête sur la classification des logements (ECL) a servi à estimer les taux d'erreur pour classer les logements des secteurs de collecte autodénombrés comme occupés ou inoccupés sur le terrain. Conformément à ces renseignements, des ajustements ont été apportés à la base de données du recensement. L'ECL a sélectionné au hasard un échantillon de 1 405 UC autodénombrées qui avaient été revisitées en juillet et en août 2006 afin de réévaluer le statut occupation/inoccupation le jour du recensement de chaque logement pour lequel aucune réponse n'avait été reçue. Selon l'ECL, 17,4 % des 934 564 logements considérés comme inoccupés étaient en fait occupés et 29,1 % des 366 527 logements non répondants qui avaient alors été considérés comme occupés ou dont le statut d'occupation était inconnu étaient en fait inoccupés. Des estimations fondées sur l'échantillon de l'ECL ont été utilisées pour rajuster le statut d'occupation pour les logements individuels, ce qui a donné lieu à une hausse de 3,6 % du nombre de logements occupés, et à une baisse de 5,2 % du nombre de logements inoccupés à l'échelle du Canada.

Après cet ajustement du statut d'occupation par l'ECL, on imputait le nombre de résidents habituels (s'il n'était pas connu) et toutes les réponses aux questions du recensement dans le cas des logements

occupés ayant une non-réponse totale, en empruntant les réponses non imputées d'un autre ménage de la même UC ayant reçu le même type de questionnaire (abrégé ou complet). Ce processus, appelé imputation de ménages au complet (IMC), a permis d'imputer 96 % des ménages ayant une non-réponse totale. La proportion restante de 4 % des ménages affichant une non-réponse totale pour lesquels aucun ménage donneur n'avait été trouvé conformément au processus d'IMC ont été imputés dans le cadre du processus principal de contrôle et d'imputation (CI). L'utilisation d'un seul donneur conformément au processus d'IMC était plus efficace du point de vue du traitement informatique et moins susceptible de produire des résultats non plausibles que le recours à plusieurs donneurs au moyen du processus principal de CI, comme on l'avait fait en 2001.

Le Rapport technique du recensement de 2006 sur la couverture, produit n° 92-567-XWF au catalogue, renferme des renseignements plus détaillés au sujet de l'ECL et de la procédure d'imputation du ménage au complet.

2.8 Contrôle et imputation

Les données recueillies lors de toute enquête ou recensement comportent des omissions ou des incohérences. Par exemple, un répondant peut être réticent à répondre à une question, peut ne pas se souvenir de la bonne réponse ou peut mal comprendre la question. Par ailleurs, le personnel du recensement peut coder les réponses incorrectement ou faire d'autres types d'erreurs au cours du traitement.

L'épuration définitive des données, faite à l'étape du contrôle et de l'imputation, a été presque entièrement automatisée. Deux types d'imputations ont été utilisés. Le premier type, appelé « imputation déterministe », consistait à attribuer des valeurs données dans certaines conditions. Des règles de contrôle détaillées étaient utilisées pour déterminer ces conditions, puis on attribuerait une valeur prédéterminée aux variables en cause dans les règles. Le deuxième type d'imputation, appelé « imputation donneur à changements minimaux », consistait à appliquer une série de règles de contrôle détaillées, qui permettaient de repérer les réponses manquantes ou incohérentes. Ces dernières étaient corrigées en modifiant par imputation le moins de variables possible. Pour l'« imputation par donneur à changements minimaux », on choisit un enregistrement qui comporte un certain nombre de caractéristiques communes à celles de l'enregistrement qui contient une erreur. On emprunte les données de cet enregistrement « donneur » et on les utilise pour changer le plus petit nombre possible de variables nécessaires pour résoudre tous les cas de réponses manquantes ou incohérentes. Le système canadien de contrôle et d'imputation du recensement (SCANCIR) était le système automatisé utilisé dans presque tous les cas d'imputation déterministe et d'imputation par donneur à changements minimaux en 2006.

2.9 Pondération

Des questions sur l'âge, le sexe, l'état matrimonial, la langue maternelle et le lien avec la Personne 1 ont été posées à 100 % de la population, comme dans les recensements précédents. Toutefois, la majorité des données du recensement ont été obtenues par le biais d'un échantillon de 20 %, au moyen des questions supplémentaires du questionnaire 2B. On a eu recours à la pondération pour extrapoler les renseignements recueillis auprès de l'échantillon de 20 % à l'ensemble de la population.

La pondération utilisée dans le cadre du Recensement de 2006 était la même que celle du Recensement de 2001, soit l'estimation par calibrage. D'abord, on attribuait des poids initiaux d'environ 5 aux ménages échantillonnés. Ces poids étaient ensuite ajustés par le plus petit nombre qui permette de faire correspondre d'aussi près que possible les estimations de l'échantillon et les chiffres de population pour plusieurs caractéristiques relatives à l'âge, au sexe, à l'état matrimonial, à l'union libre et à la taille du

ménage (p. ex., nombre de personnes de sexe masculin, nombre de personnes âgées de 15 à 19 ans). On trouve une description plus détaillée de cette procédure au [chapitre 4](#).

3 Échantillonnage dans les recensements canadiens

Dans un recensement de la population, l'échantillonnage consiste à recueillir et à traiter certaines caractéristiques à partir d'un échantillon aléatoire de logements et de personnes dénombrés lors du recensement intégral. On obtient ensuite pour l'ensemble de la population les totaux pour ces caractéristiques en pondérant les données-échantillon. Les caractéristiques recueillies pour tous les logements ou toutes les personnes faisant partie du recensement sont appelées « caractéristiques de base », tandis que les caractéristiques recueillies à partir d'un échantillon sont appelées « caractéristiques-échantillon ».

3.1 Historique de l'échantillonnage au recensement canadien

L'échantillonnage a été utilisé pour la première fois au cours du recensement canadien de 1941. Un questionnaire sur le logement a été remis à tous les 10 logements de chaque sous-district de recensement. Les réponses aux 27 questions de ce questionnaire ont été intégrées à celles du questionnaire sur la population et les ménages correspondant à ces logements, ce qui a permis de faire le recoupement des caractéristiques-échantillon et des caractéristiques de base. Lors de ce même recensement, on a utilisé l'échantillonnage au cours de l'étape du dépouillement pour obtenir des estimations provisoires du revenu des salariés, de la répartition de la population en âge de travailler et de la composition des familles au Canada. Dans ce cas, l'échantillon était constitué du dixième de tous les secteurs de dénombrement du Canada et tous les questionnaires sur la population de ces secteurs ont été dépouillés à l'avance.

L'échantillonnage a également été utilisé pour le Recensement du logement de 1951. Cette fois, on a choisi chaque cinquième logement (ceux dont le numéro d'identification se terminait par 2 ou par 7) et le questionnaire sur le logement comportait 24 questions. Au cours du Recensement de 1961, on a prélevé un échantillon de 20 % des ménages privés et on a demandé aux membres de 15 ans et plus de ces ménages de remplir un questionnaire-échantillon sur la population comportant des questions sur la migration interne, la fécondité et le revenu. L'échantillonnage n'a pas été utilisé au cours des recensements de moindre envergure de 1956 et de 1966.

Plusieurs innovations importantes ont été apportées aux méthodes de dénombrement lors du Recensement de 1971. La première a été l'utilisation de l'autodénombrement pour la majorité de la population au lieu de la méthode traditionnelle de recensement par interview. Ce changement découlait des résultats de plusieurs études effectuées au Canada et ailleurs (Fellegi, 1964; Hansen et al., 1959) selon lesquels la présence d'un agent recenseur avait un effet important sur la variance¹ de chiffres du recensement lors d'un recensement par interview. On prévoyait que l'autodénombrement permettrait de réduire la variance des chiffres du recensement attribuable à la présence de l'agent recenseur et, en même temps, donnerait au répondant le temps voulu pour remplir le questionnaire en toute confidentialité; par conséquent, on s'attendait à ce que les réponses soient plus exactes que par le passé.

Un second changement survenu au Recensement de 1971 a été la modification du contenu. Le nombre de sujets traités et de questions posées était plus élevé qu'aux recensements précédents. Une étude des coûts, du fardeau du répondant et des délais de production par rapport à la qualité des données dans le cadre d'un recensement par échantillonnage et par autodénombrement a présidé à la décision de recueillir toutes les données, sauf celles portant sur quelques caractéristiques de base, auprès d'un échantillon d'un tiers de la population au Recensement de 1971. Sauf dans les régions les plus isolées du Canada, le tiers des ménages privés a reçu le questionnaire complet comportant toutes les questions du recensement, tandis que le reste des ménages privés a reçu le questionnaire abrégé qui ne contenait que les questions

1. La « variance » d'une estimation est une mesure de sa précision. La variance est expliquée plus en détail au [chapitre 9](#).

de base soit le nom, le lien avec le chef de ménage, le sexe, la date de naissance, l'état matrimonial, la langue maternelle, le type de logement, le mode d'occupation, le nombre de pièces, l'alimentation en eau, les toilettes ainsi que certaines questions en rapport avec la couverture du recensement. Tous les ménages vivant dans des régions éloignées prédésignées et dans des logements collectifs² ont reçu le questionnaire complet. Le rapport statistique « Sampling in the Census » (Bureau fédéral de la statistique, 1968) donne une description détaillée des projets d'application de cette méthode au Recensement de 1971.

Le contenu du Recensement de 1976 était beaucoup moindre que celui du Recensement de 1971. En outre, le questionnaire de 1976 ne contenait pas les questions qui posent le plus de difficulté de collecte (p. ex., les questions sur le revenu) ni celles pour lesquelles les frais de codage sont le plus élevés (p. ex., celles sur la profession, l'industrie et le lieu de travail). Par conséquent, les avantages de l'échantillonnage en termes de coûts et d'allègement du fardeau du répondant étaient moins évidents qu'au Recensement de 1971. Néanmoins, après avoir évalué les économies éventuelles reliées à divers taux d'échantillonnage et étudié les conséquences sur le plan des relations publiques du retour au dénombrement intégral (l'échantillonnage ayant fait ses preuves en 1971), il a été décidé d'appliquer en 1976 la même méthode d'échantillonnage qu'en 1971.

Au Recensement de 1981, on a utilisé la plupart des mêmes méthodes qu'en 1971 et 1976, sauf que le taux d'échantillonnage est passé du tiers au cinquième des ménages. Les études réalisées à cette époque ont révélé que la diminution consécutive de la qualité des données (mesurée au moyen de la variance) serait acceptable et qu'elle ne serait pas suffisamment importante pour annuler les avantages découlant de la réduction des coûts du fardeau de réponse et des délais de production (voir Royce, 1983). L'échantillonnage du cinquième des ménages est la méthode retenue pour tous les recensements depuis 1981.

3.2 Plan d'échantillonnage utilisé lors du Recensement de 2006

Le 16 mai 2006, jour du recensement, une multitude de renseignements ont été recueillis auprès de chaque Canadien. La majeure partie de ces données ont été obtenues à partir d'un échantillon. Dans chaque secteur d'autodénombrement, un échantillon de un logement privé occupé sur cinq a été choisi pour recevoir un questionnaire complet (formule 2B) tandis qu'aux logements occupés ne faisant pas partie de l'échantillon, on distribuait un questionnaire abrégé (formule 2A). Les questions de base sur l'âge, le sexe, l'état matrimonial, l'union libre, la langue maternelle, le lien avec la Personne 1 ont été posées à tous les répondants de même que le type de logement. Des questions supplémentaires portant sur le logement et des questions socioéconomiques ont été posées à un échantillon de la population. On appelle « ménage » les résidents habituels d'un logement occupé. Ainsi, les termes « ménage » et « logement occupé » seront utilisés de façon interchangeable dans le présent rapport.

Tous les logements se trouvant dans les secteurs dénombrés par un intervieweur (généralement les régions éloignées ou les réserves indiennes) ont reçu le questionnaire 2D.

La plupart des personnes vivant dans des logements collectifs ont reçu un questionnaire complet (habituellement une formule 3B, sauf dans le cas des colonies huttérites et des unités composées de personnes âgées qui ont reçu une formule 2B). Toutefois, les résidents suivants des logements collectifs n'ont pas eu à répondre aux questions de l'échantillon (autrement dit, la formule 3A a été utilisée) :

a) les détenus dans les établissements de correction et les établissements pénitentiaires ou les prisons;

2. Un logement collectif est un établissement qui sert à des fins commerciales, institutionnelles ou communautaires, comme les hôtels, les hôpitaux et les camps de chantier.

- b) les patients dans les hôpitaux généraux, les établissements de soins spéciaux et les logements collectifs ou les établissements pour les personnes âgées ou les personnes souffrant de maladies chroniques ou les hôpitaux psychiatriques;
- c) les foyers collectifs pour enfants (orphelinats) et les foyers pour jeunes contrevenants;
- d) les personnes vivant dans un gîte.

On appelle « unité composée de personnes âgées » un logement au sein d'une résidence collective pour personnes âgées composé d'au moins une personne âgée considérée capable de remplir un questionnaire du recensement. Comme mentionné au point b) ci-dessus, ces personnes n'ont pas eu à répondre aux questions de l'échantillon dans le cadre du Recensement de 2001. Pour la première fois en 2006, une unité composée de personnes âgées sur cinq a été sélectionnée au hasard. Les membres de l'échantillon ainsi obtenu ont reçu un questionnaire complet à remplir. Les autres unités de personnes âgées devaient remplir un questionnaire abrégé. Ces unités composées de personnes âgées étaient considérées comme des logements privés échantillonnés par le système de pondération du Recensement de 2006 et figurent dans les divers tableaux du présent rapport. En 2006, on a dénombré 40 755 unités composées de 47 540 personnes âgées. Cependant, il convient de souligner qu'une unité composée de personnes âgées n'est pas un ménage ni un logement; en effet, c'est l'ensemble de la résidence pour personnes âgées qui constitue le ménage ou le logement. De plus, ces chiffres sont plus bas que les vrais chiffres à cause de problèmes durant la collecte et le traitement, où des résidences pour personnes âgées ont été classées par erreur comme étant des établissements de soins infirmiers. En conséquence, la qualité des données pour les unités composées de personnes âgées seulement est médiocre, et il faut les utiliser avec prudence.

La [section 1.2](#) présente les diverses méthodes de collecte des données du recensement. Chaque logement, peu importe la méthode de collecte, a été jumelé à un numéro de Registre des visites (RV). Le numéro de RV a permis de déterminer quel type de questionnaire de recensement (c.-à-d. la formule 2A ou 2B) serait livré. Un logement sur cinq a été sélectionné pour faire partie de l'échantillon du recensement et a reçu un questionnaire complet. Les quatre cinquièmes des logements restants ont reçu un questionnaire abrégé.

Dans la terminologie de l'échantillonnage, l'échantillon du recensement peut être décrit comme un échantillon stratifié avec sélection systématique de logements privés occupés, choisi en fonction d'un taux d'échantillonnage constant de 1 sur 5 dans toutes les strates (UC). En tant qu'échantillon de personnes, il peut être considéré comme un échantillon stratifié avec sélection systématique de grappes, un logement constituant une grappe. Pour une description détaillée des concepts et de la terminologie se rapportant à l'échantillonnage, voir Cochran (1977) ou Sarndal, Swensson et Wretman (1992).

4 Estimation basée sur l'échantillon du recensement

Tout échantillonnage nécessite l'application d'une méthode d'estimation pour rapporter les données-échantillon à l'échelle de la population totale. La méthode d'estimation est généralement choisie en fonction d'exigences d'ordre opérationnel et théorique. Sur le plan opérationnel, cette méthode doit être compatible avec le système de dépouillement auquel elle s'intègre, tandis que sur le plan théorique, elle doit minimiser l'erreur d'échantillonnage des estimations produites. Les sections 4.1 et 4.2 décrivent les considérations opérationnelles et théoriques qui régissent le choix des méthodes d'estimation à partir de l'échantillon du recensement. Les sections 4.3 et 4.4 présentent certaines des méthodes utilisées pour établir les poids du recensement. Les autres sections portent sur les univers de données utilisés dans le processus de pondération et expliquent brièvement pourquoi des écarts peuvent survenir entre les chiffres de population et les estimations pondérées.

4.1 Considérations d'ordre opérationnel

Une méthode d'estimation peut être décrite mathématiquement à l'aide d'une formule algébrique, ou d'un estimateur, qui indique comment l'estimation de la population est calculée en fonction des valeurs observées dans l'échantillon. Quand il s'agit de petites enquêtes qui ne portent que sur une ou deux caractéristiques, ou lorsque la formule d'estimation est très simple, il est possible de calculer les estimations-échantillon en appliquant cette formule aux données-échantillon pour chaque estimation requise. Toutefois, dans le cas d'une enquête ou d'un recensement portant sur une gamme étendue de caractéristiques, ou lorsque la formule d'estimation est assez complexe, il n'est pas possible d'appliquer une formule séparément pour chaque estimation désirée. Par exemple, une application distincte de la formule d'estimation serait nécessaire pour chaque cellule de chaque tableau du recensement publié en fonction des données de l'échantillon. Par ailleurs, les diverses estimations effectuées par calcul individuel à partir d'un même échantillon du recensement ne sont pas nécessairement cohérentes.

Par conséquent, la méthode utilisée pour le recensement (et pour de nombreuses enquêtes-échantillon) consiste à décomposer le processus d'estimation en deux étapes : a) le calcul des poids, c'est-à-dire la pondération elle-même et b) la sommation des poids afin de produire des estimations des chiffres de population. Les seules difficultés mathématiques, s'il y en a, se rapportent à l'étape a) qui n'est effectuée qu'une fois, tandis que l'étape b) se limite à un simple processus de sommation des poids au moment où une totalisation est extraite. Il faut noter que, comme le poids relié à une unité de l'échantillon est le même quelle que soit la totalisation effectuée, la cohérence entre les diverses estimations basées sur les données-échantillon est assurée.

4.2 Considérations d'ordre théorique

Selon la théorie de l'échantillonnage, il est possible de déterminer, à partir d'un plan d'échantillonnage particulier et d'une procédure d'estimation donnée, les chances qu'un certain intervalle contienne la valeur (inconnue) que l'on veut estimer pour la population entière. Le principal objet de l'estimation est de réduire au maximum l'étendue de ces intervalles de façon que les hypothèses sur les valeurs inconnues de la population soient aussi précises que possible. La mesure courante de précision utilisée pour comparer les procédures d'estimation est appelée l'erreur type. Pourvu que certaines conditions relativement peu contraignantes soient respectées, un intervalle de plus ou moins deux erreurs types par rapport à la valeur estimée renfermera la valeur de la population pour environ 95 % de tous les échantillons possibles.

Outre la réduction au minimum de l'erreur type, le deuxième objectif relatif au choix d'une méthode d'estimation de l'échantillon du recensement est de garantir autant que possible que les estimations faites à partir de l'échantillon des caractéristiques de base (c.-à-d. 2A) soient cohérentes avec les valeurs connues de la population correspondante. Heureusement, ces deux objectifs sont généralement complémentaires,

c'est-à-dire que l'erreur d'échantillonnage est habituellement réduite lorsqu'on s'assure que les estimations-échantillon de certaines caractéristiques de base sont compatibles avec les chiffres correspondants de population. Notons cependant que, bien que vrai en général, il peut arriver que le fait de forcer les estimations-échantillon de caractéristiques de base à être compatibles avec les chiffres correspondants de population pour de très petits sous-groupes puisse avoir des conséquences défavorables sur l'erreur type des estimations des caractéristiques-échantillon proprement dites.

Lorsqu'on n'a aucune information sur la population échantillonnée autre que celle qui a été recueillie pour les unités d'échantillonnage, la méthode d'estimation se limite à pondérer les unités d'échantillonnage en proportion inverse de leur probabilité de sélection; par exemple, si toutes les unités avaient une chance sur cinq d'être choisies, toutes les unités recevraient un poids de 5. En pratique, toutefois, on a généralement une certaine connaissance de la population observée, par exemple, sa taille globale et, peut-être, sa répartition en fonction d'une variable donnée (disons, par province). Ces données peuvent être utilisées pour améliorer la formule d'estimation de façon à produire des estimations qui se rapprochent davantage de la valeur inconnue de la population. Dans le cas de l'échantillon du recensement, on dispose d'un grand nombre de données très détaillées sur la population échantillonnée, soit les caractéristiques de base obtenues pour tous les ménages, à tous les niveaux géographiques. On peut se servir de cette multitude d'informations sur la population pour améliorer les estimations faites à partir de l'échantillon du recensement mais, d'autre part, ces données peuvent également occasionner des ennuis compte tenu du fait qu'il est impossible d'obtenir des estimations-échantillon des caractéristiques de base qui soient compatibles avec toutes les données sur la population à tous les niveaux géographiques. Les écarts entre les estimations-échantillon et les valeurs de la population apparaissent quand on produit un recoupement d'une variable de base et d'une variable-échantillon. La totalisation doit être fondée sur les données-échantillon; de ce fait, les totaux marginaux des variables de base sont des estimations-échantillon que l'on peut comparer avec les chiffres correspondants de population d'une autre totalisation basée sur les données intégrales. Il se peut que cette comparaison ne donne pas une concordance exacte. Ces différences sont abordées plus en détail à la [section 4.6](#) du présent rapport.

4.3 Élaboration d'une méthode d'estimation pour l'échantillon du recensement

Étant donné qu'il faut attribuer un poids à chaque unité d'échantillonnage (personne, famille ou ménage), le plus simple serait de choisir le poids 5, puisqu'il s'agit d'un échantillon de 1 sur 5. Une telle méthode serait simple et sans biais³, et si l'on n'avait pas d'autres données que celles de l'échantillon, cela pourrait être la meilleure méthode. Or, bien qu'on sache que l'échantillon contiendra presque exactement le cinquième de tous les logements (à l'exclusion des logements collectifs et de ceux qui se trouvent dans les secteurs de recensement par interview), on ne peut pas être certain qu'il contiendra exactement un cinquième de toutes les personnes, ou un cinquième de chaque genre de ménage, ou un cinquième de toutes les femmes âgées de 25 à 34 ans, etc. Par conséquent, une telle méthode n'assurerait pas la cohérence, même pour les sous-groupes les plus importants de la population. Dans ce dernier cas, les fractions seraient très proches d'un cinquième, mais elles pourraient s'en écarter de façon sensible dans le cas des petits sous-groupes. Une autre méthode simple serait de prendre certains sous-groupes importants, par exemple les groupes d'âge-sexe par province, puis, pour chacun, compter le nombre d'unités de la population appartenant au sous-groupe (N) et le nombre d'unités appartenant à l'échantillon (n), et attribuer à chaque unité d'échantillonnage appartenant au sous-groupe un poids égal à N/n. Ces sous-groupes sont souvent appelés strates « *a posteriori* ».

3. « Sans biais » veut dire que la moyenne des estimations de l'ensemble de tous les échantillons possibles serait égale à la valeur vraie de la population.

Par exemple, si l'on recensait 5 000 hommes âgés de 20 à 24 ans à l'Île-du-Prince-Édouard, et si 1 020 de ces hommes faisaient partie des logements inclus dans l'échantillon, un poids de $5\,000/1\,020 = 4,90$ serait attribué à chaque homme âgé de 20 à 24 ans dans l'échantillon de l'Île-du-Prince-Édouard. De cette façon, chaque fois que les groupes d'âge-sexe de cinq ans seraient recoupés avec une caractéristique-échantillon pour l'Île-du-Prince-Édouard, le total marginal pour le groupe d'âge-sexe d'hommes âgés de 20 à 24 ans concorderait avec le total de la population, soit 5 000. Ce genre de méthode d'estimation est appelée « estimation par la méthode du quotient ». Il est à noter que, dans cet exemple particulier, un poids simple de 5 donnerait une estimation-échantillon de 5 100 ($1\,020 \times 5$).

On appelle calibrage l'exercice qui consiste à ajuster les poids simples de 5 aussi peu que possible de manière à obtenir une concordance parfaite entre les estimations et les chiffres de population. Avant le Recensement de 1991, le calibrage se faisait grâce à la méthode itérative du quotient. Les estimations se rapportant aux ménages étaient produites à partir d'un poids calibré au niveau des ménages, alors que les estimations se rapportant aux personnes l'étaient à partir d'un poids calibré au niveau des personnes.

En 1991, on a adopté la méthode de la régression généralisée (estimateur GREG) en deux étapes, laquelle a permis d'obtenir une meilleure concordance entre les chiffres de population et les estimations correspondantes au niveau des secteurs de dénombrement (SD) qu'avec la méthode itérative du quotient. En outre, un seul poids calibré au niveau des ménages a pu servir à produire à la fois les estimations des ménages et celles des personnes. Par conséquent, on a pu éliminer les incohérences qui avaient été relevées dans certaines estimations d'avant 1991. L'estimateur GREG à deux étapes a également été utilisé en 1996.

En 2001 et en 2006, on a utilisé un estimateur de régression pseudo-optimale, car il permettait généralement d'obtenir une correspondance légèrement meilleure entre les chiffres et les estimations de population que l'estimateur GREG, tout en garantissant que les poids calibrés égalaient tous au moins 1. Voir Bankier (2002) pour obtenir une comparaison plus détaillée des estimateurs de régression.

Avec l'estimateur de régression pseudo-optimale, les poids initiaux d'environ 5 ont été ajustés aussi peu que possible pour chaque logement, de sorte qu'on s'assurait qu'il y avait une concordance parfaite entre les estimations et les chiffres de population pour autant de caractéristiques de base énoncées à l'[annexe B](#) que possible (celles-ci seront appelées contraintes ou variables auxiliaires). Il fallait que cette concordance parfaite soit réalisée au niveau des régions de pondération (RP). On retrouve plus d'information à propos des RP à la [section 7.1](#) du présent rapport.

En 2006, le Canada était divisé en environ 50 000 unités de collecte à utiliser pour recueillir les données du recensement. L'unité de collecte (UC) a une taille et des caractéristiques semblables à celles du secteur de dénombrement (SD), qui était utilisé avant le Recensement de 2006. Un échantillon systématique d'un logement sur cinq a été sélectionné dans la plupart des UC aux fins du processus de pondération du recensement. Les aires de diffusion (AD) sont un autre niveau géographique d'une taille semblable à celle des UC. Des AD complètes ont été combinées pour former des RP. En moyenne, on retrouve huit AD et sept UC échantillonnées dans une RP.

4.4 Estimateur de régression pseudo-optimale en deux étapes

Le processus de régression utilise 34 variables auxiliaires, notamment les groupes d'âge quinquennaux, l'état matrimonial, l'union libre, le sexe, la taille du ménage et le type de logement. Voir l'[annexe B](#) pour prendre connaissance des 34 variables auxiliaires. Voici les objectifs de la procédure de pondération du Recensement de 2006 :

- a) avoir une concordance **parfaite** entre les chiffres de population et les estimations au niveau des RP pour le plus grand nombre possible de variables parmi les 34 variables auxiliaires;
- b) avoir une concordance **approximative** entre les chiffres de population et les estimations au niveau des AD de plus grande taille pour les 34 variables auxiliaires.

En outre, les conditions suivantes doivent être remplies :

- c) Il doit y avoir une concordance **parfaite** entre les chiffres de population et les estimations pour le « nombre total de ménages » et le « nombre total de personnes » pour le plus grand nombre d'AD possible.
- d) Les poids de recensement finaux doivent se situer dans l'intervalle de 1 à 25 inclusivement. La limite inférieure a été fixée à 1 parce qu'on a estimé que chaque personne échantillonnée devait au moins se représenter elle-même.
- e) La méthode de calcul des poids doit être hautement automatisée, puisque les 6 602 RP comptant des ménages susceptibles d'être échantillonnés doivent être traités au cours d'une courte période. Cette méthode doit également s'ajuster automatiquement aux différents schémas de réponse dans les RP du pays.

Les poids sont calculés séparément dans chaque RP au moyen d'un système de pondération automatisé. Pour chaque RP traitée, un ensemble de paramètres définis par l'utilisateur est transmis au système. Un poids initial est attribué à chaque ménage privé échantillonné de la RP, et ces poids sont ensuite soumis à deux ou trois facteurs d'ajustement. D'abord, les ménages peuvent être stratifiés « *a posteriori* » au niveau de la RP selon la taille du ménage, puisque les ménages de petite et de grande taille sont sous-représentés dans l'échantillon. Un deuxième ajustement est ensuite apporté aux poids pour obtenir une concordance approximative entre les chiffres de population et les estimations au niveau des AD, telle que décrite à l'objectif b) ci-dessus. Enfin, un troisième ajustement est effectué afin d'obtenir une concordance parfaite entre les chiffres de population et les estimations au niveau des RP et des AD, telle que décrite aux objectifs a) et c) ci-dessus. Par souci de simplification, nous donnerons une description plus détaillée des poids initiaux et des trois ajustements avant d'aborder la question du retranchement de contraintes et des diverses raisons pour lesquelles des contraintes doivent être retranchées.

D'abord, un poids initial au niveau de l'UC est attribué à chacun des ménages privés de la RP. Le poids équivaut au nombre de ménages privés de l'UC, divisé par le nombre de ménages privés échantillonnés dans cette UC. Étant donné qu'environ un ménage sur cinq serait échantillonné, les poids initiaux ont tendance à se rapprocher de cinq. En 2001, les unités composées de personnes âgées ne faisaient pas partie du processus de pondération du recensement et étaient exclues du processus d'échantillonnage. Toutefois, étant donné qu'en 2006, les unités composées de personnes âgées étaient traitées de façon semblable aux ménages privés, elles faisaient partie de la base de sondage. Puisque la proportion d'unités composées de personnes âgées dans une UC était habituellement très faible, ce type d'unité avait généralement peu d'effet sur les résultats de la pondération. Cependant, pour un petit nombre d'UC comportant une forte proportion d'unités composées de personnes âgées, les ménages privés et les résidences pour personnes âgées étaient traités comme deux populations distinctes, ce qui fait que deux ensembles de poids initiaux ont été calculés pour chacune de ces UC, afin de réduire le biais d'échantillonnage. Une fois les poids initiaux créés, les unités composées de personnes âgées n'étaient pas traitées différemment que les ménages privés tout au long du reste du processus de pondération. Toutefois, lorsqu'on a calculé les facteurs d'ajustement de l'erreur type du [chapitre 9](#), dans le cas d'une UC

où les ménages privés et les unités composées de personnes âgées étaient traités comme deux populations distinctes, on les considérait comme deux strates d'échantillonnage plutôt qu'une.

À la première étape de l'ajustement, il arrive que les ménages soient stratifiés « *a posteriori* » en fonction de leur taille (1, 2, 3, 4, 5 ou 6 personnes ou plus) au niveau des RP. Les poids initiaux sont multipliés par un facteur de façon à obtenir les poids stratifiés « *a posteriori* ». Par exemple, selon les poids stratifiés « *a posteriori* », le nombre estimé de ménages de une personne dans une RP donnée correspondrait au nombre de ménages de une personne dans la population de cette RP. En de très rares occasions, il peut arriver qu'un poids stratifié « *a posteriori* » soit contraint afin qu'il se situe dans l'intervalle de 1 à 20 inclusivement. Une limite supérieure de 20 plutôt que de 25 est utilisée afin que d'autres ajustements puissent être apportés.

Ensuite, un facteur d'ajustement de la pondération par régression à la première étape est calculé au niveau des AD. Les 34 variables auxiliaires (âge, sexe, état matrimonial, taille du ménage et type de logement) qui doivent être appliquées au niveau des RP à la deuxième étape de l'ajustement sont classées par ordre décroissant en fonction du nombre de ménages auxquels elles s'appliquent dans la population au niveau des AD. Parmi les contraintes figurant sur cette liste, la première, la troisième, et ainsi de suite, sont placées dans un groupe, tandis que les 17 autres sont placées dans un second groupe. On fait la moyenne des facteurs d'ajustement de la pondération établis pour chaque groupe de contraintes, et on les applique aux poids stratifiés « *a posteriori* » (ou aux poids initiaux en l'absence de stratification « *a posteriori* »). Les poids de la première étape permettent généralement de réduire, mais non d'éliminer, les écarts population/estimation au niveau des AD pour les 34 contraintes.

Enfin, un facteur d'ajustement de la pondération par régression à la deuxième étape est calculé au niveau des RP. Les 34 variables auxiliaires sont appliquées au niveau des RP en même temps que deux variables auxiliaires (nombre de ménages et nombre de personnes) pour chaque AD dans la RP afin de déterminer les facteurs d'ajustement de la pondération de la deuxième étape. Ces facteurs sont appliqués aux poids de la première étape de façon à obtenir les poids finaux. Les poids finaux permettent d'éliminer ou de réduire considérablement les écarts population/estimation au niveau des RP pour les 34 variables auxiliaires.

Des contraintes sont retranchées à la première étape et à la deuxième étape pour les raisons suivantes :

- elles sont petites (elles ne s'appliquent qu'à quelques ménages dans la population);
- elles sont redondantes (ou linéairement dépendantes [LD]);
- elles sont quasi redondantes (ou quasi linéairement dépendantes [QLD]);
- elles engendrent des poids aberrants (qui se situent en dehors de l'intervalle de 1 à 25 inclusivement) lors du calcul des poids.

Par exemple, étant donné que la somme du nombre total de femmes et du nombre total d'hommes correspond au nombre total de personnes, le nombre total de femmes est une contrainte redondante ou LD qui peut être retranchée, puisqu'il suffit de satisfaire à deux des contraintes pour que la troisième soit également satisfaite. Si la contrainte « état matrimonial = veuf(ve) » est retranchée parce qu'elle est petite (puisque'il y a très peu de veufs et de veuves dans la RP), alors la somme des autres contraintes « état matrimonial » (célibataire, marié, divorcé et séparé) sera presque égale au nombre total de personnes, ce qui signifie qu'une de ces quatre contraintes pourrait être considérée comme une contrainte quasi redondante, ou QLD, et être retranchée.

D'abord, les mesures suivantes sont prises afin de relever les contraintes « petites », LD et QLD au niveau des RP :

- La taille d'une contrainte est déterminée en fonction du nombre de ménages dans la population auxquels cette contrainte s'applique. Une contrainte dont la taille est inférieure ou égale au paramètre PETITE (qui correspondait à 20, 30 ou 40 ménages en 2006) est retranchée parce que les estimations produites à partir de contraintes petites ont tendance à être très instables.
- Les contraintes LD sont ensuite retranchées.
- Par la suite, le retranchement des contraintes QLD permet de réduire le nombre-condition (l'indice de stabilité) de la matrice à inverser afin d'établir les facteurs d'ajustement de la pondération. Le nombre-condition (voir Press et coll., 1992) est le rapport entre la valeur propre la plus élevée et la valeur propre la plus faible dans la matrice à inverser. Des nombres-conditions élevés indiquent une quasi colinéarité entre les contraintes, ce qui peut causer l'instabilité des estimations. Afin de réduire le nombre-condition, on applique une méthode de sélection ascendante. La matrice est recalculée en ne tenant compte que des deux contraintes les plus fortes. Si le nombre-condition est supérieur au paramètre COND (qui correspondait à 1 000, 2 000, 4 000, 8 000 ou 16 000 en 2006), la deuxième contrainte la plus forte est retranchée. À partir de là, la contrainte suivante en importance est ajoutée à la liste des contraintes qui seront appliquées, la matrice est recalculée et le nombre-condition est déterminé. Si le nombre-condition augmente de plus que la valeur du paramètre COND, la contrainte qui vient d'être ajoutée est retranchée. Ce processus se poursuit jusqu'à ce que toutes les contraintes aient été vérifiées. Après le retranchement de ces contraintes QLD, si le nombre-condition est supérieur au paramètre MAXC (qui correspondait à 10 000, 20 000, 40 000, 80 000 ou 160 000 en 2006), d'autres contraintes sont retranchées. Celles-ci sont retranchées par ordre décroissant, selon la valeur qu'elles ont ajoutée au nombre-condition lorsqu'elles ont été incluses dans la matrice. Le nombre-condition de la matrice est recalculé chaque fois qu'une contrainte est retranchée. Lorsqu'il devient inférieur au paramètre MAXC, on cesse de retrancher des contraintes. Il convient de prendre note qu'en 2006, la valeur du paramètre MAXC était toujours 10 fois plus élevée que la valeur du paramètre COND.
- Les contraintes retranchées jusque-là ne sont pas utilisées dans les calculs de la pondération.

Plus tard, avant de calculer les facteurs d'ajustement de la pondération de la première étape dans une AD donnée, on retranche les contraintes petites qui s'y trouvent toujours. Les contraintes qui restent sont réparties dans deux groupes, comme on l'a déjà expliqué. Puis, pour chaque groupe, on relève et retranche les contraintes linéairement dépendantes (les contraintes qui sont linéairement dépendantes au niveau de l'AD ne le sont pas nécessairement au niveau de la RP). Les facteurs d'ajustement de la pondération de la première étape sont ensuite calculés pour les contraintes comprises dans chaque groupe. Si un des poids ajusté à la première étape se situe en dehors de l'intervalle de 1 à 25 inclusivement, d'autres contraintes sont retranchées. Une méthode similaire à celle utilisée pour retrancher les contraintes QLD est appliquée ici, à ceci près qu'une contrainte est retranchée si elle cause des poids aberrants. Afin d'accroître l'efficacité du calcul, on utilise la méthode de fractionnement (voir Press et coll. 1992) pour déterminer les contraintes qui doivent être retranchées.

Enfin, les facteurs d'ajustement de la pondération de la deuxième étape sont calculés en fonction des contraintes n'ayant pas été retranchées parce qu'elles étaient petites, linéairement dépendantes ou quasi linéairement dépendantes selon l'analyse initiale de la matrice à inverser. Si un des poids ajusté à la deuxième étape se situe en dehors de l'intervalle de 1 à 25 inclusivement, d'autres contraintes sont retranchées selon la méthode décrite pour la première étape.

Les poids du recensement sont calculés indépendamment pour chaque RP, de sorte qu'il est possible d'utiliser des ensembles de paramètres de pondération différents pour chacun d'eux (p. ex., PETITE,

COND, MAXC, stratification « *a posteriori* » ou non, utilisation des contraintes relatives au type de logement ou non). En 1996, un ensemble de paramètres identique était utilisé pour toutes les RP du pays. En 2001 et en 2006, étant donné la capacité de traitement accrue rendue possible par l'exploitation du système de pondération sur plusieurs ordinateurs personnels, il a été convenu de calculer les poids de chaque RP au moyen de plusieurs ensembles de paramètres différents. Deux contraintes relatives au type de logement ont vu le jour en 2006 en raison d'incohérences marquées pour ces caractéristiques dans certaines régions en 2001. Il s'agit des logements individuels non attenants et des appartements dans des immeubles de moins de cinq étages. Bien que ces nouvelles contraintes puissent atténuer les incohérences pour ces caractéristiques, elles pourraient entraîner l'élimination d'autres contraintes, ce qui donnerait lieu à une incohérence plus importante pour ces autres caractéristiques. Par conséquent, l'utilisation des contraintes relatives au type de logement a été paramétrée. Ainsi, ces contraintes pouvaient seulement être utilisées dans les RP où leur effet global sur les incohérences était positif. Vingt différents ensembles de paramètres ont été utilisés pour calculer les poids dans chaque RP en 2006. Ils représentaient les 10 ensembles de paramètres utilisés en 2001, les contraintes relatives au type de logement étant exclues (comme en 2001) et incluses. Une statistique était calculée pour chaque ensemble de paramètres afin de déterminer quel ensemble réduisait les différences entre les chiffres de population et les estimations-échantillon pour les contraintes. Les poids obtenus avec cet ensemble de paramètres étaient utilisés pour la RP correspondante. Ce processus de sélection des meilleurs poids pour chaque RP avait pour nom « sélection choisie » des paramètres.

Pour obtenir de plus amples renseignements sur les estimateurs par régression, consulter Bankier (2002) et Fuller (2002).

Des poids de régression sont établis uniquement pour les ménages privés parmi une UC échantillonnée et les unités pour personnes âgées échantillonnées ayant reçu le questionnaire complet du recensement (un cinquième de ces ménages et de ces unités a été échantillonné; quatre cinquièmes ne le fut pas). Un poids de 0 était attribué aux ménages privés et aux unités composées de personnes âgées de l'UC échantillonnée ayant reçu une formule abrégée, étant donné l'absence d'information sur les variables de l'échantillon. On attribuait un poids de 1 à tous les ménages privés et des unités composées de personnes âgées des UC non échantillonnées, car la totalité des répondants de ces secteurs fournissent des renseignements sur la formule 2B ou 2D. Les ménages collectifs recevaient aussi un poids de 1. Dans le présent rapport, et à moins d'avis contraire, on emploie le terme « ménage » pour désigner un ménage privé ou une unité composée de personnes âgées.

4.5 Traitement à deux passages

Aux recensements de 1996, de 2001 et de 2006, les réponses écrites à la question sur le lien qui étaient fournies sur les questionnaires abrégés (2A) n'ont pas été saisies en raison de contraintes budgétaires. On leur a plutôt attribué la valeur « Autre ». En revanche, les réponses écrites fournies sur les questionnaires complets (2B) à la question sur le lien ont été saisies et codées selon la méthode normale.

Le traitement à deux passages consiste à traiter les données des questionnaires complets en deux étapes. Lors de la première étape, appelée « premier passage », tous les questionnaires (complets et abrégés) sont traités ensemble, ce qui représente 100 % des données. Les réponses écrites à la question sur le lien qui sont saisies sur les questionnaires complets ne sont pas prises en considération; on leur attribue la valeur « Autre », comme dans le cas des réponses écrites fournies dans les questionnaires abrégés. Le processus de contrôle et d'imputation est le même pour les deux types de questionnaire. Lors de la deuxième étape, appelée « deuxième passage », seuls les questionnaires complets sont traités; les questionnaires abrégés ne sont pas disponibles pour l'imputation. On utilise les réponses écrites à la question sur le lien qui sont saisies sur les questionnaires complets plutôt que les réponses « Autre ».

Comme les réponses écrites sont accessibles, on suppose que la qualité des résultats est plus élevée lors du deuxième passage que lors du premier passage.

Le système de pondération utilise les résultats du premier passage pour tous les ménages afin de calculer les poids associés aux ménages. Bien qu'il soit possible d'utiliser les résultats du premier passage pour les questionnaires abrégés et les résultats du deuxième passage pour les questionnaires complets, cette méthode pourrait biaiser les estimations du recensement, à cause des différences entre le premier et le deuxième passage dans la distribution des réponses aux variables démographiques, qui sont attribuables à la présence de réponses écrites à la question sur le lien pour le deuxième passage. Les estimations du recensement qui ont été diffusées ont été produites en appliquant les poids du premier passage aux résultats imputés des questionnaires complets du deuxième passage. L'écart entre les chiffres de population (selon les données du deuxième passage pour la population échantillonnée et les résultats du premier passage pour la population restante, soit 80 % de la population) et les estimations du deuxième passage est faible pour la plupart des contraintes. Se reporter au [tableau 7.2.2.2](#), au [graphique 7.2.2.3](#) et au [graphique 7.2.2.4](#) à la [section 7.2.2](#), où sont comparés les résultats des premier et deuxième passages.

4.6 Différences entre les chiffres de population et les estimations pondérées finales

Les poids finaux des ménages sont générés de manière à ce que les chiffres de population correspondent aux estimations pondérées pour le plus grand nombre de caractéristiques possible. Parmi les caractéristiques qui sont disponibles dans les questionnaires complet et abrégé et qu'on cherche à uniformiser, mentionnons les groupes d'âge quinquennaux, le sexe, l'état matrimonial, l'union libre, la taille du ménage et le type de logement. Le processus de pondération vise à neutraliser les différences entre la population et les estimations au niveau de la région de pondération (RP), qui contient généralement de 1 000 à 3 000 logements susceptibles d'être échantillonnés.

Il existe certaines raisons qui expliquent pourquoi les estimations de l'échantillon peuvent être différentes des chiffres de population, surtout dans le cas des petites régions. Les principales sont énumérées ci-dessous :

- 1) Contraintes éliminées pendant le processus de régression : comme mentionné à la [section 4.4](#), les contraintes peuvent être éliminées parce qu'elles ont des poids aberrants, qu'elles ont de petits chiffres ou qu'elles sont linéairement dépendantes ou quasi linéairement dépendantes. Les contraintes qui sont éliminées ne font pas l'objet d'un contrôle et comportent habituellement une certaine différence entre les chiffres et les estimations de population.
- 2) Sous-régions de pondération : la région de pondération est la plus petite région géographique pour laquelle le système de pondération cherche à obtenir une correspondance entre les chiffres de population et les estimations pondérées pour le plus grand nombre de variables auxiliaires possible. Par conséquent, dans les petites régions qui font partie des RP (comme les AD ou les très petites municipalités), on observera habituellement des incohérences entre les chiffres de population et les estimations pondérées.

4.7 Les différents univers

Il existe trois types d'univers du recensement distincts pour lesquels les données du recensement peuvent être observées :

- 1) Logements privés : ménages privés et unités composés de personnes âgées qui ont fait l'objet d'un échantillonnage. Ces ménages ont servi à la création des poids finaux au niveau des ménages. La

majorité des renseignements qui figurent dans la présente publication correspondent à l'univers des logements privés.

- 2) Logements privés et logements collectifs non institutionnels : ménages privés, unités composées de personnes âgées et logements collectifs non institutionnels visés par l'échantillon, et ménages privés et unités composées de personnes âgées des UC non échantillonnées. Étant donné que les personnes supplémentaires de cet univers ont toutes reçu un questionnaire complet, les données du questionnaire 2B sont disponibles. Cet univers est utilisé dans toutes les publications du recensement au sujet des variables échantillonnées.
- 3) Logements privés et collectifs : tous les ménages privés et les unités composées de personnes âgées (faisant partie ou non de l'échantillon), ainsi que tous les logements collectifs (en logements institutionnels et non institutionnels). Les résidents des logements collectifs institutionnels répondent au questionnaire abrégé. Par conséquent, aucune données échantillonnées ne sont disponibles. Pour cette raison, cet univers au complet est utilisé dans toutes les publications du recensement au sujet des variables de base (question posée sur les questionnaires abrégés et complets), mais il ne peut pas être utilisé pour les publications échantillonnées.

La population des logements collectifs institutionnels représente une partie des différences qui seront observées par quelqu'un qui compare une publication 2B avec une publication 2A. Les comptes et les estimations pour les trois univers mentionnés précédemment se trouvent au [tableau 7.2.2.3](#).

5 Programme d'évaluation de l'échantillonnage et de la pondération

Le programme d'évaluation de l'échantillonnage et de la pondération avait pour objet de déterminer l'effet de l'échantillonnage et de la pondération sur la qualité des données-échantillon du recensement. En tout, quatre études ont été réalisées afin de mesurer la qualité des données-échantillon et des estimations-échantillon du recensement, et de fournir des renseignements utiles à la planification des recensements futurs. Les études comportaient les activités suivantes :

- a) un examen du biais d'échantillonnage;
- b) une évaluation des méthodes de pondération;
- c) une évaluation de la cohérence des estimations-échantillon et des chiffres de population;
- d) une détermination des erreurs types pour diverses caractéristiques-échantillon (20 %).

Chacune de ces études est brièvement décrite ci-après tandis que les résultats obtenus sont présentés aux [chapitres 6, 7, 8 et 9](#).

5.1 Biais d'échantillonnage

L'objectif de cette étude était d'évaluer les différences entre les estimations fondées sur les poids initiaux et les chiffres de population connus. Ces différences présentent de l'intérêt pour deux raisons. D'une part, elles peuvent être utiles pour déterminer les biais dans l'échantillon de ménages sélectionnés sur le terrain, et d'autre part, elles peuvent indiquer une incidence potentiellement négative de la non-réponse sur les questions-échantillon du recensement. Les biais dans les caractéristiques des questionnaires abrégés sont corrigés par calibrage pendant la pondération. Si les caractéristiques des questionnaires complets sont corrélées avec celles des questionnaires abrégés, le calibrage devrait aussi réduire leurs biais.

5.2 Évaluation des méthodes de pondération

Le but de cette étude était d'évaluer l'efficacité de l'estimateur de régression pseudo-optimale. À cette fin, on a examiné le niveau de concordance entre les estimations-échantillon (selon les poids finaux) et les chiffres de population à l'égard de l'ensemble des contraintes au niveau des RP. En outre, pour expliquer les incohérences observées, on a étudié le nombre et le type des contraintes retranchées au niveau des RP, ainsi que les raisons de leur retranchement, puis on a examiné la distribution des poids du recensement.

5.3 Cohérence entre les estimations-échantillon et les chiffres de population

Dans cette étude, on a examiné le niveau de concordance entre les estimations-échantillon (selon les poids finaux) et les chiffres de population pour les caractéristiques de base utilisées comme contraintes, et ce, pour diverses unités géographiques.

5.4 Variance d'échantillonnage

L'erreur type (la racine carrée de la variance) d'une estimation est une mesure de sa précision. Il est possible de calculer assez rapidement des estimations d'erreurs types pour des estimateurs lorsqu'on utilise des poids simples égaux à 5 et qu'on suppose un échantillonnage aléatoire simple. Toutefois, il est plus long d'estimer les erreurs types des estimateurs du recensement qui tiennent compte du plan d'échantillonnage et des techniques d'estimation utilisés. Ainsi, on calcule des « facteurs d'ajustement » qui représentent le rapport entre les estimations des erreurs types des estimations du recensement et les estimations simples des erreurs types. On peut alors obtenir une estimation de l'erreur type d'une estimation du recensement pour n'importe quelle caractéristique dans toute unité géographique en multipliant l'estimation simple de l'erreur type par le facteur d'ajustement approprié.

6 Biais d'échantillonnage

Dans ce chapitre, nous déterminerons si, après ajustement pour la non-réponse, l'échantillon du recensement est biaisé. Pour ce faire, nous calculerons la statistique Z pour les caractéristiques 2A telles que l'état matrimonial = célibataire, où le chiffre de population X peut être comparé avec l'estimation-échantillon $\hat{X}^{(0)}$ fondée sur les poids initiaux. Pour calculer la statistique Z , on divise la différence entre l'estimation et le chiffre de population par la racine carrée de la variance de l'estimation. Si l'échantillonnage est aléatoire et sans biais, on peut voir que $Z^{(0)}$ suivra une distribution à peu près normale (moyenne = 0; variance = 1) (voir l'[annexe C](#)).

Le [tableau 6.1](#) et le [graphique 6.1](#) présentent les valeurs Z au niveau du Canada pour les recensements de 2001 et de 2006 (ainsi que les différences $\hat{X}^{(0)} - X$) pour 34 caractéristiques très semblables aux contraintes ayant été appliquées pour le calcul des poids finaux du recensement (voir l'[annexe B](#)). Si la statistique $Z^{(0)}$ suit une distribution normale, la probabilité que $|Z^{(0)}| > 3$ est d'environ 0,0026 pour une caractéristique, ce qui laisse supposer qu'en moyenne, nous nous attendrions à ce que $0,0026 \times 34 = 0,0884$ des 34 caractéristiques correspondent à $|Z^{(0)}| > 3$. Cependant, selon le [tableau 6.1](#), 22 des 34 caractéristiques en 2006 affichent une valeur Z se situant en dehors de l'intervalle de -3 à 3 , ce qui constitue une forte indication que l'échantillon du Recensement de 2006 est biaisé. De même, en 2001, 25 des 32 caractéristiques se situaient à l'extérieur de cet intervalle.

Le [graphique 6.1](#) révèle que la statistique Z est très différente entre 2001 et 2006 pour de nombreuses caractéristiques. Étant donné que Z est une variable aléatoire, il est possible que certaines de ces différences ne soient pas statistiquement significatives. Une statistique W , définie à l'[annexe C](#), a été calculée pour chaque caractéristique afin de déterminer s'il existait des différences significatives entre les valeurs Z de 2001 et de 2006. La statistique W , sa valeur de p et 17 caractéristiques dont les valeurs Z comportent des différences statistiquement significatives (parce que leurs valeurs de p sont inférieures à 0,05) sont indiquées au [tableau 6.1](#) et au [graphique 6.1](#).

Dans le [graphique 6.1](#), on constate que le biais vers le bas dans l'échantillon de 2006 augmentait de façon significative (indiqué par un astérisque) chez les personnes de sexe masculin, les hommes âgés de 15 ans et plus, les personnes âgées de 15 à 19 ans, les personnes seules, les ménages composés de trois personnes et les ménages composés de six personnes ou plus, tandis que le biais vers le haut dans l'échantillon de 2006 augmentait de façon significative pour les ménages composés de cinq personnes. De plus, on peut constater que le biais vers le bas dans l'échantillon de 2006 diminuait de façon significative pour les personnes séparées et les ménages composés de une seule personne, tandis que le biais vers le haut dans l'échantillon de 2006 diminuait de façon significative pour les personnes de sexe féminin, les femmes âgées de 15 ans et plus, les personnes âgées de 10 à 14 ans, les personnes mariées et les ménages composés de quatre personnes. Enfin, le biais vers le haut dans l'échantillon de 2006 faisait volte-face pour se transformer en biais vers le bas pour le chiffre de population totale, chez les personnes du groupe d'âge de 5 à 9 ans et celui de 40 à 44 ans.

Le [graphique 6.1](#) révèle également un biais constant vers le bas dans les échantillons des recensements de 2001 et 2006 des personnes âgées de 20 à 39 ans, ainsi qu'un biais constant vers le haut dans les

échantillons des recensements de 2001 et de 2006 des personnes âgées de 45 ans et plus et des ménages composés de deux personnes.

Le [graphique 6.1](#) indique aussi un biais vers le haut très prononcé en 2006 pour les logements individuels non attenants, ainsi qu'un biais vers le bas un peu moins prononcé pour les appartements de moins de cinq étages.

Le biais de l'échantillon peut avoir diverses origines, notamment les erreurs des agents recenseurs (p. ex., ne pas sélectionner l'échantillon conformément aux spécifications), le biais de non-réponse (p. ex., les jeunes hommes adultes sont moins portés à remplir un questionnaire complet qu'un questionnaire abrégé), le biais de réponse (p. ex., les répondants qui donnent des réponses différentes sur la formule 2B que ce qu'ils répondraient sur la formule 2A), les erreurs de traitement, etc.

Les biais prononcés dans l'échantillon de 2006 pour les ménages composés de cinq et de six personnes étaient attribuables à la réduction du nombre de personnes du questionnaire complet de six personnes à cinq personnes, puisqu'il fallait allouer plus d'espace pour permettre la saisie de données automatisée pour les réponses écrites. Le nombre de personnes du questionnaire abrégé est demeuré à six. En 2006, les ménages composés de plus de cinq personnes qui avaient reçu un questionnaire complet ne demandaient pas toujours un deuxième formulaire complet, et il leur arrivait d'indiquer que leur ménage était composé de cinq personnes seulement. Ainsi, le biais vers le haut a fortement augmenté pour l'échantillon de ménages composés de cinq personnes, tandis qu'un biais vers le bas correspondant a été enregistré pour l'échantillon des ménages composés de six personnes. Le processus de calibrage des poids n'a pu corriger qu'en partie ces biais, qui ont également rendu difficile le calibrage pour corriger les autres biais.

La non-réponse est une autre cause possible du biais dans l'échantillon du recensement. Le pourcentage des ménages qui n'avaient pas répondu à la fin des opérations sur le terrain s'établissait à 2,8 % en 2006, comparativement à 1,6 % en 2001. Après les ajustements apportés au statut d'occupation dans le cadre de l'Enquête sur la classification des logements (voir la [section 2.7](#)), le pourcentage de logements occupés qui n'avaient pas répondu se chiffrait à 3,5 % en 2006, comparativement à 2,0 % en 2001. En 2006, l'imputation du ménage au complet a été utilisée pour imputer 96 % des données des ménages qui n'avaient pas répondu du tout, 18,6 % d'entre eux rejoignant la catégorie des ménages ayant reçu un questionnaire complet. En 2001, on n'a pas eu recours à l'imputation de ménages au complet. En 2001 et en 2006, les questionnaires complets affichant une non-réponse totale aux questions de l'échantillon ont été convertis en questionnaires abrégés. On appelait ce processus « conversion des documents ». En 2006, pour 12 638 ménages qui avaient reçu un questionnaire complet, ce dernier a été converti en questionnaire abrégé. En 2001, pour 17 692 ménages ayant reçu un questionnaire complet comportant certaines réponses au questionnaire abrégé, mais aucune réponse au questionnaire complet, le questionnaire complet a été converti au questionnaire abrégé, tandis que pour 144 282 ménages qui affichaient une non-réponse totale (on estime qu'ils avaient reçu au départ un questionnaire complet dans 20 % des cas), le questionnaire complet est devenu un questionnaire abrégé. L'importante réduction du nombre de questionnaires complets convertis en questionnaires abrégés en 2006 était attribuable à l'utilisation de la méthode d'imputation de ménages au complet pour la plupart des ménages affichant une non-réponse totale.

Ce changement a été apporté parce que l'approche de 2001 pourrait avoir entraîné des biais importants dans l'échantillon. Par exemple, en 2001, on savait que le pourcentage des logements individuels non attendants qui faisaient partie de la catégorie des ménages affichant une non-réponse totale correspondait à la moitié de celui de l'ensemble de la population. Voir la [section 7.2.2](#) pour une description plus détaillée des conséquences sur le biais d'échantillonnage de l'introduction de l'imputation de ménages au complet en 2006. La description de la [section 7.2.2](#) met en doute l'utilité de la méthode de la statistique W susmentionnée pour déterminer si les valeurs Z de 2001 et de 2006 étaient très différentes. En effet, bon nombre des différences semblaient attribuables à l'introduction de la méthode de l'imputation de ménages au complet plutôt qu'à la variabilité de l'échantillonnage. Toutefois, cette question ne concerne pas l'utilisation de la statistique W pour évaluer les différences régionales dans le biais de 2006 ci-dessous.

Une troisième source de biais possible provient des erreurs commises par le répondant ou survenues au cours du processus de saisie des données. Certaines des incohérences qui en ont découlé ont été détectées et corrigées grâce au processus de contrôle et d'imputation décrit à la [section 2.8](#).

La variation géographique du biais a aussi été étudiée. Les valeurs Z correspondant aux 34 caractéristiques ont été calculées pour la région de l'Est, le Québec, l'Ontario et la région de l'Ouest (qui comprend les trois territoires) de la même façon que pour le Canada. Le biais relatif entre ces quatre régions est présenté pour les recensements de 2006 et de 2001 dans les [graphiques 6.2](#) et [6.3](#) respectivement. Là encore, on s'est servi de la statistique W pour relever les différences statistiquement significatives entre les régions, qui sont indiquées par la présence des initiales des régions visées en bas ou en haut du graphique. Par exemple, la mention « Ou Q et On Q » indique qu'il y a une différence significative entre la région de l'Ouest et le Québec et entre l'Ontario et Québec.

Lorsque l'on compare le [graphique 6.2](#) au [graphique 6.3](#), on constate qu'il y a eu un plus grand nombre de différences régionales significatives en 2006 qu'en 2001. Il est intéressant de souligner qu'en 2006, le biais vers le bas pour le chiffre de population totale était beaucoup plus prononcé dans l'Ouest et en Ontario qu'au Québec et dans l'Est. En outre, les personnes de sexe féminin de l'échantillon affichaient un biais vers le bas dans l'échantillon de l'Ouest et de l'Ontario, et un biais vers le haut pour le Québec et l'Est.

À la [section 7.2.2](#) et au [chapitre 8](#), on verra que ces différences population/estimation peuvent souvent être réduites considérablement grâce au calibrage des poids du recensement. Par conséquent, les inférences fondées sur des estimations calibrées devraient être plus justes.

Tableau 6.1 Différences population/estimation fondées sur les poids initiaux, recensements de 2006 et de 2001

Caractéristique	Recensement de 2006						Recensement de 2001						2006 par rapport à 2001		
	Chiffre	Estimation ¹	Différence ²	Écart ³	ET ⁴	Valeur Z ⁵	Chiffre	Estimation ¹	Différence ²	Écart ³	ET ⁴	Valeur Z ⁵	W	Valeur p	Différence significative du biais ⁶
Homme	15 041 551	14 962 170	- 79 381	-0,53	6 273	-12,65	14 171 941	14 146 867	- 25 074	-0,18	6 139	-4,08	-5,83	0,000	O
Femme	15 653 041	15 653 964	923	0,01	6 095	0,15	14 699 518	14 772 915	73 397	0,50	5 940	12,36	-8,79	0,000	O
Population totale	30 694 592	30 616 134	- 78 458	-0,26	9 227	-8,50	28 871 459	28 919 783	48 324	0,17	8 991	5,37	-9,77	0,000	O
Homme ≥ 15 ans	12 263 445	12 187 117	- 76 328	-0,62	4 917	-15,52	11 340 286	11 295 995	- 44 291	-0,39	4 747	-9,33	-4,00	0,000	O
Femme ≥ 15 ans	13 005 067	13 007 175	2 108	0,02	4 557	0,46	11 998 509	12 042 929	44 420	0,37	4 342	10,23	-7,03	0,000	O
0 à 4 ans	1 640 859	1 639 505	- 1 354	-0,08	2 987	-0,45	1 636 092	1 641 720	5 628	0,34	2 986	1,88	-1,65	0,098	N
5 à 9 ans	1 760 005	1 756 879	- 3 126	-0,18	3 067	-1,02	1 910 359	1 928 604	18 245	0,96	3 213	5,68	-4,68	0,000	O
10 à 14 ans	2 025 216	2 025 458	242	0,01	3 296	0,07	1 986 213	2 010 534	24 321	1,22	3 271	7,44	-5,24	0,000	O
15 à 19 ans	2 083 373	2 070 265	- 13 108	-0,63	3 315	-3,95	1 986 163	1 983 519	- 2 644	-0,13	3 245	-0,81	-2,18	0,030	O
20 à 24 ans	2 029 449	1 978 067	- 51 382	-2,53	3 267	-15,73	1 892 572	1 851 491	- 41 081	-2,17	3 168	-12,97	-1,56	0,120	N
25 à 29 ans	1 940 880	1 905 221	- 35 659	-1,84	3 142	-11,35	1 835 744	1 810 124	- 25 620	-1,40	3 077	-8,33	-1,90	0,058	N
30 à 34 ans	1 976 478	1 956 416	- 20 062	-1,02	3 131	-6,41	2 031 513	2 013 625	- 17 888	-0,88	3 173	-5,64	-0,60	0,545	N
35 à 39 ans	2 161 430	2 148 991	- 12 439	-0,58	3 224	-3,86	2 452 299	2 446 624	- 5 675	-0,23	3 427	-1,66	-1,68	0,092	N
40 à 44 ans	2 559 477	2 541 482	- 17 995	-0,70	3 467	-5,19	2 510 847	2 513 920	3 073	0,12	3 439	0,89	-4,29	0,000	O
45 à 49 ans	2 571 429	2 579 715	8 286	0,32	3 462	2,39	2 273 676	2 283 700	10 024	0,44	3 286	3,05	-0,60	0,548	N
50 à 54 ans	2 313 657	2 327 389	13 732	0,59	3 305	4,15	2 031 050	2 041 054	10 004	0,49	3 137	3,19	0,48	0,631	N
55 à 59 ans	2 045 868	2 060 482	14 614	0,71	3 139	4,66	1 549 675	1 567 071	17 396	1,12	2 758	6,31	-1,74	0,082	N
60 à 64 ans	1 558 145	1 570 085	11 940	0,77	2 751	4,34	1 234 930	1 249 389	14 459	1,17	2 469	5,86	-1,52	0,129	N
65 à 74 ans	2 229 023	2 248 564	19 541	0,88	3 371	5,80	2 059 079	2 083 362	24 283	1,18	3 256	7,46	-1,38	0,167	N
75+ ans	1 799 303	1 807 615	8 312	0,46	2 927	2,84	1 481 247	1 495 045	13 798	0,93	2 676	5,16	-1,93	0,053	N
Célibataire	14 170 280	13 989 389	- 180 891	-1,28	8 347	-21,67	13 282 845	13 196 174	- 86 671	-0,65	8 018	-10,81	-7,40	0,000	O
Marié(e)	12 291 457	12 406 423	114 966	0,94	6 942	16,56	11 750 092	11 906 204	156 112	1,33	6 678	23,38	-4,91	0,000	O
Veuf(ve)	1 435 852	1 431 006	- 4 846	-0,34	2 321	-2,09	1 341 497	1 339 109	- 2 388	-0,18	2 254	-1,06	-0,68	0,494	N
Divorcé(e)	2 044 164	2 040 145	- 4 019	-0,20	2 989	-1,34	1 794 079	1 784 704	- 9 375	-0,52	2 824	-3,32	1,52	0,129	N
Séparé(e)	752 839	749 172	- 3 667	-0,49	1 800	-2,04	702 946	693 591	- 9 355	-1,33	1 749	-5,35	2,45	0,014	O
Union libre = oui	2 725 161	2 703 240	- 21 921	-0,80	4 419	-4,96	2 267 634	2 253 253	- 14 381	-0,63	4 090	-3,52	-0,70	0,483	N
Ménage de 1 personne	3 338 596	3 329 891	- 8 705	-0,26	3 017	-2,89	2 908 857	2 866 182	- 42 675	-1,47	2 847	-14,99	9,06	0,000	O

Tableau 6.1 Différences population/estimation fondées sur les poids initiaux, recensements de 2006 et de 2001 (suite)

Caractéristique	Recensement de 2006						Recensement de 2001						2006 par rapport à 2001		
	Chiffre	Estimation ¹	Différence ²	Écart ³	ET ⁴	Valeur Z ⁵	Chiffre	Estimation ¹	Différence ²	Écart ³	ET ⁴	Valeur Z ⁵	W	Valeur p	Différence significative du biais ⁶
Ménage de 2 personnes	4 153 415	4 182 506	29 091	0,70	3 368	8,64	3 709 282	3 739 781	30 499	0,82	3 224	9,46	-1,02	0,305	N
Ménage de 3 personnes	1 963 201	1 951 120	- 12 081	-0,62	2 611	-4,63	1 848 476	1 845 071	- 3 405	-0,18	2 541	-1,34	-2,25	0,024	O
Ménage de 4 personnes	1 843 987	1 848 858	4 871	0,26	2 506	1,94	1 812 783	1 826 921	14 138	0,78	2 481	5,70	-2,67	0,007	O
Ménage de 5 personnes	713 994	735 811	21 817	3,06	1 659	13,15	714 618	719 013	4 395	0,61	1 664	2,64	7,42	0,000	O
Ménage de 6+ personnes	338 241	302 192	- 36 049	-10,66	1 157	-31,16	332 959	328 968	- 3 991	-1,20	1 155	-3,46	19,42	0,000	O
Logements individuel non attenant	6 769 581	6 812 477	42 896	0,63	2 554	16,80
Appartement < 5 étages	2 285 965	2 258 899	- 27 066	-1,18	2 119	-12,77

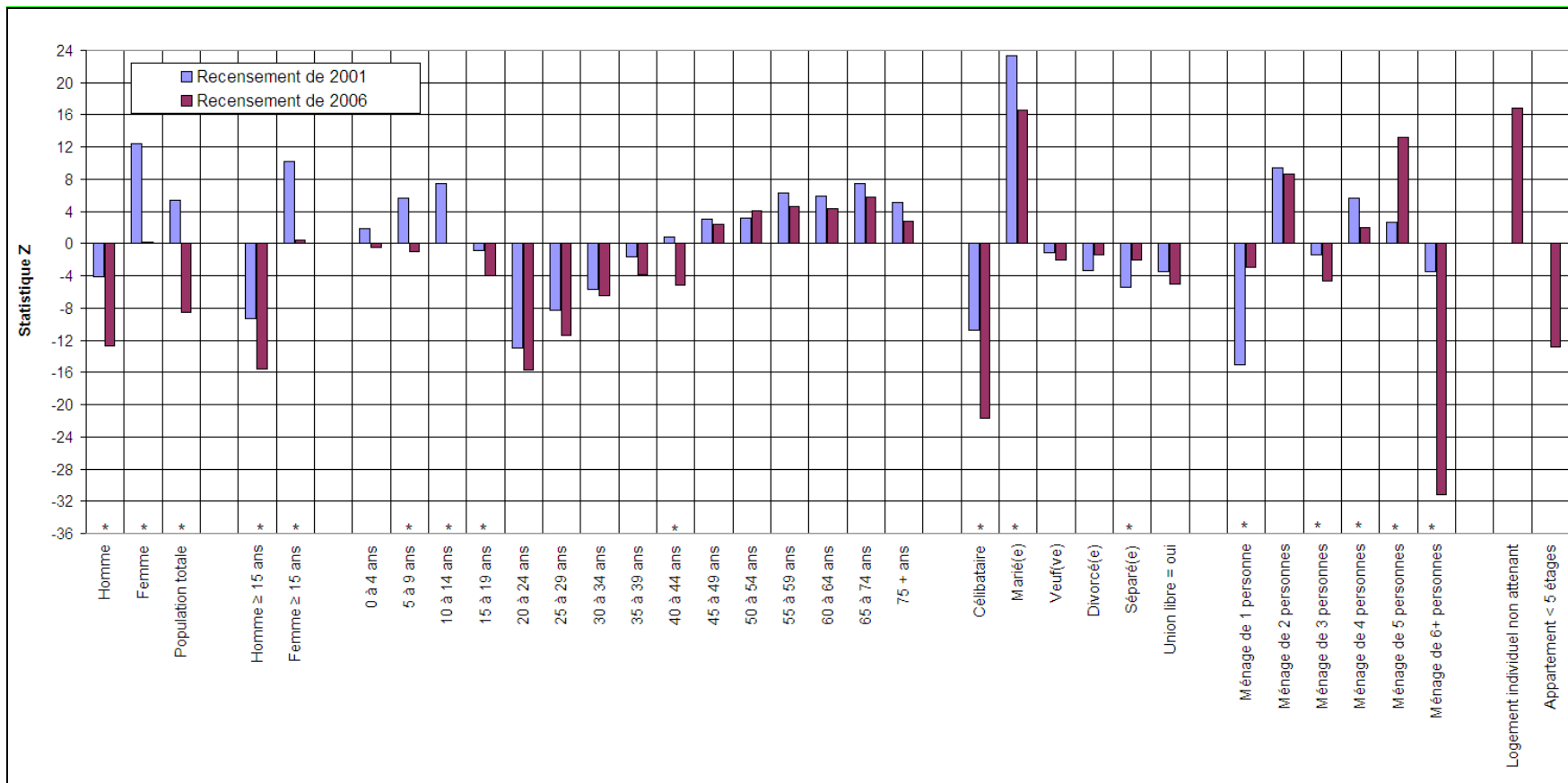
.. indisponible pour une période de référence précise

Notes :

- 1. Selon les poids initiaux.
- 2. Différence : estimation-chiffre.
- 3. Écart : (100*[estimation-chiffre]/chiffre).
- 4. ET : Erreur type de l'estimation des poids initiaux.
- 5. Valeur Z : (estimation-chiffre)/ET.
- 6. Indicateur de l'ampleur de la différence entre les biais initiaux de 2006 et de 2001 (selon la valeur p < 0,05).

Sources : Statistique Canada, recensements de 2006 et 2001.

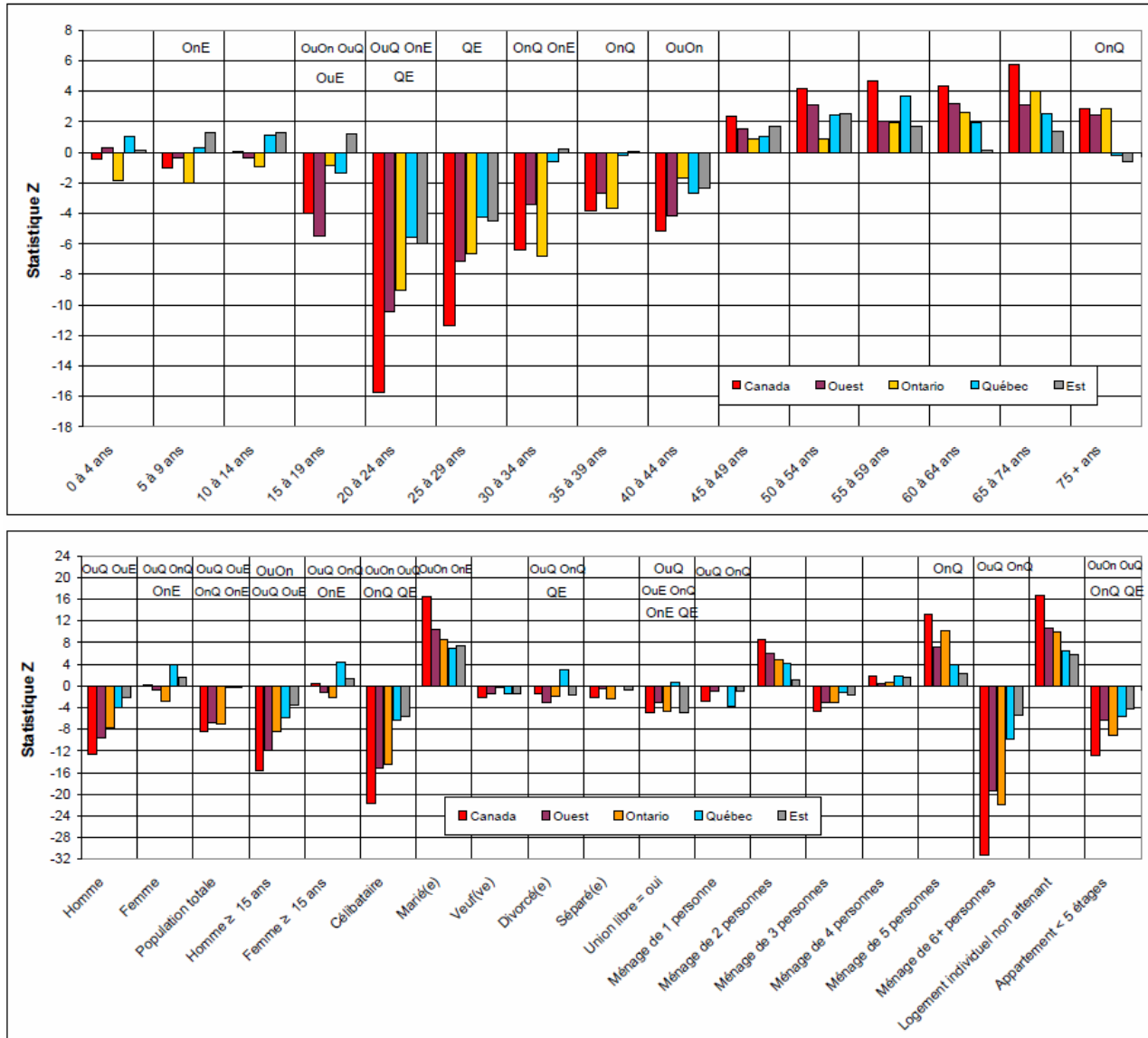
Graphique 6.1 Statistiques Z des différences population/estimation fondées sur les poids initiaux, pour le Canada, recensements de 2006 et de 2001



* Indique une différence statistiquement significative entre le biais de 2006 et celui de 2001.

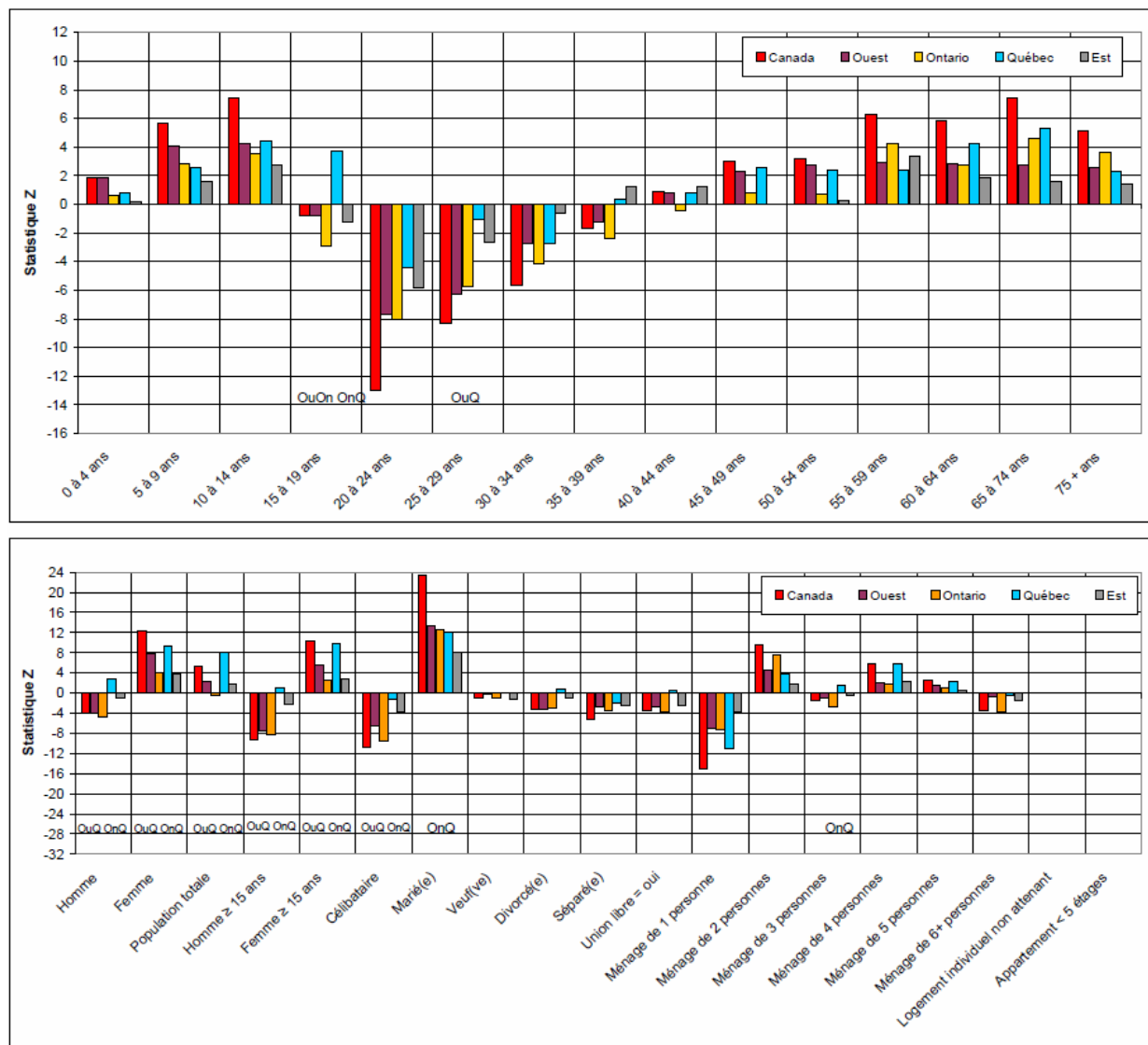
Sources : Statistique Canada, recensements de 2006 et 2001.

Graphique 6.2 Statistiques Z selon les régions en 2006



Source : Statistique Canada, Recensement de 2006.

Graphique 6.3 Statistiques Z selon les régions en 2001



Source : Statistique Canada, Recensement de 2001.

7 Évaluation des méthodes de pondération

Ce chapitre a pour but de présenter, puis d'évaluer, certains éléments du processus de pondération du recensement telles la formation des régions de pondération (RP) et la répartition des poids selon leur taille. On y examine également, pour un éventail de caractéristiques, les écarts entre les chiffres de population et les estimations-échantillon à l'échelle du Canada. De plus, on explique les résultats du premier passage comparativement à ceux du deuxième passage, ainsi que les différents univers de données pour lesquels des données de recensement peuvent être présentées. Enfin, on aborde la question du retranchement des contraintes, notamment dans le but de démontrer comment la fréquence à laquelle les contraintes sont retranchées peut influencer sur les écarts observés.

7.1 Formation des régions de pondération (RP)

Au Recensement de 2006, le pays était divisé en 6 607 RP, chacune comptant en moyenne approximativement 8 AD entières. Pour chaque RP, le programme de pondération vise une concordance entre certaines estimations-échantillon et les chiffres de population correspondants. La formation des RP consistait à regrouper des AD et devait respecter les conditions suivantes :

- a) Une RP doit respecter les limites des divisions de recensement (DR).
- b) Une RP doit contenir entre 1 000 et 3 000 ménages.
- c) Une RP doit, dans la mesure du possible, respecter les limites des régions nommées ci-après, la priorité étant donnée, par ordre d'importance, aux subdivisions de recensement (SDR) et aux secteurs de recensement (SR).
- d) Les RP doivent, dans la mesure du possible, être formées d'AD contiguës (c.-à-d. ne pas être fractionnées entre deux parties ou plus et ne pas renfermer de « trous ») et être le plus compactes possible.

Le [tableau 7.1.1](#) montre que 6 559 RP, ou 99,3 % de l'ensemble des RP, se situent à l'intérieur de l'intervalle des 1 000 à 3 000 ménages au Recensement de 2006. La situation s'est considérablement améliorée par rapport à 2001, où seulement 94,2 % des RP se situaient à l'intérieur de l'intervalle. Étant donné que l'algorithme utilisé pour produire les RP en 2006 était le même qu'en 2001, les résultats automatisés étaient semblables. Toutefois, l'amélioration est attribuable au fait que le nombre d'ajustements manuels apportés à la fin du processus était bien plus grand en 2006 qu'en 2001. Bon nombre des RP anormales ont été divisées, fusionnées ou réalignées de manière à mieux respecter les conditions mentionnées ci-dessus.

Le nombre moyen de logements par RP était de 1 869. La plus grande RP contenait 4 820 logements, ce qui constitue une amélioration par rapport à 2001, où la plus grande RP était composée de 17 043 logements. En 2006, cinq RP avaient une population de zéro. Dans ces cas, les RP contenaient des AD qui n'étaient pas visées par l'échantillonnage. Ces RP ayant une population de zéro sont au Labrador, dans les Territoires du Nord-Ouest et au Nunavut.

On obtiendra une concordance entre les estimations-échantillon et les chiffres de population uniquement pour les régions géographiques formées exclusivement de RP complètes. Ces régions comprennent les provinces et les DR, ainsi que les SDR et les SR dont aucune des RP ne fait partie d'une autre SDR ou d'un autre SR. Le [tableau 7.1.2](#) permet d'examiner le lien entre les limites des SDR et des SR du Recensement de 2006 et les limites des RP. Quatre scénarios qui s'excluent mutuellement sont possibles :

1. « Régions géographiques qui contiennent seulement une partie d'une RP, tandis que le reste de la RP contient seulement des régions géographiques complètes du même type » : ce scénario signifie que la SDR ou le SR était d'une taille suffisamment petite pour entrer au complet dans une RP, et que la même RP était seulement composée de SDR et de SR complets. Aucune SDR ni aucun SR de cette RP ne faisait aussi partie d'une autre RP. Par conséquent, la condition c) était satisfaite. Ce scénario survient souvent pour les SDR, en raison de leurs nombreuses très petites municipalités, comme les réserves indiennes et les villages, dont la population visée par l'échantillonnage est limitée ou inexistante.
2. « Régions géographiques qui contiennent seulement une partie d'une RP, tandis que le reste de la RP ne contient pas seulement des régions géographiques complètes du même type » : ce scénario signifie que la SDR ou le SR était d'une taille suffisamment petite pour entrer au complet dans une RP, mais qu'une SDR ou un SR différent au sein d'une même RP faisait aussi partie d'une autre RP. La condition c) n'est pas satisfaite.
3. « Régions géographiques qui contiennent au moins une RP complète » : ce scénario signifie que la SDR ou le SR était d'une taille suffisamment grande pour contenir des RP complètes. Aucune des RP ne faisait aussi partie d'une autre SDR ou d'un autre SR. Par conséquent, la condition c) était satisfaite. Ce scénario survient souvent dans les SR, puisque ces derniers se rencontrent dans les régions urbaines, qui sont habituellement visées par l'échantillonnage, et que les SR doivent en général avoir une taille supérieure à celle des RP.
4. « Régions géographiques qui traversent au moins une limite de RP » : ce scénario signifie que la SDR ou le SR est partagé par au moins deux RP. La condition c) n'est pas satisfaite.

D'après les chiffres présentés au [tableau 7.1.2](#), on constate que 13,2 % des SDR et 67,0 % des SR sont formés d'au moins une RP complète. Ainsi, les chances d'obtenir une concordance entre les chiffres de population et les estimations-échantillon y seront vraisemblablement meilleures. Les résultats du [tableau 7.1.2](#) sont très semblables à ceux de 2001, puisque le même algorithme automatisé a été utilisé aux deux recensements.

Pour obtenir de plus amples renseignements au sujet des régions de pondération et de leurs limites, se reporter à Kruszynski (1999).

Tableau 7.1.1 Répartition des régions de pondération selon leur taille

Logements	Recensement de 2006		Recensement de 2001	
	Nombre de RP	Pourcentage (%)	Nombre de RP	Pourcentage (%)
0 à 499	7	0,1	7	0,1
500 à 999	2	0,0	1	0,0
1 000 à 1 499	1 705	25,8	1 132	18,4
1 500 à 1 999	2 443	37,0	2 248	36,6
2 000 à 2 499	1 768	26,8	1 622	26,4
2 500 à 2 999	643	9,7	786	12,8
3 000 à 3 499	30	0,5	352	5,7
3 500+	9	0,1		
Total	6 607	100,0	6 148	100,0

Sources : Statistique Canada, recensements de 2006 et 2001.

Tableau 7.1.2 Nombre de subdivisions de recensement et de secteurs de recensement qui respectent les limites des régions de pondération, Recensement de 2006

Description	SDR		SR	
	Nombre	Pourcentage (%)	Nombre	Pourcentage (%)
Régions géographiques qui contiennent seulement une partie d'une RP, tandis que le reste de la RP contient seulement des régions géographiques complètes du même type	4 255	78,5	1 466	28,8
Régions géographiques qui contiennent seulement une partie d'une RP, tandis que le reste de la RP ne contient pas seulement des régions géographiques complètes du même type	337	6,2	115	2,3
Régions géographiques qui contiennent au moins une RP complète	714	13,2	3 409	67,0
Régions géographiques qui traversent au moins une limite de RP	112	2,1	99	1,9
Total	5 418	100,0	5 089	100,0

Source : Statistique Canada, Recensement de 2006.

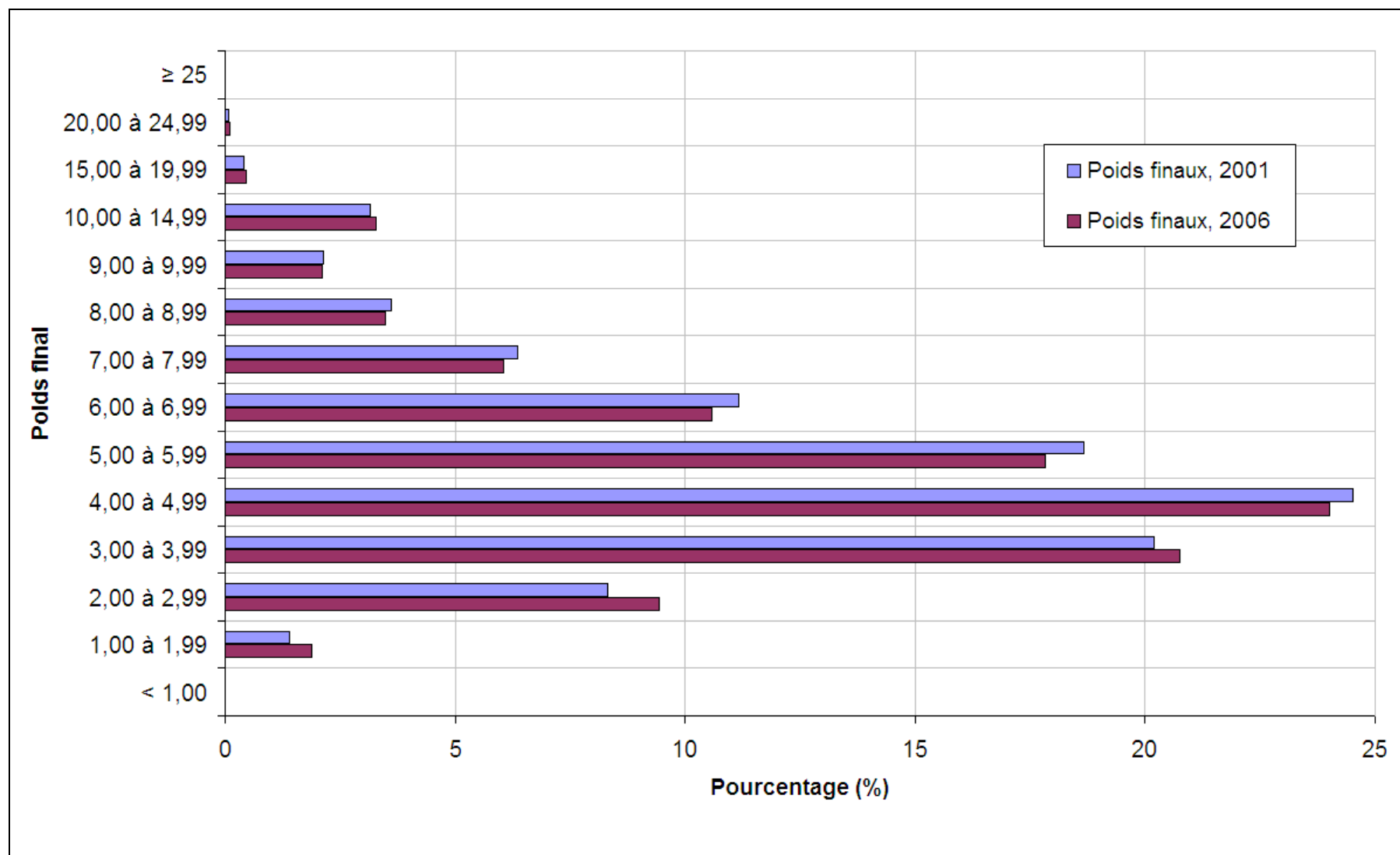
7.2 Évaluation de la méthode de pondération du recensement

7.2.1 Répartition des poids

Le [graphique 7.2.1.1](#) permet de comparer la répartition des poids finaux de 2006 avec celle de 2001. Bien que les répartitions soient presque identiques, le graphique indique qu'un peu plus de ménages avaient des poids inférieurs à 4 en 2006 qu'en 2001. En revanche, les ménages ayant des poids entre 4 et 9 étaient moins nombreux en 2006 qu'en 2001.

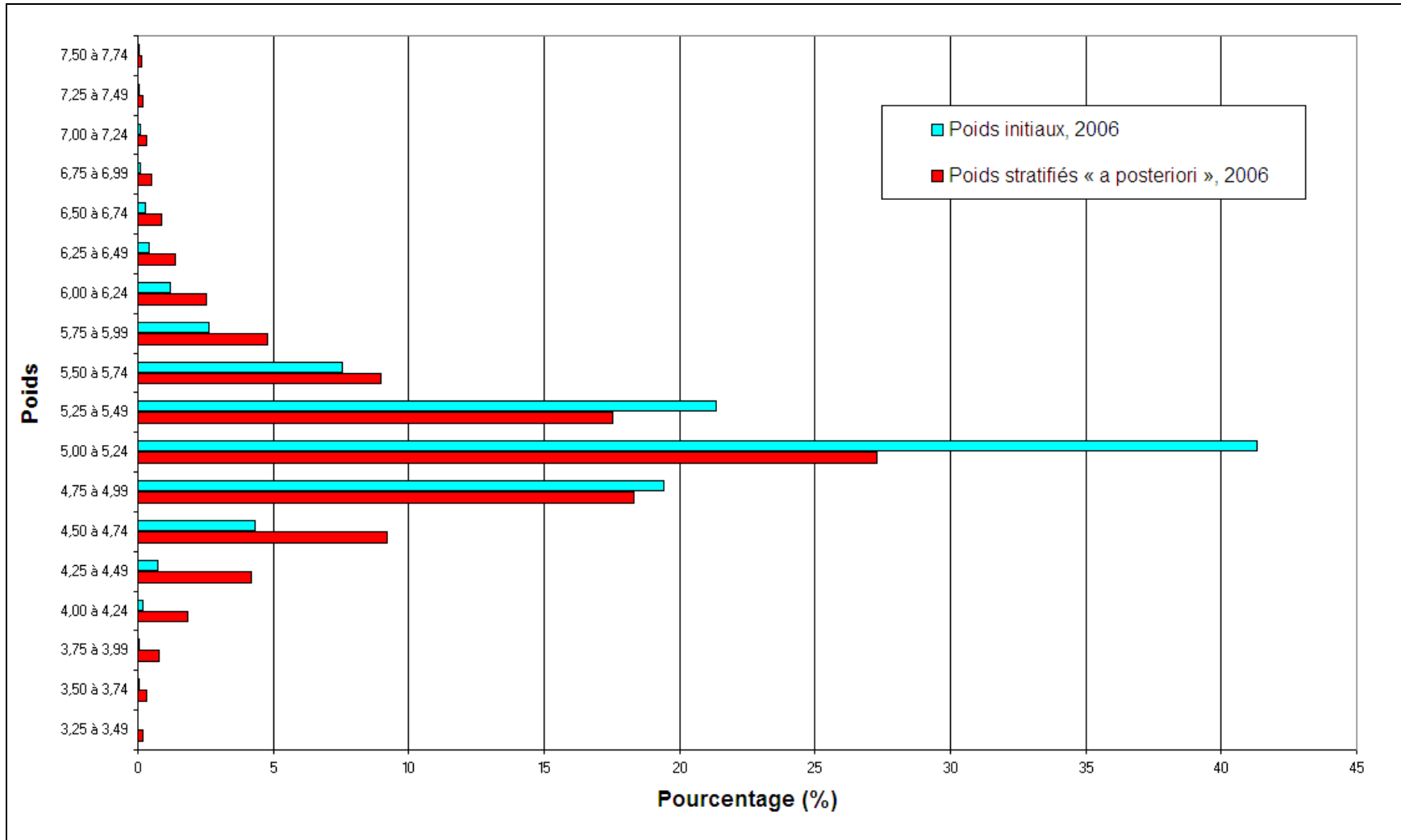
Les graphiques [7.2.1.2](#), [7.2.1.3](#) et [7.2.1.4](#) permettent de comparer les distributions des poids initiaux, stratifiés « *a posteriori* », obtenus à la première étape et finaux du Recensement de 2006. Les poids initiaux sont fortement regroupés autour de 5, étant donné que les ménages sont échantillonnés au 1/5. Les distributions des poids stratifiés « *a posteriori* », obtenus à la première étape et finaux deviennent progressivement plus étendues, à mesure que les contraintes deviennent plus restrictives.

Graphique 7.2.1.1 Comparaison des poids finaux associés aux ménages des recensements de 2006 et de 2001



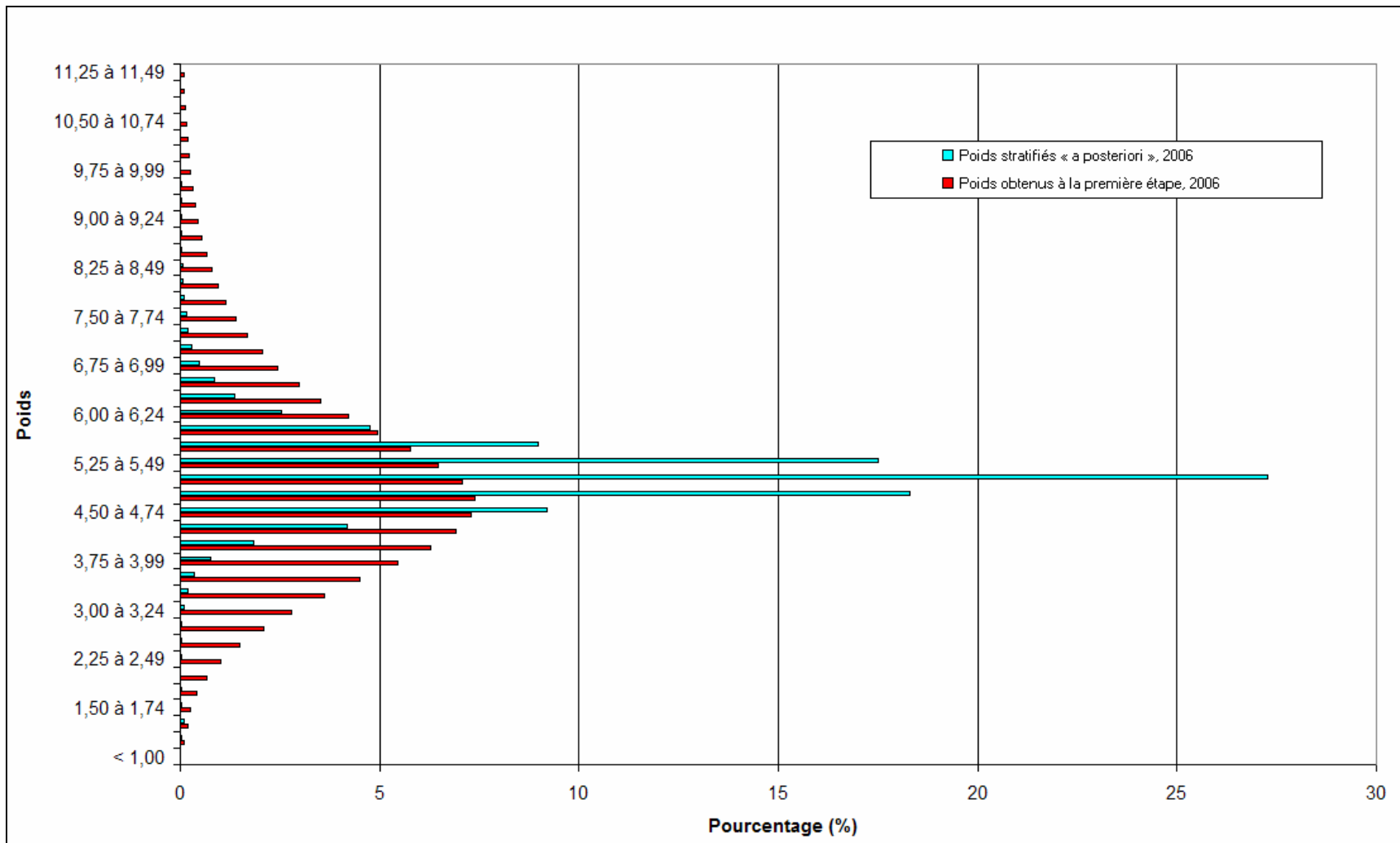
Sources : Statistique Canada, recensements de 2006 et 2001.

Graphique 7.2.1.2 Comparaison des poids initiaux et des poids stratifiés « *a posteriori* », Recensement de 2006



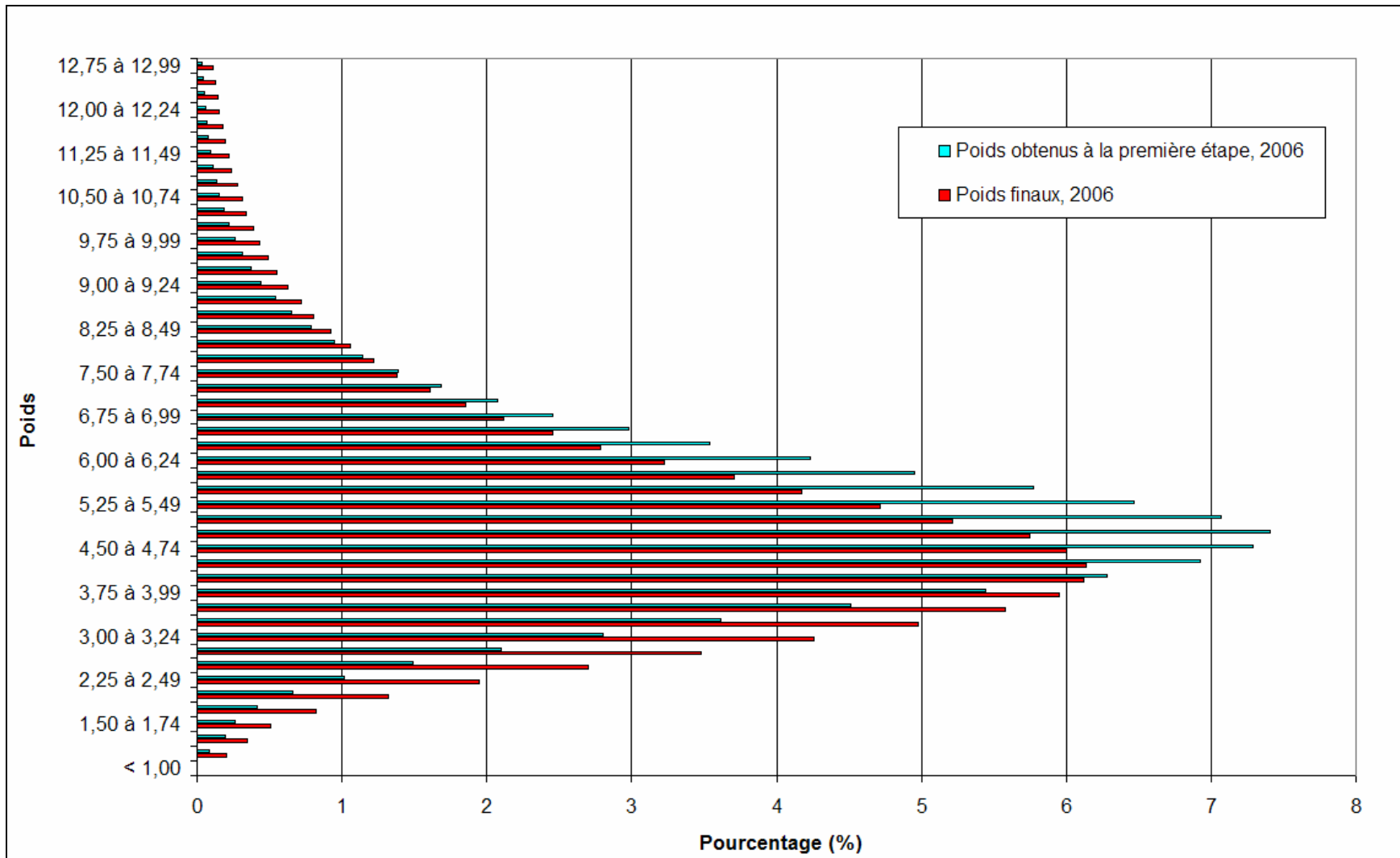
Source : Statistique Canada, Recensement de 2006.

Graphique 7.2.1.3 Comparaison des poids stratifiés « a posteriori » et des poids obtenus à la première étape, Recensement de 2006



Source : Statistique Canada, Recensement de 2006.

Graphique 7.2.1.4 Comparaison des poids obtenus à la première étape et des poids finaux, Recensement de 2006



Source : Statistique Canada, Recensement de 2006.

7.2.2 Écarts entre les chiffres de population et les estimations-échantillon

Comme nous l'avons mentionné à la [section 4.4](#), les poids finaux sont choisis de manière à réduire, ou à éliminer, les écarts entre les chiffres de population et les estimations-échantillon correspondantes pour 34 contraintes au niveau des RP (voir l'[annexe B](#)). Certains écarts peuvent toutefois subsister étant donné que les contraintes sont parfois retranchées (voir les [sections 4.4](#) et [7.2.3](#)). La différence entre les chiffres de population et les estimations-échantillon se définit comme suit :

$$\text{écart population/estimation} = \frac{\text{estimation-échantillon} - \text{chiffre de population}}{\text{chiffre de population}} \times 100$$

Le numérateur de cette expression (estimation-échantillon – chiffre de population) est appelé différence population/estimation. La comparaison entre les estimations-échantillon et les chiffres de population est fondée sur les logements privés occupés dans les UC échantillonnées.

Le [tableau 7.2.2.1](#) ainsi que les [graphiques 7.2.2.1](#) et [7.2.2.2](#) montrent les différences et les écarts population/estimation observés à l'échelle du Canada en 2006 et en 2001 pour les 34 contraintes au niveau des RP au moyen des poids initiaux ou finaux. Le [graphique 7.2.2.1](#) est semblable au [graphique 6.1](#), puisqu'il est fondé sur les poids initiaux, mais il révèle les écarts entre la population et les estimations plutôt que les valeurs Z. Par conséquent, une grande partie de la discussion sur le [graphique 6.1](#) s'applique au [graphique 7.2.2.1](#).

Le [tableau 7.2.2.1](#) et le [graphique 7.2.2.1](#) indiquent également ce que le biais d'échantillonnage aurait été en 2006 si l'approche de 2001, axée sur la conversion des documents plutôt que sur l'imputation de ménages au complet (IMC), avait été utilisée pour traiter les questionnaires complets ayant une non-réponse totale. Pour répondre à cette question, les questionnaires complets ayant fait l'objet de l'imputation de ménages au complet ont été traités comme des questionnaires abrégés, et les poids initiaux ont été recalculés au niveau des UC pour tenir compte de ce fait. Les poids initiaux recalculés ont été appliqués à l'échantillon réduit pour créer de nouvelles différences entre la population et les estimations, qui figurent à la colonne appelée « sans IMC » au [tableau 7.2.2.1](#). Ces différences apparaissent également comme des écarts dans le [graphique 7.2.2.1](#) et sont appelées « 2006 sans IMC » dans la légende. Les différences entre la population et les estimations selon la méthode de l'imputation de ménages au complet au moyen des poids initiaux originaux et l'échantillon non réduit sont indiquées dans la colonne intitulée « avec IMC » au [tableau 7.2.2.1](#) et « 2006 avec IMC » dans la légende du [graphique 7.2.2.1](#).

En général, on constate que les différences de 2006 « sans IMC » du [tableau 7.2.2.1](#) sont beaucoup plus semblables aux différences de 2001 qu'à celles de 2006 « avec IMC ». De plus, les différences entre la population et les estimations sont souvent plus petites « avec IMC » en 2006 que « sans IMC » en 2006 (p. ex., c'est le cas des personnes de sexe féminin, des personnes âgées de moins de 15 ans ou celles âgées de 45 ans et plus; des personnes mariées, veuves, divorcées ou séparées; des ménages composés de 1, 2, 4 ou 5 personnes; des logements individuels non attenants et des appartements de moins de cinq

étages). Par conséquent, l'introduction de l'imputation de ménages au complet en 2006 pour traiter les ménages affichant une non-réponse totale s'était avérée bénéfique dans l'ensemble.

Bien qu'ils ne soient pas indiqués au [tableau 7.2.2.1](#) et au [graphique 7.2.2.1](#), les poids initiaux pour l'échantillon réduit ont été calculés une deuxième fois séparément pour les ménages composés de 1, 2, 3, 4, 5 et 6 personnes ou plus au niveau de l'UC. Conformément à cette approche, les différences de 2006 « sans IMC » étaient beaucoup plus semblables aux différences de 2006 « avec IMC », ce qui porte à croire que l'imputation de ménages au complet donne des résultats semblables à ceux qu'on aurait obtenu au moyen de la conversion des documents et des poids initiaux si ces derniers avaient été stratifiés « *a posteriori* » selon la taille du ménage.

Le [tableau 7.2.2.1](#) montre que les valeurs absolues des écarts population/estimation d'après les poids finaux en 2006 sont sensiblement plus grandes qu'en 2001 pour les groupes d'âge de 15 à 19 ans et de 25 à 34 ans, mais semblables ou plus petites pour les autres groupes d'âge. Les écarts absolus en 2006 sont également plus prononcés dans le cas des ménages composés de 4, 5 ou 6 personnes ou plus. Comme mentionné au [chapitre 6](#), ce phénomène est probablement attribuable en grande partie au fait que le nombre de personnes sur le questionnaire papier 2B a été réduit de six à cinq en 2006. Lorsqu'on compare les [graphiques 7.2.2.1](#) et [7.2.2.2](#), il est clair que les écarts population/estimation de 2006 fondés sur les poids finaux sont nettement plus faibles que ceux fondés sur les poids initiaux, sauf dans le cas des ménages de 5 personnes et de 6 personnes ou plus. Ce phénomène est probablement attribuable à la difficulté de corriger des biais initiaux aussi prononcés, tout en corrigeant parallèlement les contraintes restantes. Les écarts pour ces deux contraintes demeurent significativement réduits lorsque les poids finaux sont utilisés au lieu des poids initiaux. Par ailleurs, il convient de souligner que les écarts fondés sur les poids finaux pour les deux types de caractéristiques du logement (logements individuels non attenants et appartements de moins de cinq étages) ont été considérablement réduits par rapport à ceux qui étaient fondés sur les poids initiaux, malgré l'absence d'un tel contrôle dans toutes les RP. La réduction de l'écart pour ces caractéristiques a probablement fait augmenter l'écart pour d'autres caractéristiques qui ont été éliminées à leur place. Les conséquences exactes sur les autres caractéristiques ne peuvent être observées en raison des nombreux facteurs en cause pour leur création. Le [graphique 7.2.2.2-A](#) est le même que le [graphique 7.2.2.2](#), sauf qu'il est agrandi de manière à ce que les écarts puissent être visualisés plus facilement pour les autres contraintes. Le [graphique 7.2.2.2-A](#) révèle que, hormis les contraintes relatives à la taille du ménage, la contrainte « Union libre = oui » affiche l'écart le plus prononcé.

Le [tableau 7.2.2.2](#) et le [graphique 7.2.2.3](#) présentent les résultats des premier et deuxième passages quant aux différences et aux écarts population/estimation de 2006 fondés sur les poids finaux pour les 34 contraintes au niveau des RP, pour le Canada (voir la [section 4.5](#)). Les écarts du premier passage sont moins prononcés que ceux du deuxième passage, puisque les poids du recensement ont été calculés en fonction des résultats du premier passage. Le [graphique 7.2.2.4](#) examine la différence entre les résultats du premier passage et du deuxième passage pour les recensements de 2006 et de 2001. Il indique que, à l'exception des contraintes relatives à l'union libre, au veuvage et à la séparation, l'écart entre les estimations du premier passage et du deuxième passage est beaucoup moins prononcé en 2006 qu'en 2001. Ce phénomène est peut-être en partie attribuable au processus d'imputation de ménages au complet, qui pourrait avoir donné lieu à une cohérence accrue entre les données du premier passage et du deuxième passage par rapport à 2001.

Le [tableau 7.2.2.2](#) démontre qu'il n'y a aucune différence entre la population et l'estimation pour le nombre total de personnes dans les résultats du premier passage et du deuxième passage. Il faudrait prendre note que ceci représente une combinaison de personnes des ménages privés et des unités composées de personnes âgées. Toutefois, lorsqu'on observe les résultats des premier et deuxième passages pour ces deux univers séparément, la population totale pour les ménages privés est alors surestimée de 1 982 personnes et la population totale pour les unités composées de personnes âgées est sous-estimée du même nombre.

Le [tableau 7.2.2.3](#) présente les chiffres et les estimations pour les trois univers distincts pour lesquels les données du recensement peuvent être observées. La [section 4.7](#) comporte une description plus détaillée à ce sujet. Le présent rapport sur la pondération s'intéresse principalement aux données qui proviennent de l'univers des logements privés. Le [tableau 7.2.2.3](#) révèle la différence entre les chiffres de population et les estimations lorsque les logements collectifs et les établissements institutionnels sont pris en compte, puisque ces derniers sont inclus dans les tableaux publiés du recensement.

Tableau 7.2.2.1 Comparaison des écarts population/estimation des recensements de 2001 et de 2006, pour le Canada

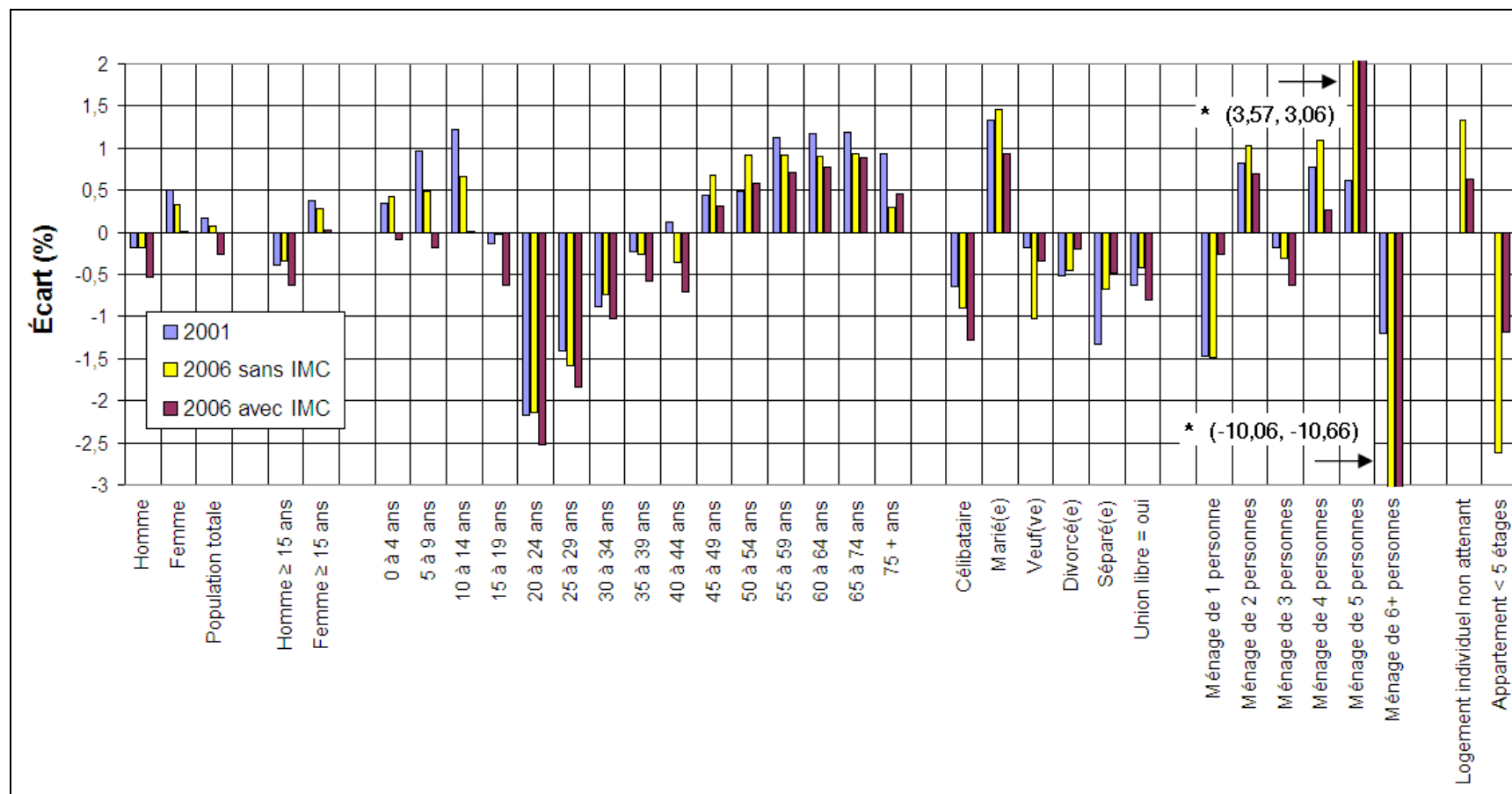
Caractéristique	Recensement de 2006				Recensement de 2001		
	Différences entre les poids initiaux**		Poids finaux		Poids initiaux	Poids finaux	
	Avec IMC	Sans IMC	Différence	Écart	Différence	Différence	Écart
Homme	-79 381	-27 123	-49	0,00	-25 074	0	0,00
Femme	923	50 176	49	0,00	73 397	0	0,00
Homme ≥ 15 ans	-76 328	-41 932	116	0,00	-44 291	51	0,00
Personne ≥ 15 ans	-74 220	-5 817	243	0,00	130	0	0,00
Total des ménages	-1 056	-1 245	0	0,00	-1 040	0	0,00
Population totale	-78 458	23 053	0	0,00	48 324	0	0,00
0 à 4 ans	-1 354	6 915	-68	0,07	5 628	559	0,03
5 à 9 ans	-3 126	8 560	-360	-0,02	18 245	-792	-0,04
10 à 14 ans	242	13 395	186	0,01	24 321	234	0,01
15 à 19 ans	-13 108	-319	1 856	0,09	-2 644	779	0,04
20 à 24 ans	-51 382	-43 328	-226	0,01	-41 081	-504	-0,03
25 à 29 ans	-35 659	-30 595	-1 359	0,07	-25 620	-785	-0,04
30 à 34 ans	-20 062	-14 450	-359	-0,02	-17 888	7	0,00
35 à 39 ans	-12 439	-5 644	391	0,02	-5 675	-556	-0,02
40 à 44 ans	-17 995	-8 963	348	0,01	3 073	100	0,00
45 à 49 ans	8 286	17 325	328	0,01	10 024	687	0,03
50 à 54 ans	13 732	21 134	342	0,01	10 004	-87	0,00
55 à 59 ans	14 614	18 772	249	0,01	17 396	81	0,01
60 à 64 ans	11 940	14 123	-522	-0,03	14 459	933	0,08
65 à 74 ans	19 541	20 930	300	0,01	24 283	271	0,01
75+ ans	8 312	5 197	-1 105	-0,06	13 798	-926	-0,06
Célibataire	-180 891	-127 391	-42	0,00	-86 671	-53	0,00
Marié(e)	114 966	179 374	-184	0,00	156 112	-57	0,00
Veuf(ve)	-4 846	-14 702	-252	-0,02	-2 388	557	0,04
Divorcé(e)	-4 019	-9 166	141	0,01	-9 375	206	0,01
Séparé(e)	-3 667	-5 061	336	0,04	-9 355	-653	-0,09
Union libre = oui	-21 921	-11 523	2 995	0,11	-14 381	4 115	0,18
Ménage de 1 personne	-8 705	-49 455	-3 370	-0,10	-42 675	-4 175	-0,14
Ménage de 2 personnes	29 091	42 615	-1 658	-0,04	30 499	-906	-0,02
Ménage de 3 personnes	-12 081	-5 989	-3 727	-0,19	-3 405	-5 010	-0,27
Ménage de 4 personnes	4 871	20 108	6 025	0,33	14 138	2 414	0,13
Ménage de 5 personnes	21 817	25 493	15 862	2,22	4 395	8 818	1,23
Ménage de 6+ personnes	-36 049	-34 017	-13 132	-3,88	-3 991	-1 142	-0,34
Logement individuel non attenant	42 896	89 944	8 648	0,13
Appartement < 5 étages	-27 066	-60 002	-1 089	-0,05

** L'imputation de ménages au complet a été utilisée dans le calcul des poids du Recensement de 2006.

.. indisponible pour une période de référence précise

Sources : Statistique Canada, recensements de 2006 et 2001.

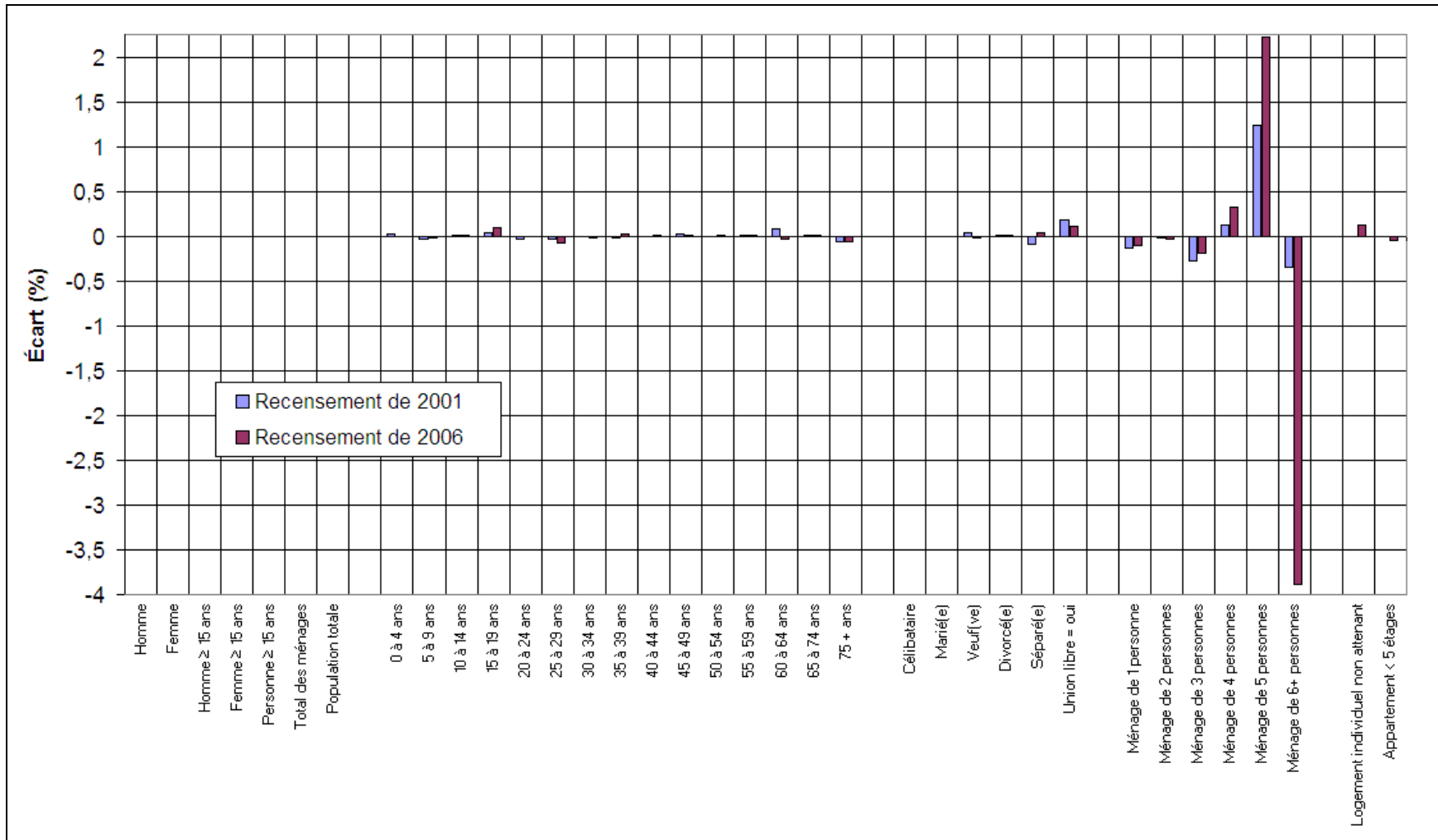
Graphique 7.2.2.1 Comparaison des écarts entre les poids initiaux avec et sans imputation de ménages au complet



* Les valeurs insérées dans le graphique se rapportent aux barres qui dépassent la limite du graphique. Les valeurs correspondent aux barres de gauche à droite.

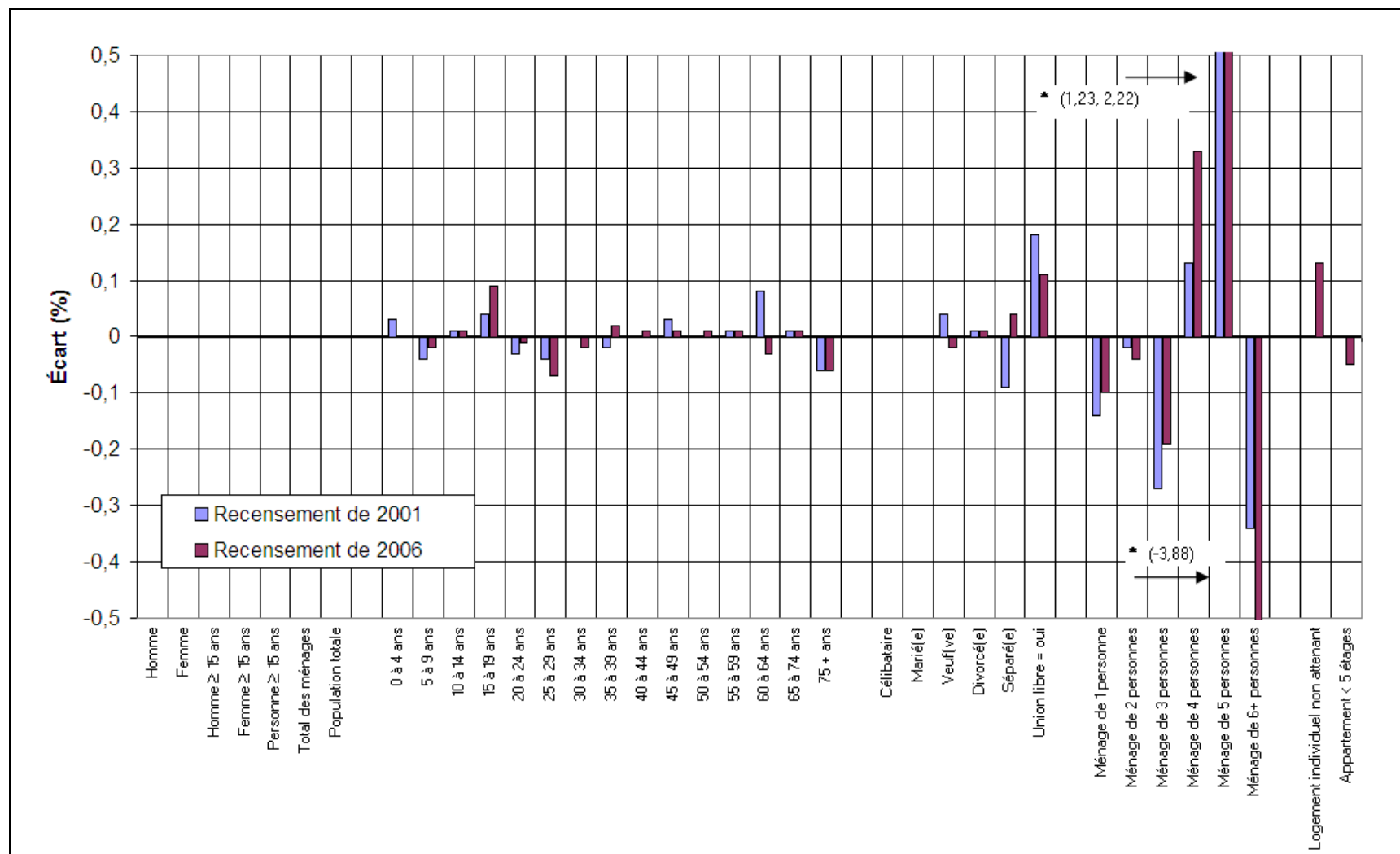
Sources : Statistique Canada, recensements de 2006 et 2001.

Graphique 7.2.2.2 Écarts population/estimation fondés sur les poids finaux



Sources : Statistique Canada, Recensements de 2006 et 2001.

Graphique 7.2.2.2-A Écarts population/estimation fondés sur les poids finaux (agrandissement)



* Les valeurs insérées dans le graphique se rapportent aux barres qui dépassent la limite du graphique. Les valeurs correspondent aux barres de gauche à droite.

Sources : Statistique Canada, recensements de 2006 et 2001.

Tableau 7.2.2.2 Comparaison des écarts population/estimation entre le premier passage et le deuxième passage fondés sur les poids finaux, pour le Canada, Recensement de 2006

Caractéristique	Recensement de 2006 : Passage 1				Recensement de 2006 : Passage 2				Passage 2 – Passage 1	
	Chiffre	Estimation	Différence ¹	Écart ²	Chiffre	Estimation	Différence ¹	Écart ²	Différence ³	Écart ⁴
Homme	15 041 551	15 041 502	-49	0	15 041 422	15 040 736	-686	0	-637	0
Femme	15 653 041	15 653 090	49	0	15 653 170	15 653 856	686	0	637	0
Population totale	30 694 592	30 694 592	0	0	30 694 592	30 694 592	0	0	0	0
Homme ≥ 15 ans	12 263 445	12 263 561	116	0	12 263 025	12 260 877	-2 148	-0,02	-2 264	-0,02
Femme ≥ 15 ans	13 005 067	13 005 193	126	0	13 004 908	13 004 626	-282	0	-408	0
Population totale ≥ 15 ans	25 268 512	25 268 754	242	0	25 267 933	25 265 503	-2 430	-0,01	-2 672	-0,01
0 à 4 ans	1 640 859	1 640 791	-68	0	1 641 102	1 642 103	1 001	0,06	1 069	0,07
5 à 9 ans	1 760 005	1 759 645	-360	-0,02	1 760 149	1 760 221	72	0	432	0,02
10 à 14 ans	2 025 216	2 025 402	186	0,01	2 025 408	2 026 764	1 356	0,07	1 170	0,06
15 à 19 ans	2 083 373	2 085 229	1 856	0,09	2 083 289	2 084 962	1 673	0,08	-183	-0,01
20 à 24 ans	2 029 449	2 029 223	-226	-0,01	2 029 402	2 028 899	-503	-0,02	-277	-0,01
25 à 29 ans	1 940 880	1 939 521	-1 359	-0,07	1 940 768	1 939 125	-1 643	-0,08	-284	-0,01
30 à 34 ans	1 976 478	1 976 119	-359	-0,02	1 976 505	1 976 221	-284	-0,01	75	0
35 à 39 ans	2 161 430	2 161 821	391	0,02	2 161 366	2 161 265	-101	0	-492	-0,02
40 à 44 ans	2 559 477	2 559 825	348	0,01	2 559 271	2 558 507	-764	-0,03	-1 112	-0,04
45 à 49 ans	2 571 429	2 571 757	328	0,01	2 571 359	2 571 322	-37	0	-365	-0,01
50 à 54 ans	2 313 657	2 313 999	342	0,01	2 313 669	2 314 099	430	0,02	88	0
55 à 59 ans	2 045 868	2 046 117	249	0,01	2 045 821	2 045 799	-22	0	-271	-0,01
60 à 64 ans	1 558 145	1 557 623	-522	-0,03	1 558 054	1 557 336	-718	-0,05	-196	-0,01
65 à 74 ans	2 229 023	2 229 323	300	0,01	2 229 016	2 229 129	113	0,01	-187	-0,01
75 ans et plus	1 799 303	1 798 198	-1 105	-0,06	1 799 413	1 798 841	-572	-0,03	533	0,03
Célibataire	14 170 280	14 170 238	-42	0	14 170 125	14 168 822	-1 303	-0,01	-1 261	-0,01
Marié(e)	12 291 457	12 291 273	-184	0	12 291 559	12 291 750	191	0	375	0
Veuf(ve)	1 435 852	1 435 600	-252	-0,02	1 435 992	1 436 293	301	0,02	553	0,04
Divorcé(e)	2 044 164	2 044 305	141	0,01	2 044 209	2 045 165	956	0,05	815	0,04
Séparé(e)	752 839	753 175	336	0,04	752 707	752 563	-144	-0,02	-480	-0,06
Union libre = oui	2 725 161	2 728 156	2 995	0,11	2 726 070	2 733 383	7 313	0,27	4 318	0,16
Ménage de 1 personne	3 338 596	3 335 226	-3 370	-0,10	3 338 596	3 335 226	-3 370	-0,10	0	0
Ménage de 2 personnes	4 153 415	4 151 757	-1 658	-0,04	4 153 415	4 151 757	-1 658	-0,04	0	0
Ménage de 3 personnes	1 963 201	1 959 474	-3 727	-0,19	1 963 201	1 959 474	-3 727	-0,19	0	0
Ménage de 4 personnes	1 843 987	1 850 012	6 025	0,33	1 843 987	1 850 012	6 025	0,33	0	0
Ménage de 5 personnes	713 994	729 856	15 862	2,22	713 994	729 856	15 862	2,22	0	0
Ménage de 6+ personnes	338 241	325 109	-13 132	-3,88	338 241	325 109	-13 132	-3,88	0	0
Logement individuel non attenant	6 769 581	6 778 229	8 648	0,13	6 769 581	6 778 229	8 648	0,13	0	0
Appartement < 5 étages	2 285 965	2 284 876	-1 089	-0,05	2 285 965	2 284 876	-1 089	-0,05	0	0

Notes :

1. Différence : estimation - chiffre.

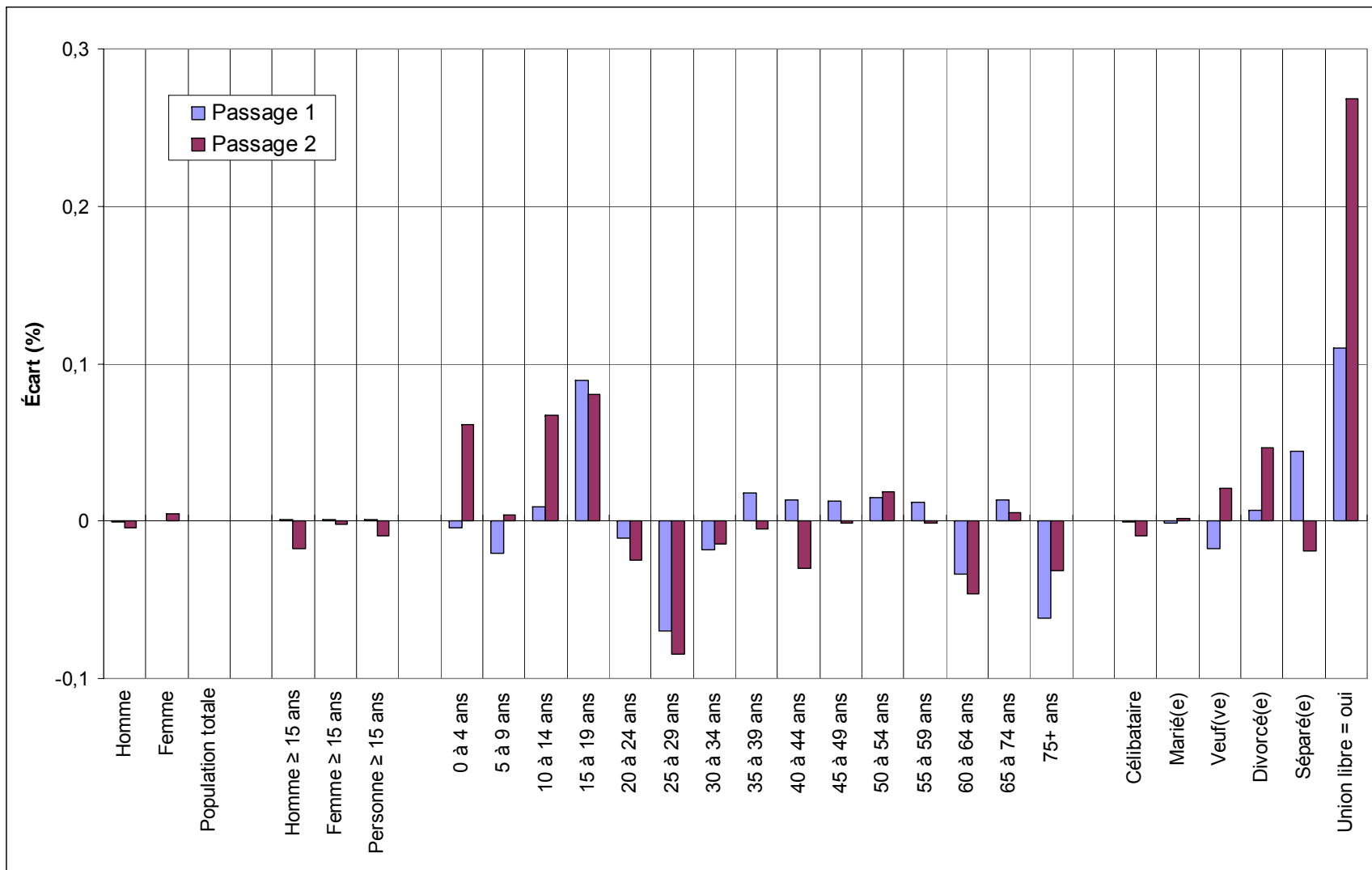
3. Différence : différence Passage 2 - différence Passage 1.

2. Écart : (100 * [estimation - chiffre]/chiffre).

4. Écart : (100 * [différence Passage 2 - différence Passage 1]/différence Passage 1).

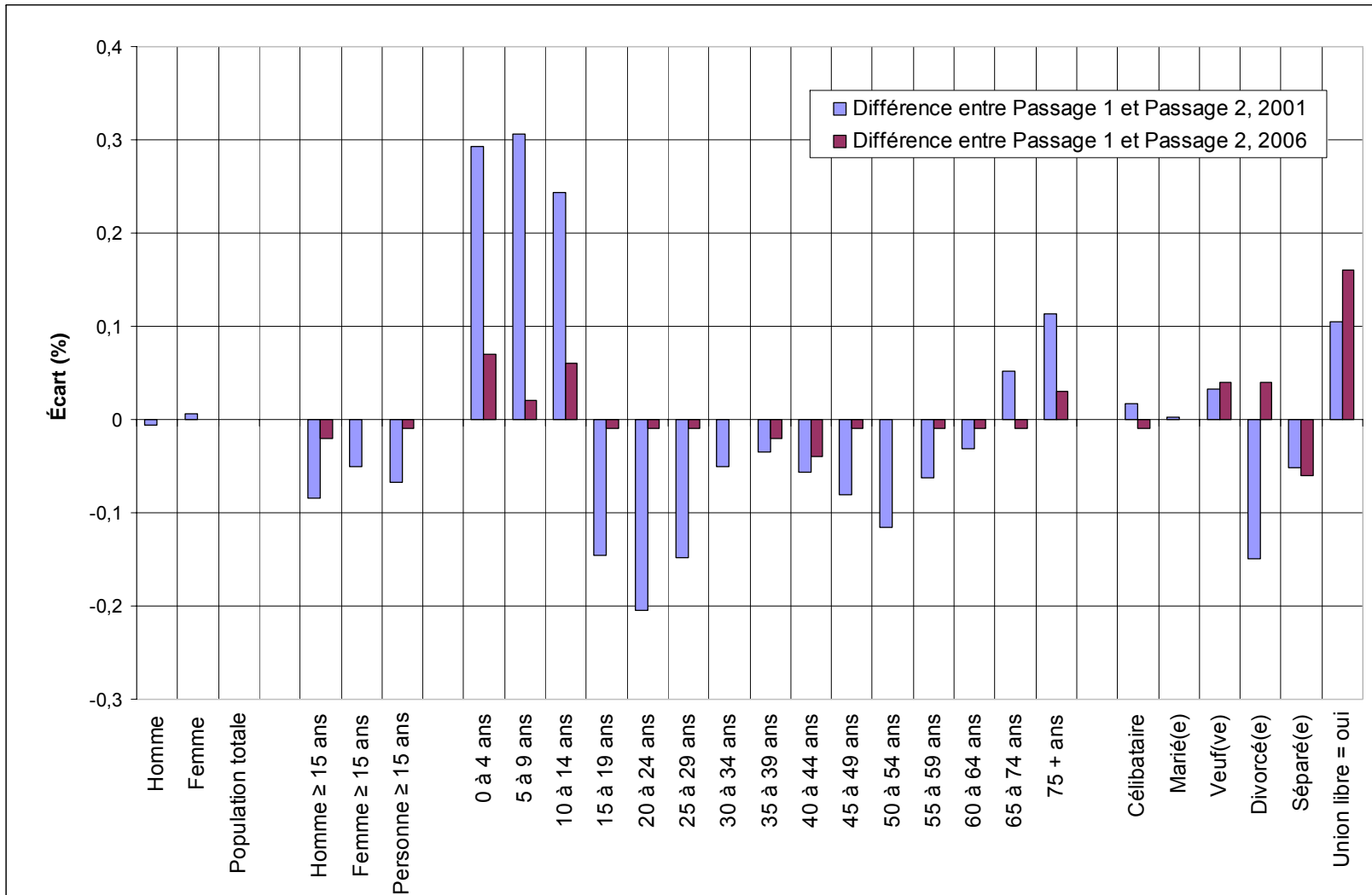
Source : Statistique Canada, Recensement de 2006.

Graphique 7.2.2.3 Comparaison des écarts population/estimation entre le premier passage et le deuxième passage fondés sur les poids finaux, pour le Canada, Recensement de 2006



Source : Statistique Canada, Recensement de 2006.

Graphique 7.2.2.4 Comparaison des écarts population/estimation, différences entre le premier passage et le deuxième passage, recensements de 2006 et de 2001



Sources : Statistique Canada, recensements de 2006 et 2001.

Tableau 7.2.2.3 Comparaison des univers – Chiffres de population et estimations, Recensement de 2006

Caractéristique	Passage 2 – Logements privés		Passage 2 – Logements privés/collectifs		Passage 2 – Logements privés/collectifs/établissements institutionnels		Établissements institutionnels
	Chiffre	Estimation	Chiffre	Estimation	Chiffre	Estimation	
Homme	15 041 422	15 040 736	15 326 954	15 326 268	15 475 970	15 475 284	149 016
Femme	15 653 170	15 653 856	15 914 076	15 914 762	16 136 927	16 137 613	222 851
Population totale	30 694 592	30 694 592	31 241 030	31 241 030	31 612 897	31 612 897	371 867
Homme ≥ 15 ans	12 263 025	12 260 877	12 472 933	12 470 785	12 618 649	12 616 501	145 716
Femme ≥ 15 ans	13 004 908	13 004 626	13 193 720	13 193 438	13 414 410	13 414 128	220 690
Personne ≥ 15 ans	25 267 933	25 265 503	25 666 653	25 664 223	26 033 059	26 030 629	366 406
0 à 4 ans	1 641 102	1 642 103	1 689 395	1 690 396	1 690 539	1 691 540	1 144
5 à 9 ans	1 760 149	1 760 221	1 808 205	1 808 277	1 809 373	1 809 445	1 168
10 à 14 ans	2 025 408	2 026 764	2 076 777	2 078 133	2 079 926	2 081 282	3 149
15 à 19 ans	2 083 289	2 084 962	2 134 246	2 135 919	2 140 493	2 142 166	6 247
20 à 24 ans	2 029 402	2 028 899	2 072 397	2 071 894	2 080 384	2 079 881	7 987
25 à 29 ans	1 940 768	1 939 125	1 977 418	1 975 775	1 985 580	1 983 937	8 162
30 à 34 ans	1 976 505	1 976 221	2 011 584	2 011 300	2 020 228	2 019 944	8 644
35 à 39 ans	2 161 366	2 161 265	2 197 813	2 197 712	2 208 273	2 208 172	10 460
40 à 44 ans	2 559 271	2 558 507	2 597 151	2 596 387	2 610 458	2 609 694	13 307
45 à 49 ans	2 571 359	2 571 322	2 606 788	2 606 751	2 620 598	2 620 561	13 810
50 à 54 ans	2 313 669	2 314 099	2 344 230	2 344 660	2 357 304	2 357 734	13 074
55 à 59 ans	2 045 821	2 045 799	2 072 074	2 072 052	2 084 621	2 084 599	12 547
60 à 64 ans	1 558 054	1 557 336	1 578 195	1 577 477	1 589 868	1 589 150	11 673
65 à 74 ans	2 229 016	2 229 129	2 255 529	2 255 642	2 288 363	2 288 476	32 834
75+ ans	1 799 413	1 798 841	1 819 228	1 818 656	2 046 889	2 046 317	227 661
Célibataire	14 170 125	14 168 822	14 541 272	14 539 969	14 666 870	14 665 567	125 598
Marié(e)	12 291 559	12 291 750	12 415 528	12 415 719	12 470 398	12 470 589	54 870
Veuf(ve)	1 435 992	1 436 293	1 451 805	1 452 106	1 612 819	1 613 120	161 014
Divorcé(e)	2 044 209	2 045 165	2 066 245	2 067 201	2 087 387	2 088 343	21 142
Séparé(e)	752 707	752 563	766 180	766 036	775 423	775 279	9 243
Union libre = oui	2 726 070	2 733 383	2 789 627	2 796 940	2 789 627	2 796 940	0
Ménage de 1 personne	3 338 596	3 335 226	3 367 367	3 363 997	3 367 367	3 363 997	0
Ménage de 2 personnes	4 153 415	4 151 757	4 186 669	4 185 011	4 186 669	4 185 011	0
Ménage de 3 personnes	1 963 201	1 959 474	1 985 694	1 981 967	1 985 694	1 981 967	0
Ménage de 4 personnes	1 843 987	1 850 012	1 865 954	1 871 979	1 865 954	1 871 979	0
Ménage de 5 personnes	713 994	729 856	730 325	746 187	730 325	746 187	0
Ménage de 6+ personnes	338 241	325 109	370 002	356 870	370 002	356 870	0
Logement individuel non attenant	6 769 581	6 778 229	6 871 318	6 879 966	6 871 318	6 879 966	0
Appartement < 5 étages	2 285 965	2 284 876	2 289 388	2 329 459	2 289 388	2 329 459	0

Source : Statistique Canada, Recensement de 2006.

7.2.3 Retranchement des contraintes

Dans le cadre du Recensement de 2006, on a examiné 20 ensembles de combinaisons de paramètres au moyen du système de pondération pour chaque région de pondération (RP), puis on a sélectionné l'ensemble de paramètres qui a donné lieu aux meilleurs résultats pour chacune des RP (voir la [section 4.4](#)).

L'[annexe B](#) présente une liste complète des 34 contraintes utilisées. Trente-deux de ces contraintes faisaient partie de chacune des mises à l'essai des divers ensembles de paramètres. Deux de ces contraintes, relatives aux logements individuels non attenants et aux appartements des immeubles de moins de cinq étages, étaient nouvelles en 2006 et ont seulement été ajoutées comme contraintes pour certaines combinaisons de paramètres.

Le [tableau 7.2.3.1](#) montre la fréquence à laquelle chacune des 34 contraintes a été retranchée dans les 6 602 RP échantillonnées en 2006 et les 6 141 RP échantillonnées en 2001. La raison du retranchement d'une contrainte (c.-à-d. parce qu'elle est petite, linéairement dépendante ou quasi linéairement dépendante, ou parce qu'elle cause des poids aberrants [se reporter à la [section 4.4](#)]) permet de mieux comprendre l'ampleur des écarts population/estimation de certaines contraintes fournies dans le [graphique 7.2.2.2](#). L'analyse qui suit portera surtout sur les résultats du Recensement de 2006. D'abord, on note qu'une contrainte telle que « Âge 0 à 4 » peut être retranchée fréquemment parce qu'elle est linéairement dépendante (ce qui signifie « redondante » et qu'elle peut quand même donner une légère différence population/estimation. De plus, si une contrainte comme « Union libre = oui » ou « Ménages de 5 personnes » est retranchée fréquemment parce qu'elle cause des poids aberrants ou parce qu'elle est quasi linéairement dépendante (comme c'est le cas pour les ménages de 1, de 3 ou de 4 personnes); cela peut causer de grands écarts population/estimation, comme on peut le voir dans le [graphique 7.2.2.2](#).

Les deux contraintes relatives au type de logement (logements individuels non attenants et appartements de moins de cinq étages) étaient nouvelles en 2006 et ont été traitées différemment des 32 contraintes qui étaient aussi utilisées en 2001. Le niveau de non-réponse pour la variable du type de logement a été analysé au niveau des AD. Ces deux contraintes ont été automatiquement éliminées pour 399 RP qui englobaient une AD ayant été associée à un niveau significatif de non-réponse pour cette variable, ce qui rendrait peu fiables les estimations pour ces caractéristiques. Pour les 6 203 RP restantes, l'utilisation de ces contraintes a été incluse comme paramètre. Dans le cas de 10 des 20 combinaisons de paramètres pour lesquelles les RP avaient été traitées, on s'efforçait de tenir compte de ces caractéristiques. Pendant le processus de sélection choisie, on a sélectionné les poids finaux de 3 688 RP à partir d'une combinaison de paramètres qui avait pour objet de tenir compte de ces deux contraintes. Par exemple, la contrainte relative aux logements individuels non attenants a été éliminée du système de pondération pour 304 RP seulement, mais elle a tout de même été prise en compte pour 3 384 RP.

Le [tableau 7.2.3.2](#) résume les renseignements qui figurent au [tableau 7.2.3.1](#). Le nombre total de contraintes éliminées est plus élevé en 2006, puisque les RP sont plus nombreuses (6 602

RP en 2006, 6 141 RP en 2001), mais le nombre moyen de contraintes au niveau des RP éliminées par RP est demeuré relativement constant entre 2001 et 2006.

Le [tableau 7.2.3.2](#) présente aussi un résumé des chiffres relatifs à la fréquence de retranchement des contraintes « nombre de ménages » et « nombre de personnes » au niveau des AD. Supposons, à titre d'exemple, qu'une RP renferme 8 AD, elle compterait alors 16 contraintes au niveau des AD. Dans l'ensemble, le nombre moyen de contraintes au niveau des AD ayant été retranchées a diminué (0,8 en 2006, 1,1 en 2001). La baisse la plus marquée a été observée dans la catégorie PETITE, où seulement 248 contraintes ont été retranchées en 2006, comparativement à 1 354 en 2001. Cette situation est en partie attribuable au choix d'un plus grand nombre d'ensembles de paramètres.

Tableau 7.2.3.1 Fréquence à laquelle des contraintes au niveau des régions de pondération ont été retranchées en 2001 et en 2006 lors de l'ajustement des poids finaux

Caractéristique	Recensement de 2006					Recensement de 2001				
	Petite	LD	QLD	Aber.	Total	Petite	LD	QLD	Aber.	Total
Homme	0	0	0	1	1	0	0	0	0	0
Femme	0	0	0	0	0	0	0	0	0	0
Population totale	0	0	0	1	1	0	0	0	0	0
Homme ≥ 15 ans	0	3	22	51	76	0	4	24	27	55
Personne ≥ 15 ans	0	1	0	3	4	0	0	0	0	0
Total des ménages	0	0	0	0	0	0	0	0	0	0
0 à 4 ans	16	4 342	6	191	4 555	29	4 286	2	124	4 441
5 à 9 ans	60	1 217	11	342	1 630	68	406	4	251	729
10 à 14 ans	66	961	5	279	1 311	79	1 359	2	141	1 581
15 à 19 ans	12	678	7	142	839	18	492	6	131	647
20 à 24 ans	2	228	18	150	398	2	243	15	125	385
25 à 29 ans	2	1 190	7	110	1 309	3	877	9	94	983
30 à 34 ans	1	657	9	169	836	3	158	5	83	249
35 à 39 ans	2	121	3	139	265	3	6	1	35	45
40 à 44 ans	0	6	2	48	56	2	0	0	19	21
45 à 49 ans	0	4	1	41	46	2	2	3	41	48
50 à 54 ans	1	5	1	41	48	2	7	1	38	48
55 à 59 ans	0	10	2	62	74	3	238	7	79	327
60 à 64 ans	2	1 431	42	108	1 583	5	1 751	65	130	1 951
65 à 74 ans	5	33	34	60	132	5	2	32	49	88
75+ ans	28	2 201	10	52	2 291	42	2 308	8	38	2 396
Célibataire	0	0	0	5	5	1	0	0	2	3
Marié(e)	0	0	0	10	10	1	1	0	2	4
Veuf(ve)	11	615	25	183	834	6	593	15	128	742
Divorcé(e)	0	12	14	106	132	3	15	11	94	123
Séparé(e)	33	5 933	12	30	6 008	20	5 510	3	34	5 567
Union libre = oui	9	0	0	318	327	16	0	0	278	294
Ménage de 1 personne	0	175	1 716	49	1 940	2	194	1 869	22	2 087
Ménage de 2 personnes	0	1	320	26	347	1	2	310	15	328
Ménage de 3 personnes	2	46	2 828	45	2 921	7	40	2 537	42	2 626
Ménage de 4 personnes	40	264	1 416	157	1 877	50	187	1 102	98	1 437
Ménage de 5 personnes	473	1 377	162	473	2 485	401	1 206	143	281	2 031
Ménage de 6+ personnes	2 228	4 133	118	24	6 503	1 941	3 960	121	9	6 031
Logement individuel non attenant**	256	3	3	42	304
Appartement < 5 étages**	550	2	0	141	693

** Seules 3 688 des 6 602 RP ont utilisé cette contrainte pour les calculs des poids.

.. indisponible pour une période de référence précise

Notes :

Petite : Contraintes petites

QLD : Contraintes quasi linéairement dépendantes

aberrants

LD : Contraintes linéairement dépendantes

Aber. : Contraintes qui ont donné lieu à des poids

Sources : Statistique Canada, recensements de 2006 et 2001.

Tableau 7.2.3.2 Fréquence à laquelle des contraintes au niveau des régions de pondération et des aires de diffusion ont été retranchées en 2001 et en 2006 lors de l'ajustement des poids finaux – Statistiques sommaires

	Recensement de 2006					Recensement de 2001				
	Petite	LD	QLD	Aber.	Total	Petite	LD	QLD	Aber.	Total
Contraintes au niveau des RP										
Total des contraintes retranchées	2 993	25 644	6 791	3 416	38 844	2 715	23 847	6 295	2 410	35 267
Contraintes retranchées par RP	0,5	3,9	1,0	0,5	5,9	0,4	3,9	1,0	0,4	5,7
Contraintes au niveau des AD										
Total des contraintes retranchées	248	190	4 154	869	5 461	1 354	357	4 191	917	6 819
Contraintes retranchées par AD	0,0	0,0	0,6	0,1	0,8	0,2	0,1	0,7	0,1	1,1

Notes :

Petite : Contraintes petites

LD : Contraintes linéairement dépendantes

QLD : Contraintes quasi linéairement dépendantes

Aber. : Contraintes qui ont donné lieu à des poids aberrants

Les contraintes relatives au type de logement (logements individuels non attenants, appartements de moins de cinq étages) n'ont pas été incluses dans ces chiffres, puisqu'elles n'étaient pas incluses dans toutes les RP.

Sources : Statistique Canada, recensements de 2006 et 2001.

8 Cohérence entre les estimations-échantillon et les chiffres de population

Le [chapitre 7](#) (voir le [Tableau 7.2.2.1](#)), faisait état des écarts entre les chiffres de population et les estimations-échantillon correspondantes fondés sur les poids finaux pour le Canada, calculés selon la méthode suivante :

$$\text{écart population/estimation} = \frac{\text{estimation-échantillon} - \text{chiffre de population}}{\text{chiffre de population}} \times 100$$

La comparaison entre les estimations-échantillon et les chiffres de population est fondée sur les logements privés occupés des UC échantillonnées.

Dans le présent chapitre, les écarts population/échantillon des recensements de 2001 et de 2006 seront étudiés en fonction des niveaux géographiques suivants :

- a) les aires de diffusion (AD);
- b) les régions de pondération (RP);
- c) les subdivisions de recensement (SDR);
- d) les secteurs de recensement (SR);
- e) les divisions de recensement (DR).

Au niveau des RP, nous observons que des écarts population/estimation nuls sont garantis pour les contraintes qui sont conservées par le système de pondération. En général, les régions géographiques constituées de RP complètes affichent de faibles écarts population/estimation. Le [tableau 7.1.2](#) révèle que 13,2 % des SDR et 67,0 % des SR sont constitués d'au moins une RP complète. En outre, étant donné la façon dont les RP sont formées, 100 % des DR sont constituées de RP complètes.

Les graphiques et tableaux de ce chapitre présentent les centiles des écarts population/estimation pour 33 caractéristiques qui, à quelques exceptions près, sont identiques aux 34 contraintes au niveau des RP appliquées aux poids du recensement (voir l'[annexe B](#)). Définissons le terme « centile » au moyen d'un exemple. Ainsi, le [tableau 8.2.1](#) affiche un 10^e centile de -11,12 % pour les ménages de six personnes ou plus en 2006. Ce chiffre signifie que 10 % des RP ont un écart de -11,12 % ou moins. Un 90^e centile de 6,75 % signifie que 10 % des RP ont un écart de 6,75 % ou plus. Les écarts population/estimation pour les régions géographiques dont le chiffre de population est inférieur ou égal à 50 pour une caractéristique donnée sont exclus des tableaux et graphiques présentés dans ce chapitre. Ces écarts se sont révélés relativement élevés et auraient pu altérer considérablement les centiles présentés dans ce chapitre.

Dans les prochaines sections, les écarts de 2006 seront comparés à ceux de 2001 pour divers niveaux de géographie.

8.1 Aires de diffusion

Le Canada est divisé en 54 626 AD, dont 52 448 renfermaient les ménages échantillonnés dans les 6 602 RP pendant le processus de pondération. La population moyenne d'une AD sera de 580 personnes.

Lorsqu'on compare les [graphiques 8.1.1](#) et [8.1.2](#) aux autres graphiques de ce chapitre, il ressort clairement que les écarts population/estimation sont un peu plus élevés au niveau des AD qu'au niveau des RP, des SDR, des SR ou des DR. Ce phénomène n'est pas surprenant, puisque les RP sont constituées d'AD complètes et que les RP correspondent au niveau le plus bas auquel les estimations-échantillon concorderont aux chiffres de population pour la plupart des caractéristiques.

Dans l'ensemble, la répartition des écarts au niveau des AD en 2006 était semblable à celle de 2001. Ainsi, les écarts observés en 2006 étaient tantôt légèrement plus prononcés, tantôt légèrement moins marqués qu'en 2001. Toutefois, les écarts sont légèrement plus prononcés pour la plupart des contraintes relatives aux groupes d'âge en 2006 comparativement à 2001.

8.2 Régions de pondération

Le Canada est divisé en 6 607 RP, dont 6 602 sont échantillonnées. Les RP ont une population moyenne de 4 785 personnes et sont habituellement constituées de 8 AD entières. Bien qu'on s'en serve pour calculer les poids du recensement, on ne publie pas de données au niveau des RP.

Le [tableau 8.2.1](#) montre que les 10^e, 25^e, 50^e, 75^e et 90^e centiles correspondent à zéro pour toutes pour les caractéristiques de la population, tant pour le Recensement de 2006 que pour celui de 2001. Dans le cas des caractéristiques des ménages, la plupart des 25^e, 50^e et 75^e centiles correspondent également à zéro, alors que certains des 10^e et 90^e centiles sont différents de zéro. Ces résultats ne sont pas surprenants puisque les RP constituent le plus bas niveau auquel les estimations-échantillon doivent correspondre aux chiffres de population pour les contraintes de pondération. La différence la plus marquée concerne les écarts plus prononcés pour les ménages composés de 5 personnes et de 6 personnes ou plus, ce à quoi on peut s'attendre compte tenu de l'écart au niveau du Canada pour ces contraintes. Les centiles des deux contraintes relatives au type de logement ne correspondent pas à zéro au niveau des RP, puisque leur utilisation était paramétrée et n'était pas incluse dans toutes les RP.

8.3 Subdivisions de recensement

Le Canada est divisé en 5 418 SDR. Une SDR correspond à une municipalité ou à une région considérée comme étant équivalente à des fins statistiques (p. ex., les réserves indiennes). La population moyenne des SDR se chiffre à 5 859 personnes, mais pour ce qui est de leur grandeur, elles peuvent ressembler à une très petite ville ou à une très grande ville. Le [tableau 7.1.2](#) montre que 13,2 % des SDR sont formées d'au moins une RP complète.

Les [graphiques 8.3.1](#) et [8.3.2](#) présentent les écarts population/estimation pour toutes les SDR échantillonnées au Canada. Dans l'ensemble, la répartition des écarts au niveau des SDR est

semblable en 2006 et en 2001. Cependant, les écarts sont légèrement plus élevés en 2006 qu'en 2001 pour les contraintes relatives aux groupes d'âge dans le cas des 10^e et 90^e centiles. La tendance ne s'applique pas aux 25^e et 75^e centiles des contraintes relatives aux groupes d'âge. On observe également certains écarts marqués en 2006 pour les ménages composés de 5 personnes et de 6 personnes ou plus. On ne s'en étonne pas, étant donné les écarts prononcés observés dans le [graphique 6.1](#) en fonction des poids initiaux.

8.4 Secteurs de recensement

Les SR se retrouvent uniquement dans les grands centres urbains dont le noyau urbain compte 50 000 habitants ou plus. Il y a 5 089 SR au Canada. Les SR comptent habituellement entre 1 500 et 8 000 personnes, la moyenne étant d'environ 4 500 personnes. Le [tableau 7.1.2](#) montre que 67,0 % des SR sont formés d'au moins une RP complète.

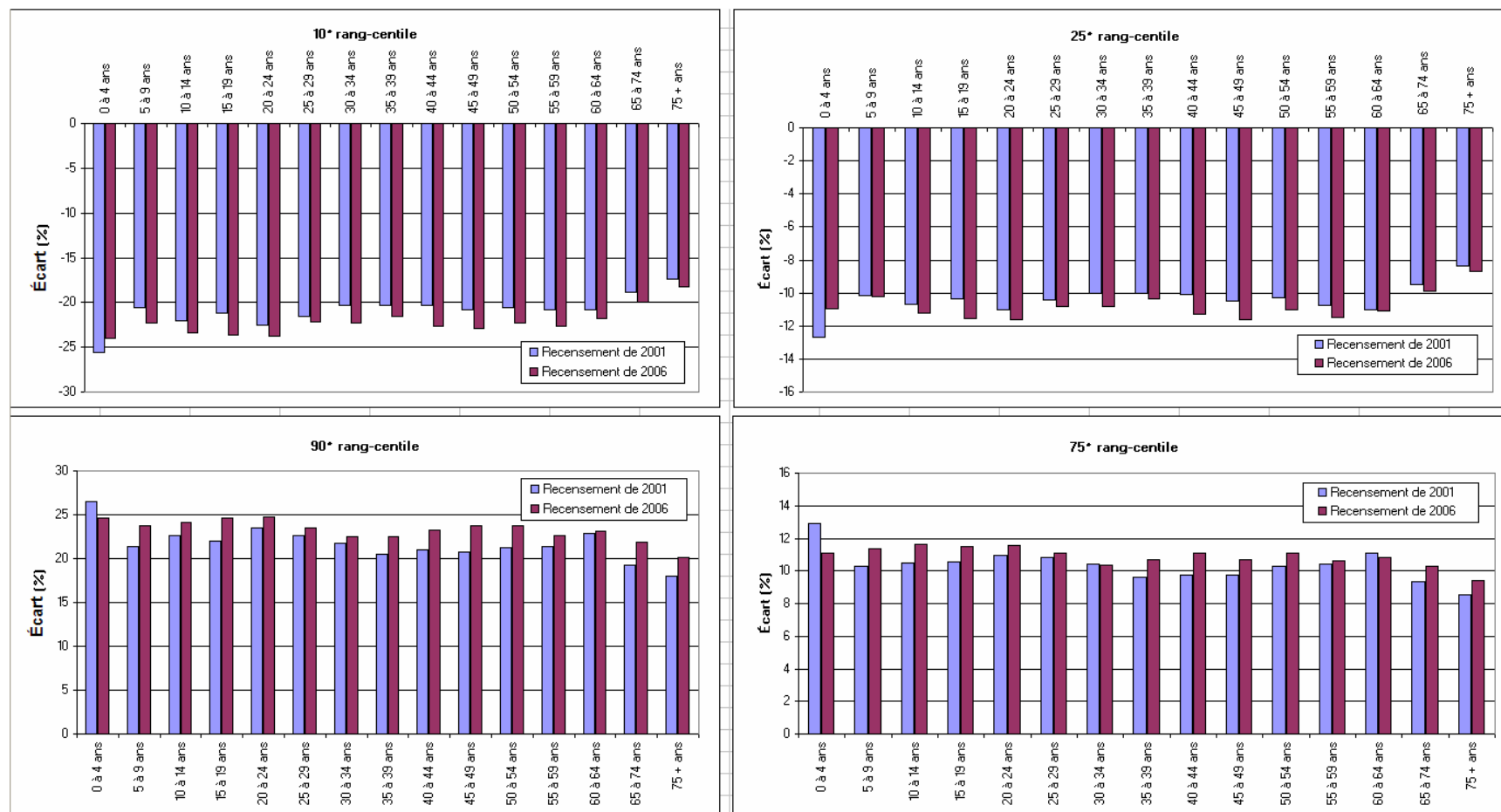
Le [graphique 8.4.1](#) présente les écarts population/estimation pour tous les SR échantillonnés au Canada. Il n'est pas étonnant que les écarts soient semblables en 2001 et en 2006 pour la plupart des caractéristiques. À l'instar des SDR, les ménages composés de 5 personnes et de 6 personnes ou plus sont associés à des écarts prononcés aux 10^e et 90^e centiles.

8.5 Divisions de recensement

Le Canada est divisé en 288 DR, qui comptent en moyenne 110 000 personnes. Une DR peut correspondre à un comté, à une municipalité régionale, à un district régional ou à toute autre région créée en vertu d'une loi provinciale ou territoriale.

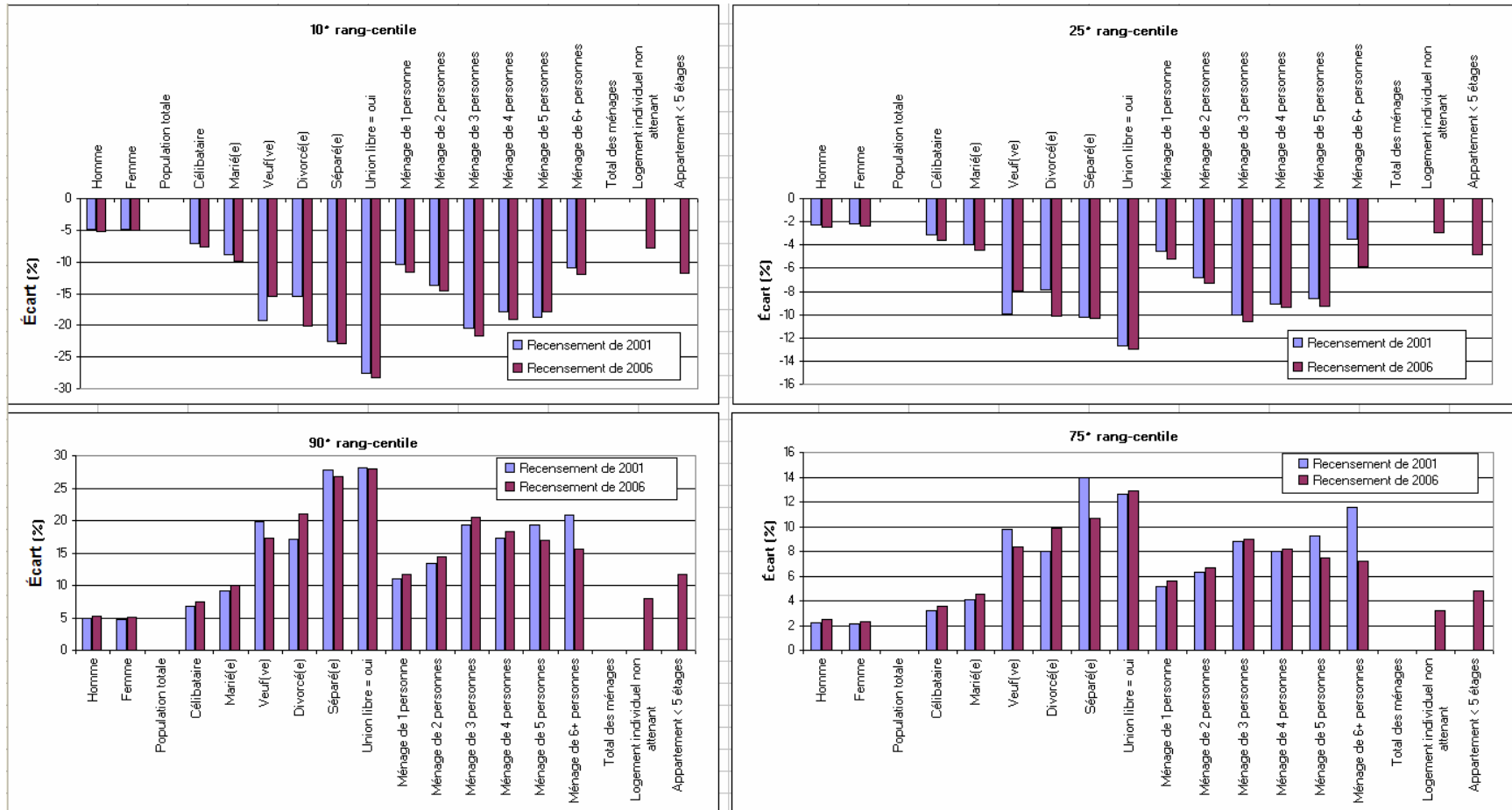
Le [tableau 8.5.1](#) présente un résumé des résultats des écarts population/estimation des recensements de 2006 et de 2001 pour les DR échantillonnées. Toutes les DR sont formées de RP complètes. Ainsi, les caractéristiques qui étaient rarement retranchées affichent une cohérence parfaite ou quasi parfaite au niveau des DR. À l'exception des caractéristiques des ménages composés de 5 personnes et de 6 personnes ou plus, les écarts pour les caractéristiques qui étaient éliminées plus souvent demeurent très faibles.

Graphique 8.1.1 Centiles des écarts population/estimation pour les aires de diffusion de 2006 et de 2001 (groupes d'âge)



Sources : Statistique Canada, recensements de 2006 et 2001.

Graphique 8.1.2 Centiles des écarts population/estimation pour les aires de diffusion de 2006 et de 2001 (autres caractéristiques)



Sources : Statistique Canada, recensements de 2006 et 2001.

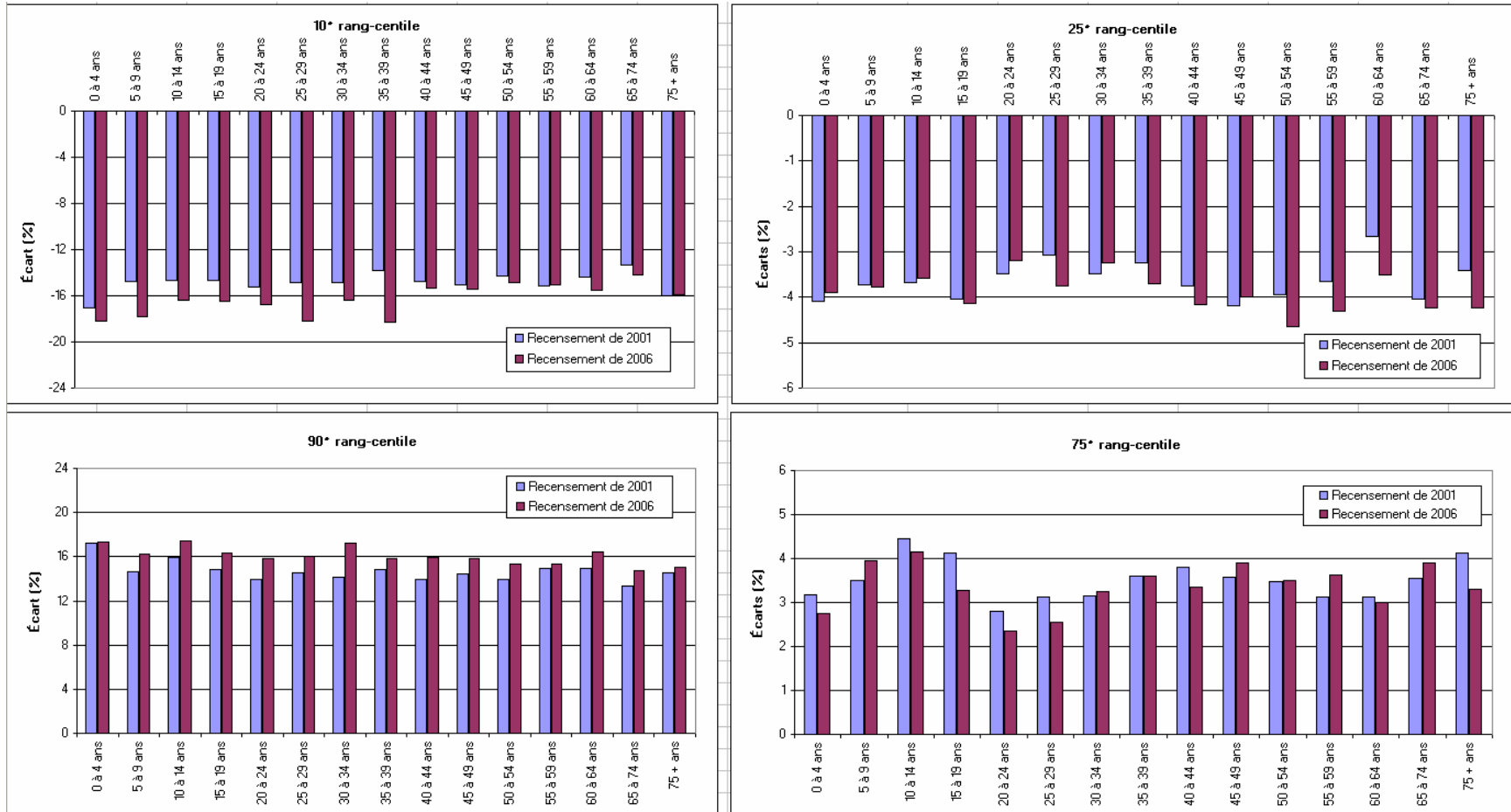
Tableau 8.2.1 Centiles des écarts population/estimation pour les régions de pondération

Caractéristique	Rang-centile de 2006					Rang-centile de 2001				
	10 ^e	25 ^e	50 ^e	75 ^e	90 ^e	10 ^e	25 ^e	50 ^e	75 ^e	90 ^e
Homme	0	0	0	0	0	0	0	0	0	0
Femme	0	0	0	0	0	0	0	0	0	0
Population totale	0	0	0	0	0	0	0	0	0	0
0 à 4 ans	0	0	0	0	0	0	0	0	0	0
5 à 9 ans	0	0	0	0	0	0	0	0	0	0
10 à 14 ans	0	0	0	0	0	0	0	0	0	0
15 à 19 ans	0	0	0	0	0	0	0	0	0	0
20 à 24 ans	0	0	0	0	0	0	0	0	0	0
25 à 29 ans	0	0	0	0	0	0	0	0	0	0
30 à 34 ans	0	0	0	0	0	0	0	0	0	0
35 à 39 ans	0	0	0	0	0	0	0	0	0	0
40 à 44 ans	0	0	0	0	0	0	0	0	0	0
45 à 49 ans	0	0	0	0	0	0	0	0	0	0
50 à 54 ans	0	0	0	0	0	0	0	0	0	0
55 à 59 ans	0	0	0	0	0	0	0	0	0	0
60 à 64 ans	0	0	0	0	0	0	0	0	0	0
65 à 74 ans	0	0	0	0	0	0	0	0	0	0
75+ ans	0	0	0	0	0	0	0	0	0	0
Célibataire	0	0	0	0	0	0	0	0	0	0
Marié(e)	0	0	0	0	0	0	0	0	0	0
Veuf(ve)	0	0	0	0	0	0	0	0	0	0
Divorcé(e)	0	0	0	0	0	0	0	0	0	0
Séparé(e)	0	0	0	0	0	0	0	0	0	0
Union libre = oui	0	0	0	0	0	0	0	0	0	0
Ménage de 1 personne	-0,79	0	0	0	0,12	-1,11	0	0	0	0,03
Ménage de 2 personnes	0	0	0	0	0	0	0	0	0	0
Ménage de 3 personnes	-2,04	-0,14	0	0	0,65	-1,8	-0,18	0	0	0,40
Ménage de 4 personnes	-0,28	0	0	0	1,24	-0,06	0	0	0	0,16
Ménage de 5 personnes	0	0	0	0	15,35	0	0	0	0	7,89
Ménage de 6+ personnes	-11,12	-4,36	0	3,06	6,75	-6,07	-1,57	1,16	4,63	7,98
Total des ménages	0	0	0	0	0	0	0	0	0	0
Logement individuel non attendant	-2,02	-0,75	0,02	0,96	2,71
Appartement < 5 étages	-6,84	-2,35	0,04	2,56	7,09

.. indisponible pour une période de référence précise

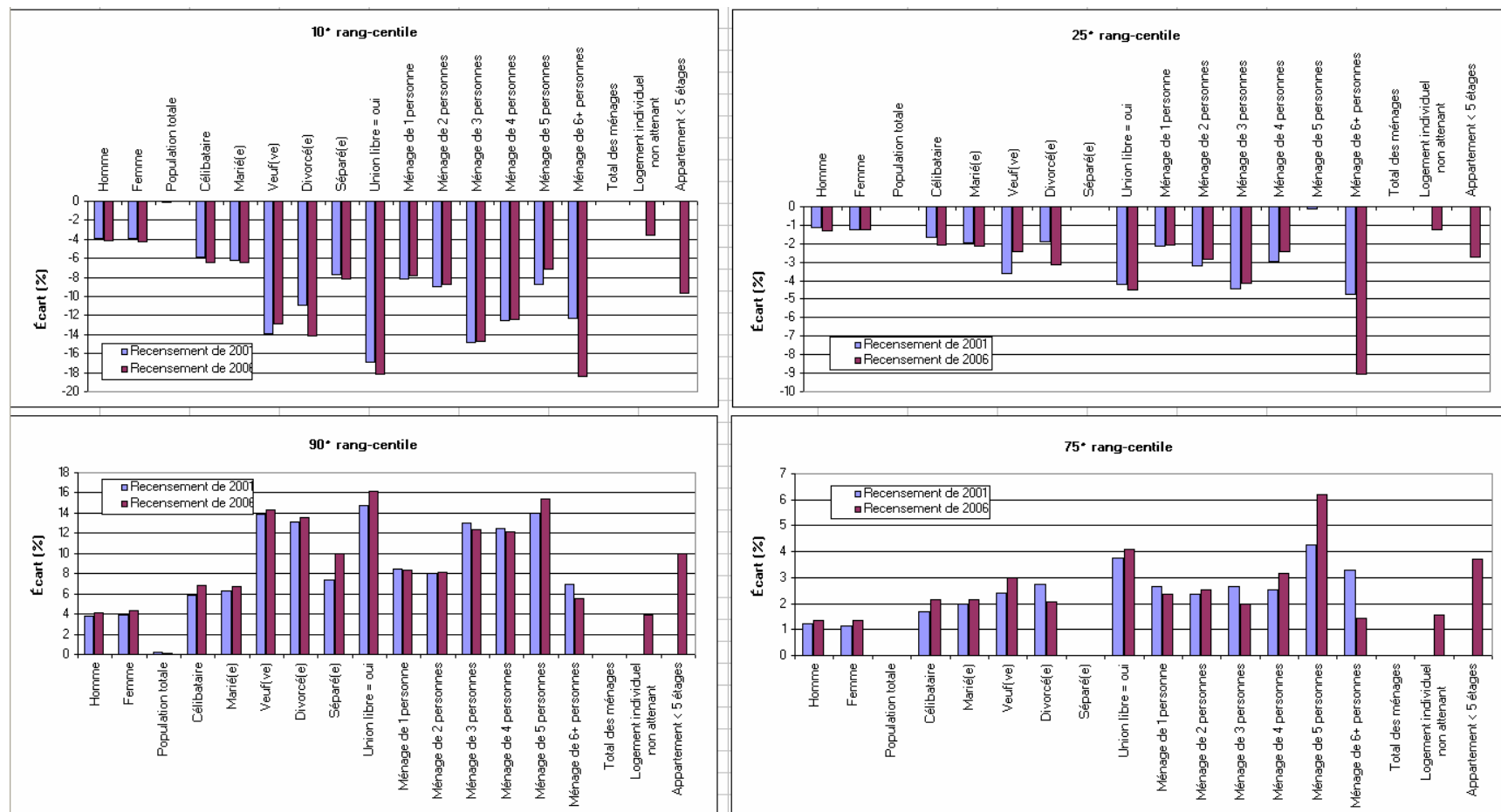
Sources : Statistique Canada, recensements de 2006 et 2001.

Graphique 8.3.1 Centiles des écarts population/estimation pour les subdivisions de recensement (groupes d'âge)



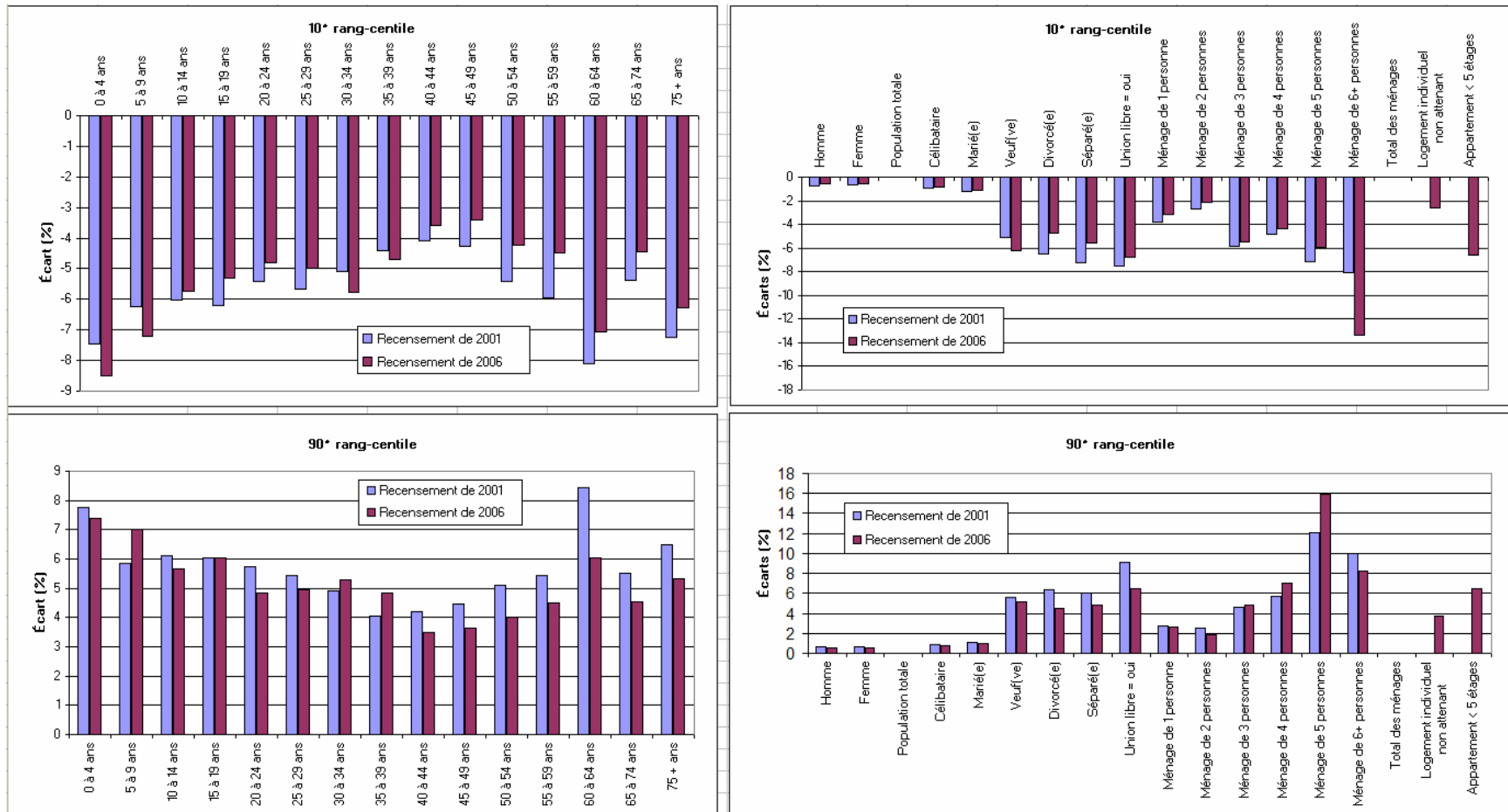
Sources : Statistique Canada, recensements de 2006 et 2001.

Graphique 8.3.2 Centiles des écarts population/estimation pour les subdivisions de recensement (autre caractéristiques)



Sources : Statistique Canada, recensements de 2006 et 2001.

Graphique 8.4.1 Centiles des écarts population/estimation pour les secteurs de recensement



Sources : Statistique Canada, recensements de 2006 et 2001.

Tableau 8.5.1 Centiles des écarts population/estimation pour les divisions de recensement

Caractéristique	Rang-centile de 2006					Rang-centile de 2001				
	10 ^e	25 ^e	50 ^e	75 ^e	90 ^e	10 ^e	25 ^e	50 ^e	75 ^e	90 ^e
Homme	0	0	0	0	0	0	0	0	0	0
Femme	0	0	0	0	0	0	0	0	0	0
Population totale	0	0	0	0	0	0	0	0	0	0
0 à 4 ans	-2,07	-0,17	0	0,23	2,28	-0,81	0	0	0	0,72
5 à 9 ans	-1,67	-0,20	0	0	1,06	-0,51	0	0	0	0,23
10 à 14 ans	-0,51	0	0	0	0,99	-0,20	0	0	0	0,36
15 à 19 ans	-0,03	0	0	0	0,76	-0,05	0	0	0	0,06
20 à 24 ans	-1,57	0	0	0	0,84	-0,41	0	0	0	0,29
25 à 29 ans	-2,65	-0,46	0	0	0,81	-0,54	0	0	0	0,44
30 à 34 ans	-1,28	0	0	0	0,75	-0,04	0	0	0	0,08
35 à 39 ans	-0,18	0	0	0	0,48	0	0	0	0	0
40 à 44 ans	0	0	0	0	0	0	0	0	0	0
45 à 49 ans	0	0	0	0	0	0	0	0	0	0
50 à 54 ans	0	0	0	0	0	0	0	0	0	0
55 à 59 ans	0	0	0	0	0	0	0	0	0	0
60 à 64 ans	-0,25	0	0	0	0,50	-0,18	0	0	0	0,49
65 à 74 ans	0	0	0	0	0	0	0	0	0	0
75+ ans	-0,48	0	0	0	0,21	-0,65	0	0	0	0,28
Célibataire	0	0	0	0	0	0	0	0	0	0
Marié(e)	0	0	0	0	0	0	0	0	0	0
Veuf(ve)	-0,41	0	0	0	0,48	-0,14	0	0	0	0,04
Divorcé(e)	0	0	0	0	0,05	-0,08	0	0	0	0,31
Séparé(e)	-1,72	0	0	0	2,21	-0,96	0	0	0	0,94
Union libre = oui	-0,74	0	0	0	1,10	-0,21	0	0	0	0,84
Ménage de 1 personne	-0,40	-0,13	-0,01	0	0,03	-0,34	-0,15	-0,04	0	0,01
Ménage de 2 personnes	-0,05	0	0	0	0	-0,04	0	0	0	0
Ménage de 3 personnes	-1,24	-0,66	-0,21	0	0,39	-1,05	-0,59	-0,23	0	0,29
Ménage de 4 personnes	-0,53	-0,10	0,03	0,55	1,29	-0,27	-0,07	0	0,16	0,67
Ménage de 5 personnes	-0,63	0,17	2,42	5,16	8,55	-0,79	0	0,99	2,47	5,16
Ménage de 6+ personnes	-18,92	-10,78	-4,98	-2,22	1,53	-9,17	-3,86	-0,76	1,65	3,87
Total des ménages	0	0	0	0	0	0	0	0	0	0
Logement individuel non attendant	-0,62	-0,22	0,06	0,38	0,71
Appartement < 5 étages	-3,35	-1,30	0,01	1,42	3,92

.. indisponible pour une période de référence précise

Sources : Statistique Canada, recensements de 2006 et 2001.

9 Variance d'échantillonnage

L'erreur d'échantillonnage a deux composantes : la variance et le biais. La variance mesure la variabilité de l'estimation par rapport à l'estimation moyenne qui résulterait de répétitions hypothétiques du processus d'enquête. Le biais est la différence entre la valeur moyenne d'une estimation qui résulterait de répétitions hypothétiques du processus d'enquête, et la valeur réelle de la variable estimée. Le [chapitre 6](#) présente les résultats du biais d'échantillonnage et décrit la nature et l'importance du biais dans l'échantillon avant la pondération. Même avec une méthode d'échantillonnage totalement dépourvue de biais, les résultats seraient sujets à la variance du simple fait que les estimations ont été produites à partir d'un échantillon. On peut estimer la variance à partir de données-échantillon⁴. La variance d'échantillonnage a été étudiée afin d'estimer l'effet des méthodes d'échantillonnage et d'estimation sur les chiffres du recensement établis à partir de données-échantillon.

Statistique Canada produit des milliers de tableaux à partir des données-échantillon (c'est-à-dire celles recueillies par la formule 2B). Théoriquement, l'estimation de la variance d'échantillonnage constitue une mesure de la précision de chaque estimation présentée dans ces tableaux. Cette mesure tiendrait compte à la fois du plan de sondage et de la méthode d'estimation. En pratique cependant, il est impossible de calculer une telle mesure pour toutes les estimations du recensement en raison du coût élevé du traitement des données. On calcule donc la variance d'échantillonnage pour un sous-ensemble seulement des estimations du recensement. Cela permet ensuite d'évaluer les effets combinés du plan de sondage et de la méthode d'estimation sur la variance d'échantillonnage. On peut ensuite ajuster les estimations simples de la variance d'échantillonnage, qui sont peu coûteuses à produire, pour tenir compte de cet effet et produire une estimation de la variance d'échantillonnage pour n'importe quelle estimation du recensement.

La racine carrée de la variance d'échantillonnage, appelée erreur type, peut être calculée approximativement à l'aide des données des [tableaux 9.1](#) et [9.2](#). Le [tableau 9.1](#) donne les valeurs non ajustées (simples) de l'erreur type des estimations-échantillon du recensement. Les chiffres présentés dans ce tableau ont été établis en supposant qu'on a utilisé un échantillon aléatoire simple au 1/5 et une pondération simple de 5. Les erreurs types sont présentées au [tableau 9.1](#) en fonction à la fois de la taille de l'estimation du recensement et de la taille de la région géographique. Par exemple, pour un total estimé de 250 personnes dans une région géographique comptant 1 000 personnes, l'erreur type non ajustée serait de 25.

4. Malheureusement, la variance d'échantillonnage ne nous fournit aucune indication de l'importance de l'erreur non due à l'échantillonnage.

Le [tableau 9.1](#) contient les erreurs types associées à un nombre restreint de valeurs du « total estimé » et du nombre total de personnes, de ménages, de logements ou de familles dans la région. La formule qui suit peut servir à calculer l'erreur type non ajustée de n'importe quel total estimé, pour une région de n'importe quelle taille :

$$ETNA =$$

où ETNA est l'erreur type non ajustée; E; le total estimé, et N, le nombre total de personnes, de ménages, de logements ou de familles dans la région. Par exemple, pour un total estimé de 750 personnes dans une région comptant 9 000 personnes, l'erreur type non ajustée serait de :

$$= 52$$

Le [tableau 9.2](#) fournit des facteurs d'ajustement⁵ par lesquels on doit multiplier les erreurs types non ajustées pour tenir compte des effets combinés du plan de sondage et de la méthode d'estimation. Pour calculer ces facteurs d'ajustement, on a calculé les estimations de la variance d'échantillonnage des estimations obtenues par régression pour différentes catégories de toutes les caractéristiques⁶ énumérées au [tableau 9.2](#). Ce calcul a été fait pour chaque RP échantillonnée. Les estimations de la variance d'échantillonnage aux niveaux provincial et national ont été obtenues en faisant la somme des estimations au niveau des RP. Les facteurs d'ajustement pour chaque caractéristique dans chacune des catégories ont été calculés en divisant la racine carrée de ces estimations par l'erreur type non ajustée correspondante. Des facteurs d'ajustement ont été calculés aux niveaux provincial et national pour chaque caractéristique en faisant la moyenne des facteurs d'ajustement de toutes ses catégories. Par exemple, les facteurs d'ajustement pour le « Sexe » correspondent à la moyenne de ceux qui s'appliquent aux catégories Hommes et Femmes. Pour la majorité des caractéristiques, les catégories sont regroupées en fonction de facteurs d'ajustement similaires, et le facteur du groupe approprié devrait être utilisé pour chaque catégorie. Lorsque le tableau renvoie à plusieurs catégories, le plus gros facteur d'ajustement doit être utilisé. Pour de plus amples renseignements sur la façon dont ces facteurs d'ajustement ont été calculés, voir Hovington (2004).

Pour estimer l'erreur type d'une estimation-échantillon donnée du recensement, l'utilisateur doit trouver dans le [tableau 9.2](#) le facteur d'ajustement s'appliquant à la caractéristique en question et multiplier ce facteur par l'erreur type non ajustée choisie à partir du [tableau 9.1](#). Si la caractéristique n'est pas énumérée au [tableau 9.2](#), l'utilisateur doit choisir le facteur d'ajustement 1 indiqué pour la catégorie « Toutes les autres ». Pour chaque caractéristique du [tableau 9.2](#), les facteurs d'ajustement sont fournis aux niveaux national et provincial, ainsi qu'au niveau des RP. Sauf dans le cas d'une région infraprovinciale, il faut choisir le « Facteur national ou provincial ». Au [tableau 9.2](#), les facteurs d'ajustement associés aux différentes provinces sont indiqués seulement s'ils diffèrent significativement de ceux du niveau national. Ce n'est le cas que pour quelques caractéristiques relatives à la langue. Il convient de prendre note que, étant donné qu'aucun échantillonnage n'a été fait au Nunavut, les facteurs d'ajustement de toutes les caractéristiques dans ce territoire doivent

5. On appelle communément « effet de plan » le carré des facteurs d'ajustement.

6. Par exemple, « 10 000 \$ à 19 999 \$ » est l'une des catégories pour lesquelles des estimations de la variance d'échantillonnage ont été calculées pour la caractéristique « Nombre de personnes dans les intervalles de revenu total ».

correspondre à zéro. Comme on a eu recours à l'échantillonnage au Territoire du Yukon et aux Territoires du Nord-Ouest, il faut utiliser le facteur d'ajustement « autres provinces » s'il est offert. Si un facteur d'ajustement est requis pour une estimation du recensement associée à une région infraprovinciale, les centiles des facteurs au niveau des RP constitueront des valeurs plus précises. Les centiles donnent l'étendue de tous les facteurs d'ajustement calculés dans l'étude au niveau des RP pour les différentes catégories d'une caractéristique. N % des facteurs d'ajustement au niveau des RP sont inférieurs au N^e centile et (100 – N) % d'entre eux sont supérieurs au N^e centile. Par exemple, 90 % des facteurs d'ajustement au niveau des RP sont inférieurs au 90^e centile et 10 % d'entre eux sont supérieurs à celui-ci. Selon qu'on souhaite obtenir une estimation de l'erreur type plus ou moins prudente, on choisira le centile en conséquence. Par exemple, le 99^e centile fournirait une estimation très prudente, et le 75^e centile une estimation un peu moins prudente.

Les règles qui suivent s'appliquent lors du calcul de l'erreur type ajustée :

- a) Pour déterminer l'erreur type d'une estimation ayant trait aux familles ou aux ménages, c'est le nombre de familles ou de ménages dans la région, non pas le nombre de personnes, qui permet de repérer la colonne appropriée au [tableau 9.1](#).
- b) À moins d'avis contraire, les facteurs d'ajustement des caractéristiques des familles qui comprennent un époux, une épouse, un parent seul ou une personne repère sont les mêmes que pour les caractéristiques de la population. À titre d'exemple, le facteur d'ajustement de la caractéristique « Plus haut niveau de scolarité atteint chez l'époux, l'épouse, ou le parent seul d'une famille de recensement » est le même que pour la caractéristique de la population « Plus haut niveau de scolarité atteint ».
- c) Pour le classement recoupé de deux caractéristiques ou plus, il faut utiliser le facteur d'ajustement le plus élevé pour ces caractéristiques.
- d) Tous les facteurs d'ajustement de l'erreur type visent les estimations relatives au nombre de personnes, de ménages, de logements ou de familles et non pas, par exemple, des valeurs monétaires. Par exemple, les facteurs d'ajustement du revenu du ménage visent les estimations du nombre de ménages dont le revenu fait partie d'une fourchette donnée, et non aux estimations telles que le revenu moyen des ménages.

L'exemple qui suit illustre la façon de calculer l'erreur type ajustée. Supposons que l'estimation étudiée soit le nombre de personnes qui ont immigré au Canada entre 1996 et 2006. L'estimation du Recensement de 2006 pour cette caractéristique était 1 954 605. Le chiffre de population obtenu pour le Canada au Recensement de 2006 pour les variables échantillonnées était 31 241 030. Étant donné que ni l'un ni l'autre de ces chiffres ne se rapproche des valeurs fournies au [tableau 9.1](#), on doit employer la formule qui sert à calculer l'erreur type non ajustée. Ainsi, on obtient une estimation de 2 707. D'après le [tableau 9.2](#), le facteur d'ajustement au niveau national pour la caractéristique « période d'immigration » après 1990 est 1,67. Il s'ensuit que l'erreur type ajustée pour cette estimation est $2\,707 \times 1,67 = 4\,520$.

L'estimation-échantillon et son erreur type peuvent être utilisées pour déterminer un intervalle à l'intérieur duquel on peut s'attendre, avec un degré de certitude établi, à trouver le chiffre de population inconnu. L'échantillon particulier choisi pour cette enquête est l'un des nombreux échantillons de même taille qui auraient pu être sélectionnés selon le même plan de sondage. Les estimations établies pour ces différents échantillons seraient différentes de l'un à l'autre. Si des intervalles allant de deux erreurs types sous l'estimation à deux erreurs types au-dessus de l'estimation étaient établis à partir de chacune des différentes estimations possibles, environ 19 intervalles sur

20 incluraient la valeur normalement obtenue lors d'un recensement complet. Un tel intervalle est appelé intervalle de confiance de 95 % ($19 \div 20 = 95\%$). Afin de garantir un seuil de confiance de 95 %, ces intervalles doivent toutefois être calculés à l'aide des vraies erreurs types des estimations-échantillon. Les erreurs types ajustées qui sont calculées à partir des [tableaux 9.1](#) et [9.2](#) ne sont que des estimations des véritables erreurs types. Toutefois, pour les estimations-échantillon aux niveaux provincial et national, les erreurs types ajustées devraient être assez près des véritables erreurs types pour calculer des intervalles de confiance de 95 % d'une assez bonne précision. Au niveau infraprovincial, les erreurs types ajustées peuvent ne pas être suffisamment précises pour servir à cette fin.

Si on utilise l'erreur type calculée ci-dessus, l'intervalle de confiance à 95 % pour le nombre de personnes ayant immigré au Canada entre 1996 et 2006 serait $1\,954\,605 \pm 2(4\,520)$ ou $1\,954\,605 \pm 9\,040$.

Il convient de souligner que les estimations des petites régions sont parfois peu fiables, comme le démontre l'exemple suivant. Une communauté de 500 personnes où l'on avait estimé à 50 le nombre de personnes qui avaient immigré au Canada entre 1996 et 2006 aurait une erreur type de 15 selon le [tableau 9.1](#). Étant donné que cette population est inférieure à celle du niveau provincial, un facteur d'ajustement au niveau de la RP doit être sélectionné à partir du [tableau 9.2](#). En s'appuyant sur le chiffre le plus prudent du 99^e centile, on obtiendrait une erreur type corrigée de $15 \times 2,46 = 36,9$, ce qui donnerait un intervalle de confiance d'approximativement 95 % de $50 \pm 2(36,9)$ ou $50 \pm 73,8$. Autrement dit, la valeur réelle de la population de cette communauté des personnes qui ont immigré au Canada entre 1996 et 2006 pourrait se situer n'importe où de 0 à 123, selon un intervalle de confiance de 95 %. Même lorsqu'on utilise un chiffre un peu moins prudent selon le facteur d'ajustement au 75^e centile (1,52), on obtient un intervalle de confiance à 95 % qui varie de 5 à 95.

Tableau 9.1 Estimations non ajustées des erreurs types des estimations-échantillon

Nombre estimatif de personnes, ménages ou logements dans la région pour la caractéristique à l'étude	Nombre estimatif de personnes, de ménages ou de logements dans la région							
	500	1 000	2 500	5 000	10 000	25 000	50 000	100 000
50	15	15	15	15	15	15	15	15
100	20	20	20	20	20	20	20	20
250	20	25	30	30	30	30	30	30
500	0	30	40	40	45	45	45	45
1 000	...	0	50	55	60	60	65	65
2 500	0	70	85	95	100	100
5 000	0	100	125	135	140
10 000	0	155	180	190
25 000	0	225	275
50 000	0	315
100 000	0
	250 000	500 000	1 000 000	2 500 000	5 000 000	10 000 000	25 000 000	30 000 000
50	15	15	15	15	15	15	15	15
100	20	20	20	20	20	20	20	20
250	30	30	30	30	30	30	30	30
500	45	45	45	45	45	45	45	45
1 000	65	65	65	65	65	65	65	65
2 500	100	100	100	100	100	100	100	100
5 000	140	140	140	140	140	140	140	140
10 000	195	200	200	200	200	200	200	200
25 000	300	310	310	315	315	315	315	315
50 000	400	425	435	445	445	445	445	445
100 000	490	565	600	620	625	630	630	630
250 000	0	705	865	950	975	985	995	995
500 000	...	0	1 000	1 265	1 340	1 380	1 400	1 400
1 000 000	0	1 550	1 790	1 900	1 960	1 965
2 500 000	0	2 235	2 740	3 000	3 030
5 000 000	0	3 160	4 000	4 085
10 000 000	0	4 900	5 165
15 000 000	4 900	5 475

... n'ayant pas lieu de figurer

Source : Statistique Canada, Recensement de 2006.

Tableau 9.2 Facteurs d'ajustement de l'erreur type au niveau national ou provincial et des régions de pondération

Caractéristique	Facteurs nationaux ou provinciaux	Rangs centiles des facteurs au niveau des RP						
		1	25	50	75	90	95	99
Caractéristique de la population								
Âge								
0 à 4, 5 à 9, 10 à 14, 15 à 19, 20 à 29, 30 à 34 ans	0,17	0,00	0,00	0,00	0,13	0,32	0,60	0,78
35 à 44, 45 à 54, 55 à 59, 60 à 64, 65+ ans	0,12	0,00	0,00	0,00	0,11	0,18	0,27	0,71
Sexe								
	0,07	0,00	0,00	0,04	0,07	0,10	0,13	0,18
Union libre								
En union libre	0,19	0,00	0,00	0,00	0,09	0,23	0,65	1,03
Pas en union libre	0,16	0,00	0,00	0,00	0,04	0,19	0,39	0,87
État matrimonial								
Célibataire, marié(e)	0,07	0,00	0,00	0,00	0,05	0,10	0,13	0,20
Séparé(e), divorcé(e), veuf(ve)	0,13	0,00	0,00	0,00	0,11	0,22	0,32	0,68
Plus haut grade, certificat ou diplôme								
	1,12	0,62	1,03	1,15	1,25	1,36	1,43	1,63
Principal domaine d'études								
	1,10	0,29	0,86	0,98	1,14	1,26	1,34	1,55
Lieu de naissance								
Né(e) au Canada	1,30	0,24	0,98	1,27	1,48	1,64	1,74	1,97
Né(e) à l'extérieur du Canada	0,98	0,56	0,97	1,07	1,19	1,31	1,38	1,55
Citoyenneté								
Canadienne, de naissance	1,21	0,19	0,90	1,17	1,41	1,59	1,74	2,00
Autre	1,16	0,06	0,78	1,10	1,37	1,61	1,73	2,00
Nombre de citoyennetés								
Canadienne seulement	1,26	0,26	1,10	1,30	1,48	1,64	1,76	2,02
Une ou deux autres	1,67	0,00	1,17	1,38	1,64	1,92	2,09	2,52
Période d'immigration								
Avant 1950, 1951 à 1960, 1961 à 1970, 1971 à 1980, 1981 à 1990	1,37	0,44	0,91	1,06	1,26	1,43	1,55	1,79
1991 à 1995, 1996 à 2001, 2001 à 2006	1,67	0,00	0,91	1,18	1,52	1,87	2,09	2,46
Âge à l'immigration								
	1,20	0,61	0,95	1,15	1,32	1,46	1,54	1,74
Mobilité (1 an auparavant)								
Personne n'ayant pas déménagé	1,70	0,51	1,32	1,57	1,80	1,97	2,08	2,36
Personne ayant déménagé (migrant, non-migrant)	1,86	0,32	1,46	1,71	1,94	2,16	2,29	2,67

Tableau 9.2 Facteurs d'ajustement de l'erreur type au niveau national ou provincial et des régions de pondération (suite)

Caractéristique	Facteurs nationaux ou provinciaux	Rangs centiles des facteurs au niveau des RP						
		1	25	50	75	90	95	99
Mobilité (5 ans auparavant)								
Personne n'ayant pas déménagé	1,58	0,59	1,32	1,51	1,70	1,86	1,96	2,10
Personne ayant déménagé (migrant, non-migrant)	1,75	0,63	1,49	1,70	1,90	2,12	2,28	2,73
Population des immigrants ou des non-immigrants								
Population des immigrants	1,30	0,79	1,29	1,44	1,60	1,76	1,87	2,11
Population des non-immigrants	1,21	0,19	0,89	1,15	1,39	1,58	1,71	1,97
Minorité visible								
Chinois, Sud-Asiatique, Noir, Philippin, Latino-Américain, Arabe, Coréen, Japonais, Minorité visible, n.i.a., Minorité visible multiple	2,04	0,00	1,16	1,45	1,88	2,32	2,59	3,16
Autochtone	1,41	0,54	1,36	1,61	1,85	2,07	2,21	2,52
Autre	1,53	0,22	1,03	1,41	1,75	2,04	2,20	2,48
Origine ethnique								
	1,56	0,76	1,43	1,63	1,79	1,95	2,04	2,22
Langue parlée à la maison – anglais								
Nouveau-Brunswick, Colombie-Britannique, Ontario, Alberta	1,59	0,17	1,31	1,68	1,93	2,11	2,24	2,48
Québec	1,61	0,79	1,57	1,80	2,01	2,18	2,33	2,62
Autre province	1,24	0,00	0,82	1,37	1,73	2,01	2,15	2,49
Canada	1,16
Langue parlée à la maison – français								
Nouvelle-Écosse, Québec	1,28	0,00	0,99	1,48	1,78	1,99	2,13	2,82
Nouveau-Brunswick	0,99	0,49	1,38	1,65	1,90	2,21	2,32	2,63
Autre province	1,64	0,00	1,37	1,67	1,96	2,26	2,45	2,88
Canada	0,77
Première langue officielle parlée – anglais								
Québec	1,43	0,77	1,32	1,55	1,75	1,92	2,02	2,25
Terre-Neuve	0,68	0,00	0,29	0,61	1,02	1,39	1,65	2,28
Autre province	1,21	0,18	0,92	1,22	1,45	1,64	1,76	2,01
Canada	0,79
Première langue officielle parlée – français								
Nouveau-Brunswick	0,95	0,57	1,23	1,44	1,60	1,79	1,97	2,32
Autre province	1,28	0,23	1,21	1,42	1,65	1,88	2,01	2,31
Canada	0,76
Première langue officielle parlée – autre								
Anglais ou français	1,64	0,00	1,18	1,43	1,70	1,99	2,24	2,62

Tableau 9.2 Facteurs d'ajustement de l'erreur type au niveau national ou provincial et des régions de pondération (suite)

Caractéristique	Facteurs nationaux ou provinciaux	Rangs centiles des facteurs au niveau des RP						
		1	25	50	75	90	95	99
Ni l'une ni l'autre	1,03	0,00	0,74	0,90	1,09	1,27	1,39	1,66
Langue officielle parlée – anglais								
Québec	1,36	0,00	1,24	1,41	1,58	1,74	1,87	2,25
Autre province	1,28	0,37	1,11	1,33	1,50	1,64	1,74	1,90
Canada	0,84
Langue officielle parlée – français								
Nouveau-Brunswick, Québec	1,25	0,56	1,23	1,42	1,56	1,68	1,76	1,94
Autre province	1,62	0,00	1,16	1,38	1,68	2,05	2,29	2,91
Canada	0,94
Langue officielle parlée – autre								
Bilingue	1,22	0,72	1,23	1,36	1,48	1,60	1,70	1,88
Ni l'une ni l'autre	1,44	0,00	1,04	1,26	1,52	1,77	1,94	2,34
Langue maternelle – anglais								
Ontario, Alberta, Colombie-Britannique	1,34	0,00	1,11	1,39	1,62	1,82	1,93	2,14
Québec	1,39	0,57	1,24	1,45	1,66	1,85	1,96	2,21
Autre province	1,10	0,00	0,92	1,19	1,44	1,66	1,79	2,02
Canada	1,07
Langue maternelle – français								
Nouveau-Brunswick	0,85	0,00	1,04	1,31	1,52	1,75	1,87	2,31
Québec	1,14	0,05	1,00	1,34	1,58	1,77	1,90	2,16
Autre province	1,35	0,54	1,21	1,39	1,60	1,83	1,98	2,30
Canada	0,75
Langue maternelle – autre que l'anglais ou le français								
	1,76	0,00	1,02	1,24	1,56	1,95	2,24	2,87
Langue de travail – anglais								
Québec	1,15	0,74	1,13	1,24	1,34	1,45	1,52	1,68
Autre province	0,78	0,24	0,67	0,77	0,88	1,01	1,10	1,26
Canada	0,75
Langue de travail – français								
Nouveau-Brunswick, Québec	0,88	0,38	0,75	0,87	1,06	1,25	1,33	1,48
Autre province	1,22	0,00	1,07	1,20	1,34	1,52	1,67	1,97
Canada	0,72
Langue de travail – autre								
	1,22	0,00	0,75	1,01	1,25	1,54	1,81	2,26
Industrie	1,23	0,66	0,94	1,11	1,24	1,36	1,45	1,66

Tableau 9.2 Facteurs d'ajustement de l'erreur type au niveau national ou provincial et des régions de pondération (suite)

Caractéristique	Facteurs nationaux ou provinciaux	Rangs centiles des facteurs au niveau des RP						
		1	25	50	75	90	95	99
Profession	1,17	0,69	0,87	0,97	1,12	1,24	1,31	1,47
Travail en 2005	1,08	0,68	1,06	1,17	1,26	1,36	1,42	1,57
Semaines travaillées en 2005	1,02	0,57	0,92	1,07	1,20	1,29	1,35	1,48
Heures travaillées pendant la semaine de référence	1,09	0,51	0,86	0,98	1,12	1,22	1,29	1,44
Travail à plein temps ou à temps partiel								
Travail à plein temps	0,82	0,35	0,71	0,81	0,89	0,98	1,03	1,13
Travail à temps partiel	1,09	0,58	0,98	1,07	1,15	1,22	1,27	1,39
Année du dernier travail								
En 2006, en 2005, avant 2005	0,86	0,39	0,82	0,97	1,12	1,24	1,30	1,44
N'a jamais travaillé	1,20	0,66	1,09	1,19	1,28	1,38	1,46	1,61
Catégorie de travailleur								
Travailleur rémunéré	0,89	0,43	0,85	1,02	1,25	1,41	1,50	1,71
Travailleur autonome, entreprise non constituée en société, travailleur familial non rémunéré	1,26	0,66	1,11	1,23	1,35	1,47	1,57	1,85
Travaux ménagers sans rémunération	1,20	0,65	1,10	1,20	1,30	1,39	1,45	1,62
Situation professionnelle								
Personne occupée	0,91	0,00	0,83	0,98	1,16	1,29	1,37	1,54
Personne au chômage	1,22	0,00	1,02	1,15	1,28	1,42	1,53	1,85
Personne inactive	1,07	0,64	1,02	1,16	1,28	1,41	1,49	1,68
Mode de transport au travail								
Avec sa propre voiture, à pied, transports publics, covoiturage, taxi	0,77	0,39	0,74	0,85	0,93	1,03	1,09	1,25
Bicyclette, motocyclette	1,21	0,00	1,01	1,13	1,24	1,36	1,46	1,77
Autre	1,21	0,60	1,03	1,15	1,28	1,42	1,54	1,95
Lieu de travail – province	0,56	0,00	0,70	0,85	1,02	1,22	1,37	1,74
Lieu de travail – Classification des secteurs statistiques (zone d'influence métropolitaine [ZIM])								
ZIM forte ou modérée	0,95	0,48	0,99	1,09	1,19	1,31	1,40	1,67
ZIM faible ou hors ZIM	0,79	0,00	0,99	1,16	1,34	1,54	1,68	2,05
Dans une AR ou une RMR	0,88	0,49	0,99	1,14	1,28	1,41	1,51	1,81

Tableau 9.2 Facteurs d'ajustement de l'erreur type au niveau national ou provincial et des régions de pondération (suite)

Caractéristique	Facteurs nationaux ou provinciaux	Rangs centiles des facteurs au niveau des RP						
		1	25	50	75	90	95	99
Dans les territoires	0,49	0,00	1,09	1,28	1,49	1,79	2,00	2,30
Lieu de travail – type de navettage								
Travaille dans la même SDR que sa résidence	1,06	0,45	0,96	1,11	1,24	1,36	1,43	1,56
Travaille dans une autre SDR que sa résidence	1,08	0,67	1,06	1,15	1,25	1,35	1,44	1,65
Catégorie de lieu de travail								
Travaillait à la maison, pas de lieu de travail fixe	1,27	0,62	1,13	1,24	1,35	1,46	1,53	1,67
Travaillait à l'extérieur du Canada	1,23	0,00	0,96	1,12	1,27	1,43	1,55	1,85
Lieu de travail habituel	0,99	0,46	0,85	0,95	1,05	1,14	1,19	1,31
Nombre de personnes dans les fourchettes de revenu total (\$)								
0 à 9 999	0,73	0,16	0,56	0,67	0,77	0,87	0,92	1,05
10 000 à 19 999, 20 000 à 29 999, 30 000 à 39 999, 40 000 à 49 999, 50 000 à 59 999, 60 000 à 69 999, 70 000 à 74 999, 75 000 ou plus	1,12	0,73	0,87	0,95	1,09	1,17	1,22	1,31
Situation dans la famille de recensement								
Époux, épouse	0,12	0,00	0,05	0,08	0,11	0,15	0,17	0,23
Enfant	0,49	0,23	0,37	0,43	0,50	0,58	0,64	0,75
Mère seule, père seul, non-membre d'une famille de recensement	0,80	0,44	0,71	0,83	1,02	1,16	1,23	1,41
Autre	0,46	0,00	0,21	1,01	1,33	1,66	1,87	2,40
Dans une famille de recensement –								
Présence d'un mari, d'une femme ou d'un conjoint de fait								
Oui, époux ou épouse	0,16	0,00	0,06	0,10	0,15	0,19	0,22	0,28
Oui, partenaire de même sexe	1,63	0,00	1,40	1,55	1,73	1,92	2,02	2,39
Oui, époux de même sexe	1,28	0,00	1,04	1,21	1,37	1,51	1,61	2,12
Oui, partenaire de sexe opposé	0,31	0,00	0,07	0,17	0,28	0,40	0,70	1,07
Non	0,94	0,43	0,81	0,90	1,00	1,10	1,18	1,41
Situation dans la famille économique								
Époux, épouse (y compris les partenaires de même sexe)	0,16	0,00	0,08	0,13	0,23	0,81	1,01	1,28
Parent seul, enfant	0,58	0,16	0,51	0,61	0,75	0,85	0,90	1,02
Autre membre de la famille	1,26	0,00	0,83	1,03	1,28	1,47	1,60	1,88
Toutes les autres caractéristiques de la population								
	1,00
Caractéristique du ménage et du logement								
Mode d'occupation								
	0,77	0,55	0,80	0,90	0,99	1,07	1,11	1,19

Tableau 9.2 Facteurs d'ajustement de l'erreur type au niveau national ou provincial et des régions de pondération (suite)

Caractéristique	Facteurs nationaux ou provinciaux	Rangs centiles des facteurs au niveau des RP						
		1	25	50	75	90	95	99
Période de construction	1,03	0,80	1,04	1,11	1,17	1,23	1,28	1,39
Nombre de pièces	1,09	0,78	1,04	1,10	1,17	1,23	1,27	1,39
Nombre de chambres	1,02	0,47	0,95	1,05	1,12	1,20	1,25	1,38
Type de construction								
Logement individuel non attenant, appartement dans un immeuble	0,47	0,20	0,38	0,66	0,84	0,94	1,01	1,24
Autre type de construction	0,89	0,30	0,89	1,03	1,15	1,25	1,34	1,55
Taille du ménage								
Ménage de 1 personne, de 2 personnes	0,06	0,00	0,00	0,00	0,00	0,05	0,10	0,24
Autre	0,32	0,00	0,00	0,00	0,09	1,05	1,26	1,73
Principal soutien du ménage	0,00
Âge du principal soutien du ménage								
15 à 19 ans	1,12	0,00	0,76	0,85	0,97	1,09	1,16	1,38
20 à 24, 25 à 29, 30 à 34, 35 à 39, 40 à 44, 45 à 49, 50 à 54, 55 à 59, 60 à 64	0,68	0,46	0,61	0,67	0,75	0,85	0,93	1,11
65+ ans	0,36	0,23	0,31	0,35	0,39	0,44	0,47	0,58
Sexe du principal soutien du ménage	0,70	0,47	0,64	0,71	0,81	0,89	0,93	1,01
Nombre de soutiens du ménage								
Un soutien du ménage	1,02	0,84	0,98	1,04	1,09	1,15	1,18	1,26
Plus d'un soutien du ménage	1,04	0,00	1,00	1,08	1,20	1,36	1,47	1,85
La personne repère est un soutien du ménage	1,89	1,03	1,61	1,82	2,00	2,18	2,28	2,51
Une personne qui ne vit pas là est un soutien du ménage	1,60	0,35	1,21	1,46	1,73	2,04	2,24	2,79
Nombre de ménages dans les tranches de loyer brut (tranches de 100 \$)	1,09	0,56	0,87	0,99	1,11	1,21	1,28	1,43
Nombre de ménages par loyer brut comme pourcentage des fourchettes de revenu des ménages								
Moins de 10 %, entre 10 et 20 %, 20 et 30 %, 30 et 40 %, 40 et 50 %, plus de 50 %	1,10	0,62	0,84	0,90	0,96	1,02	1,06	1,17
Nombre de ménages dans les tranches des principaux paiements du propriétaire (tranches de 200 \$)	1,02	0,70	0,87	0,95	1,07	1,15	1,19	1,28
Nombre de ménages dans les principaux paiements du propriétaire comme pourcentage des fourchettes de revenu des ménages								

Tableau 9.2 Facteurs d'ajustement de l'erreur type au niveau national ou provincial et des régions de pondération (suite)

Caractéristique	Facteurs nationaux ou provinciaux	Rangs centiles des facteurs au niveau des RP						
		1	25	50	75	90	95	99
Moins de 10 %, entre 10 et 20 %, 20 et 30 %, 30 et 40 %, 40 et 50 %, plus de 50 %	1,09	0,71	0,85	0,90	0,95	1,00	1,04	1,14
Personne responsable des paiements du ménage								
La personne est le principal soutien	0,00
Autre soutien	0,89	0,00	0,88	1,00	1,19	1,36	1,47	1,85
Nombre de ménages dans les fourchettes de revenu des ménages (fourchettes de 10 000 \$)	1,05	0,63	0,84	0,90	0,97	1,07	1,14	1,25
Nombre de ménages dans les fourchettes de valeurs des logements	0,92	0,66	0,97	1,07	1,15	1,22	1,27	1,40
Condominium enregistré								
Fait partie	0,88	0,62	0,96	1,06	1,16	1,27	1,34	1,51
Ne fait pas partie	0,72	0,47	0,79	0,90	0,99	1,07	1,11	1,19
État du logement								
Entretien régulier, réparations majeures ou mineures	1,09	0,90	1,06	1,12	1,17	1,23	1,26	1,35
Toutes les autres caractéristiques du ménage et du logement	1,00
Caractéristique de la famille de recensement								
Activité de l'époux, de l'épouse ou du parent seul								
Époux ou épouse dans la population active	0,61	0,36	0,56	0,65	0,73	0,80	0,85	0,93
Parent seul dans la population active	0,99	0,69	0,91	1,01	1,11	1,22	1,29	1,47
Âge des enfants à la maison	0,37	0,00	0,00	0,10	0,28	0,68	0,98	2,53
Activité professionnelle en 2005 de l'époux, de l'épouse ou du parent seul								
A travaillé en 2005	1,06	0,80	1,06	1,16	1,26	1,37	1,45	1,67
N'a pas travaillé en 2005	0,97	0,72	0,94	1,06	1,22	1,37	1,46	1,73
Toutes les autres caractéristiques de la famille de recensement	1,00
(Caractéristique de la famille économique)								
Nombre de ménages dans les fourchettes de revenu du travail autonome (\$)								
0 à 9 999	1,01	0,06	0,75	0,94	1,10	1,22	1,30	1,43
10 000 à 19 999, 20 000 à 29 999, 30 000 à 39 999, 40 000 à 49 999, 50 000 à 59 999, 60 000 à 69 999, 70 000 à 74 999, 75 000 ou plus	1,24	0,55	0,83	0,93	1,07	1,24	1,34	1,56

Tableau 9.2 Facteurs d'ajustement de l'erreur type au niveau national ou provincial et des régions de pondération (suite)

Caractéristique	Facteurs nationaux ou provinciaux	Rangs centiles des facteurs au niveau des RP						
		1	25	50	75	90	95	99
Situation de faible revenu								
Au-dessus du seuil	1,80	1,02	1,57	1,77	1,97	2,15	2,28	2,54
Au-dessous du seuil	1,93	1,38	1,73	1,89	2,09	2,27	2,39	2,64
Impôts fonciers inclus dans le versement hypothécaire								
Impôts inclus, impôts non inclus	1,06	0,89	1,03	1,09	1,15	1,21	1,24	1,34
Salaires et traitements (\$)								
	1,11	0,70	0,88	1,01	1,15	1,25	1,31	1,44
Langue maternelle de la personne repère de la famille – anglais								
Terre-Neuve, Nouvelle-Écosse, Île-du-Prince-Édouard, Territoire du Yukon ¹	0,35	0,13	0,26	0,34	0,43	0,51	0,56	0,83
Québec	1,01	0,72	1,02	1,10	1,18	1,28	1,34	1,50
Autres provinces	0,63	0,33	0,52	0,65	0,80	0,93	1,01	1,13
Canada	0,62
Langue maternelle de la personne repère de la famille – français								
Québec	0,48	0,16	0,31	0,46	0,71	0,90	1,00	1,14
Nouveau-Brunswick, Territoire du Yukon ¹	0,65	0,29	0,74	1,02	1,13	1,22	1,26	1,33
Autres provinces	1,03	0,66	1,04	1,13	1,22	1,31	1,37	1,56
Canada	0,56
Langue maternelle de la personne repère de la famille – autre que l'anglais ou le français								
	1,09	0,00	0,94	1,08	1,21	1,35	1,45	1,71
Toutes les autres caractéristiques de la famille économique								
	1,00

... n'ayant pas lieu de figurer

Note :

- En vigueur à partir du 20 octobre 2008, les noms « Territoire du Yukon » en français et « Yukon Territory » en anglais sont devenus « Yukon » en français et en anglais. Les utilisateurs doivent noter qu'il n'y a aucun changement aux abréviations ainsi qu'aux codes numérique et alphabétique du Yukon. Les abréviations demeurent les mêmes : « Yn » en français et « Y.T. » en anglais, et « 60 » comme code numérique et « YT » comme code alphabétique.

Source : Statistique Canada, Recensement de 2006.

10 Conclusion

L'échantillonnage fait maintenant partie intégrante du processus de recensement et est bien accepté. Son utilisation permet de faire des économies substantielles et de réduire de façon appréciable le fardeau du répondant, ou encore, d'élargir la portée d'un recensement sans frais supplémentaires. Le prix de ces avantages est l'introduction d'une erreur d'échantillonnage dans les estimations du recensement obtenues à partir de l'échantillon. Cet effet se fait particulièrement sentir sur les estimations de faible valeur du recensement, qu'il s'agisse des chiffres se rapportant à des catégories peu usuelles au niveau national ou provincial, ou de ceux qui correspondent à des régions géographiques de petite taille. Il convient de souligner que les erreurs de réponse et les erreurs de traitement contribuent également à l'erreur totale des estimations du recensement et ce sont les estimations de faible valeur qui sont particulièrement sensibles aux effets de ces erreurs non dues à l'échantillonnage. Par conséquent, bon nombre de ces faibles valeurs n'auraient qu'une fiabilité relative même si elles étaient fondées sur un recensement intégral. Pour ce qui est du Recensement de 2006, les estimations inférieures ou égales à 100 établies à partir des données-échantillon sont généralement très peu fiables, tandis que les estimations allant jusqu'à 500 tendent à avoir des erreurs types de plus de 10 % de leur valeur.

Pour de nombreuses caractéristiques, un certain biais a été détecté dans l'échantillon. Il est possible qu'une petite partie de ce biais ait été causé par le traitement des données et à l'étape du contrôle et de l'imputation. Le reste du biais s'expliquerait par un ou plusieurs facteurs tels que le biais de non-réponse, le biais de réponse ou la sélection par les agents recenseurs d'un échantillon biaisé. Le calibrage des estimations de l'échantillon en fonction des chiffres de population connus dans le cadre des procédures de pondération du recensement a contribué à atténuer les répercussions de ces biais.

Annexe A Glossaire

Les définitions des termes, variables et concepts du recensement sont présentées ici telles qu'elles figurent dans le *Dictionnaire du Recensement de 2006* (produit n° 92-566-XWF au catalogue). L'utilisateur doit se reporter au *Dictionnaire du Recensement de 2006* pour obtenir les définitions complètes et des observations additionnelles se rapportant aux concepts, comme des renseignements sur les variables directes et dérivées, ainsi que leur univers respectif.

Aire de diffusion (AD) : Une aire de diffusion (AD) est une petite unité géographique relativement stable formée de un ou de plusieurs îlots de diffusion avoisinants. Il s'agit de la plus petite région géographique normalisée pour laquelle toutes les données du recensement sont diffusées. Les AD couvrent tout le territoire du Canada.

Division de recensement (DR) : Division de recensement (DR) est le terme général de régions créées en vertu des lois provinciales (comme les comtés, les municipalités régionales de comté et les *regional districts*) ou des régions équivalentes. Les divisions de recensement sont des régions géographiques intermédiaires entre la municipalité (subdivision de recensement) et la province/territoire.

Logement collectif : Établissement commercial, institutionnel ou communautaire que le recenseur peut identifier comme tel grâce à une enseigne ou en s'adressant à la personne qui en a la charge, à un résident, à un voisin, etc. Sont inclus dans cette catégorie les pensions et maisons de chambres, les hôtels, motels et maisons de chambres pour touristes, les maisons de repos, les hôpitaux, les résidences de personnel, les casernes (bases militaires), les camps de chantier, les prisons, les centres d'accueil, les foyers collectifs, etc. Il peut s'agir d'un logement occupé par des résidents habituels ou uniquement par des résidents étrangers et/ou temporaires.

Logement privé : Ensemble distinct de pièces d'habitation ayant une entrée privée donnant sur l'extérieur ou sur un corridor, un hall, un vestibule ou un escalier commun à l'intérieur. L'entrée doit donner accès au logement sans que l'on ait à passer par les pièces d'habitation de quelqu'un d'autre.

Logement privé occupé : Logement privé occupé de façon permanente par une personne ou un groupe de personnes. Sont également inclus dans cette catégorie les logements privés dont les résidents habituels sont temporairement absents le jour du recensement. Sauf indication contraire, toutes les données présentées dans les produits sur le logement ont trait aux logements privés occupés et non aux logements privés inoccupés ou aux logements occupés par des résidents étrangers et/ou temporaires uniquement.

Ménage : Personne ou groupe de personnes (autres que des résidents étrangers) occupant un même logement et n'ayant pas de domicile habituel ailleurs au Canada. Il peut se composer d'un groupe familial (famille de recensement) avec ou sans autres personnes, de deux familles ou plus partageant le même logement, d'un groupe de personnes non apparentées ou d'une personne seule. Les membres d'un ménage qui sont temporairement absents le jour du recensement (p.ex., qui résident temporairement ailleurs) sont considérés comme faisant partie de leur ménage habituel. Pour les besoins du recensement, chaque personne est membre d'un seul et unique ménage. À moins d'indications contraires, toutes les données contenues dans les rapports sur les ménages se rapportent aux ménages privés seulement.

Ménage privé : Personne ou groupe de personnes (autres que des résidents étrangers) occupant un logement privé et n'ayant pas de domicile habituel ailleurs au Canada.

Secteur de recensement (SR) : Les secteurs de recensement (SR) sont de petites régions géographiques relativement stables qui comptent habituellement entre 2 500 et 8 000 habitants. Ils sont créés au sein de régions métropolitaines de recensement et d'agglomérations de recensement dont le noyau urbain compte 50 000 habitants ou plus d'après le recensement précédent.

Un comité de spécialistes locaux (par exemple, des planificateurs, des travailleurs sociaux, des travailleurs du secteur de la santé et des éducateurs) délimite initialement les secteurs de recensement de concert avec Statistique Canada. Une fois qu'une région métropolitaine de recensement (RMR) ou qu'une agglomération de recensement (AR) a été divisée en secteurs de recensement, les secteurs de recensement sont maintenus même si, ultérieurement, la population du noyau urbain de la RMR ou de l'AR devient inférieure à 50 000 habitants.

Subdivision de recensement (SDR) : Subdivision de recensement (SDR) est un terme générique qui désigne les municipalités (telles que définies par les lois provinciales/territoriales) ou les territoires considérés comme étant des équivalents municipaux à des fins statistiques (par exemple, les réserves indiennes, les établissements indiens et les territoires non organisés).

Unité de collecte (UC) : Petite unité géographique utilisée pour la collecte des données du recensement. Les unités de collecte englobent le territoire du Canada au complet.

Annexe B Contraintes au niveau des RP et des AD appliquées aux poids des recensements de 2006 et de 2001

Contraintes au niveau des RP et des AD (personnes)	Contraintes au niveau des RP et des AD (ménages)
-Total des personnes	-Ménages de 1 personne
-Total des personnes âgées de 15 ans et plus	-Ménages de 2 personnes
	-Ménages de 3 personnes
-Hommes	-Ménages de 4 personnes
-Hommes âgés de 15 ans et plus	-Ménages de 5 personnes
	-Ménages de 6 personnes ou plus
	-Nombre total de ménages
-Personnes âgées de 0 à 4 ans	
-Personnes âgées de 5 à 9 ans	-Type de logement (nouveau en 2006)
-Personnes âgées de 10 à 14 ans	-Appartements de moins de cinq étages (nouveau en 2006)
-Personnes âgées de 15 à 19 ans	
-Personnes âgées de 20 à 24 ans	
-Personnes âgées de 25 à 29 ans	Autres contraintes au niveau des AD
-Personnes âgées de 30 à 34 ans	-Total des ménages dans l'AD
-Personnes âgées de 35 à 39 ans	-Total des personnes dans l'AD
-Personnes âgées de 40 à 44 ans	
-Personnes âgées de 45 à 49 ans	
-Personnes âgées de 50 à 54 ans	
-Personnes âgées de 55 à 59 ans	
-Personnes âgées de 60 à 64 ans	
-Personnes âgées de 65 à 74 ans	
-Personnes âgées de 75 ans et plus	
-Personnes mariées	
-Personnes célibataires	
-Personnes divorcées	
-Veufs et veuves	
-Personnes séparées	
-Union libre = oui	

Annexe C Statistiques utilisées dans l'étude du biais d'échantillonnage

On mentionne au [chapitre 6](#) que selon le mode d'échantillonnage aléatoire, la valeur suivra une distribution à peu près normale (0,1). Une justification de ceci est donnée ici. L'échantillonnage a été fait indépendamment pour chaque UC. Par conséquent, $\hat{X}^{(0)}$ est la somme de H variables aléatoires indépendantes, où H est le nombre de UC au Canada. Comme il y a 46 510 UC au Canada, alors H est très élevé. Par conséquent, selon le théorème central limite, $(\hat{X}^{(0)} - E(\hat{X}^{(0)})) / \sqrt{V(\hat{X}^{(0)})}$ suivra une distribution à peu près normale (0,1) (voir Kendall et Stuart, 1963, p. 193), tout comme $Z^{(0)} = (\hat{X}^{(0)} - X) / \sqrt{V(\hat{X}^{(0)})}$ si $E(\hat{X}^{(0)}) = X$. $Z^{(0)}$, cependant, n'aurait pas une moyenne de 0 si les échantillons de ménages au niveau des UC présentaient un biais important, pour quelque raison que ce soit. $E(\hat{X}^{(0)}) = X$

Calculons maintenant une autre statistique permettant de vérifier si le biais est le même entre deux régions ou deux recensements. Soit $\hat{X}_1^{(0)}$ et $\hat{X}_2^{(0)}$ des estimateurs (fondés sur les poids initiaux) des chiffres de population connus X_1 et X_2 pour deux régions géographiques ou deux recensements. Soit $RB(\hat{X}_1^{(0)}) = (E(\hat{X}_1^{(0)}) - X_1) / X_1$ et $RB(\hat{X}_2^{(0)}) = (E(\hat{X}_2^{(0)}) - X_2) / X_2$ les biais relatifs de $\hat{X}_1^{(0)}$ et $\hat{X}_2^{(0)}$. Nous voulons vérifier si l'hypothèse nulle $H_0 : RB(\hat{X}_1^{(0)}) = RB(\hat{X}_2^{(0)})$ est vraie. Pour ce faire, on peut utiliser la statistique

où $rb(\hat{X}_1^{(0)}) = (\hat{X}_1^{(0)} - X_1) / X_1$ et $rb(\hat{X}_2^{(0)}) = (\hat{X}_2^{(0)} - X_2) / X_2$ sont des estimateurs non biaisés de $RB(\hat{X}_1^{(0)})$ et $RB(\hat{X}_2^{(0)})$ respectivement. Par conséquent, si l'hypothèse nulle H_0 ci-dessus est vraie, l'espérance de W est zéro. Il convient également de prendre note que le dénominateur de W est l'erreur type du numérateur de W (il n'y a pas de terme de covariance parce que les estimations établies pour des régions différentes ou des recensements différents sont indépendantes), de sorte que W a une variance de 1. Maintenant si $\hat{X}_1^{(0)}$ suit une distribution à peu près normale (encore selon le théorème central limite), $rb(\hat{X}_1^{(0)})$ suivra aussi une distribution à peu près normale, tout comme $rb(\hat{X}_2^{(0)})$ et $rb(\hat{X}_1^{(0)}) - rb(\hat{X}_2^{(0)})$. Par conséquent, W suit une distribution à peu près normale (0,1) si l'hypothèse nulle H_0 est vraie.

Bibliographie

- BANKIER, M. 2002. *Pondération pour le Recensement du Canada de 2001*, Ottawa, Statistique Canada. Actes du Symposium de 2002 de Statistique Canada.
- BUREAU FÉDÉRAL DE LA STATISTIQUE. 1968. *Sampling in the Census*, Ottawa, Statistique Canada. Rapport interne.
- COCHRAN, W. 1977. *Sampling Techniques*, 3^e éd., Toronto, John Wiley and Sons.
- Dick, Peter. 2007. 'The 2006 Dwelling Classification Survey'. Report for the Federal-Provincial-Territorial Meeting on Demography September 21, 2007.
- FELLEGI, Ivan P. 1964. « Response Variance and its Estimation », *Journal of the American Statistical Association*, vol. 59, décembre, p. 1016 à 1041.
- FULLER, Wayne A. 2002. « Estimation par régression appliquée à l'échantillonnage », *Techniques d'enquête*, vol. 28, n^o 1, p. 5 à 25.
- HANSEN, M.H., W.N. HURWITZ et M.A. BERSHAD. 1959. « Measurement Errors in Censuses and Surveys », *Bulletin of the International Statistical Institute*, vol. 38, p. 359 à 374.
- HOVINGTON, Édith. 2004. *Étude de l'effet de plan de la variance d'échantillonnage pour le recensement de 2001*, Ottawa, Statistique Canada. Rapport interne.
- KENDALL, Maurice G., Sir, et Alan STUART. 1963. *The Advanced Theory of Statistics*, London, Charles Griffin and Company Limited, vol. 1.
- KRUSZYNSKI, G. 1999. *Evaluation of the 1996 Weighting Areas*, Ottawa, Statistique Canada, Division de la géographie. Rapport interne.
- PRESS, W.H., S.A. TEUKOLSKY, W.T. VETTERLING et B.P. FLANNERY. 1992. *Numeric Recipes in C*, New York, Cambridge University Press.
- ROYCE, Don. 1983. *The Use of Sampling in the 1981 Canadian Census*, Ottawa, Statistique Canada. Rapport interne.
- SARNDAL, C., B. SWENSSON et J. WRETMAN. 1992. *Model Assisted Survey Sampling*, New York, Springer-Verlag.
- STATISTIQUE CANADA. 2002a. *Dictionnaire du recensement de 2001*, produit n^o 92-378-XIF au catalogue de Statistique Canada, Ottawa, « série des produits de référence du recensement de 2001 ».
- . 2002b. *Le recensement de 2001 en bref*, produit n^o 92-379-XIF au catalogue de Statistique Canada, Ottawa, « série des produits de référence du recensement de 2001 ».
- . 2006. *Dictionnaire du Recensement de 2006*, produit n^o 92-566-XWF au catalogue de Statistique Canada, version mise à jour le 14 février 2008, Ottawa, Ontario, <http://www12.statcan.ca/francais/census06/reference/dictionary/index.cfm> (site consulté le 5 novembre 2008).

Centre de contact national de Statistique Canada

Courrier électronique : infostats@statcan.gc.ca

8 h 30 à 16 h 30 du lundi au vendredi

Téléphone sans frais (Canada et États-Unis) :

- 1-800-263-1136 – Service de renseignements
- 1-800-363-7629 – Service national d'appareils de télécommunications pour les malentendants
- 1-877-287-4369 – Télécopieur

Appels locaux ou internationaux :

- 1-613-951-8116 – Service de renseignements
- 1-613-951-0581 – Télécopieur