



Numéro 92-395-XIF au catalogue

## Échantillonnage et pondération

Rapport technique du recensement de 2001



Statistique  
Canada

Statistics  
Canada

Canada

PUBLICATIONS ÉLECTRONIQUES DISPONIBLES À  
**[www.statcan.ca](http://www.statcan.ca)**



## Échantillonnage et pondération

	Page
<b>INTRODUCTION.....</b>	<b>3</b>
<b>1. COLLECTE DES DONNÉES DU RECENSEMENT .....</b>	<b>5</b>
1.1 Généralités.....	5
1.2 Méthodes de collecte.....	6
<b>2. TRAITEMENT DES DONNÉES DU RECENSEMENT .....</b>	<b>7</b>
2.1 Généralités.....	7
2.2 Dépouillement régional.....	8
2.3 Imagerie.....	9
2.4 Vérification interactive.....	9
2.5 Codage automatisé et codage interactif.....	10
2.6 Contrôle et imputation.....	10
2.7 Ajustements de la couverture en fonction des logements inoccupés et des logements non répondants.....	11
2.8 Pondération.....	11
<b>3. ÉCHANTILLONNAGE DANS LES RECENSEMENTS CANADIENS.....</b>	<b>13</b>
3.1 Historique de l'échantillonnage au recensement canadien.....	13
3.2 Plan d'échantillonnage utilisé lors du recensement de 2001.....	14
<b>4. ESTIMATIONS BASÉES SUR L'ÉCHANTILLON DU RECENSEMENT .....</b>	<b>16</b>
4.1 Considérations d'ordre opérationnel.....	16
4.2 Considérations d'ordre théorique.....	16
4.3 Élaboration d'une méthode d'estimation pour l'échantillon du recensement.....	17
4.4 Estimateur par régression généralisée en deux étapes.....	18
4.5 Traitement à deux passages.....	21
<b>5. PROGRAMME D'ÉVALUATION DE L'ÉCHANTILLONNAGE ET DE LA PONDÉRATION .....</b>	<b>23</b>
5.1 Biais d'échantillonnage.....	23
5.2 Évaluation des méthodes de pondération.....	23
5.3 Cohérence entre les estimations-échantillon et les chiffres de population.....	24
5.4 Variance d'échantillonnage.....	24
<b>6. BIAIS D'ÉCHANTILLONNAGE.....</b>	<b>25</b>
<b>7. ÉVALUATION DES MÉTHODES DE PONDÉRATION.....</b>	<b>32</b>
7.1 Formation des régions de pondération (RP).....	32
7.2 Évaluation de la méthode de pondération du recensement.....	34
7.2.1 Répartition des poids d'échantillonnage.....	34
7.2.2 Écart entre les chiffres de population et les estimations-échantillon.....	37
7.2.3 Retranchement des contraintes.....	43

	<b>Page</b>
<b>8. COHÉRENCE ENTRE LES ESTIMATIONS-ÉCHANTILLON ET LES CHIFFRES DE POPULATION .....</b>	<b>48</b>
8.1 Aires de diffusion .....	49
8.2 Régions de pondération.....	49
8.3 Subdivisions de recensement.....	50
8.4 Secteurs de recensement.....	50
8.5 Divisions de recensement.....	51
<b>9. VARIANCE D'ÉCHANTILLONNAGE .....</b>	<b>59</b>
<b>10. CONCLUSION.....</b>	<b>73</b>
 <b>ANNEXES</b>	
Annexe A. Glossaire .....	74
Annexe B. Contraintes au niveau des RP et des AD appliquées aux poids d'échantillonnage des recensements de 2001 et de 1996 .....	76
Annexe C. Statistiques utilisées dans l'Étude du biais d'échantillonnage .....	78
Annexe D. Produits et services du recensement de 2001 .....	79
<b>BIBLIOGRAPHIE.....</b>	<b>80</b>

# Introduction

Le recensement de 2001 a nécessité la participation de l'ensemble de la population canadienne, soit environ 30 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 2001 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 2001* porte sur la méthode d'échantillonnage et de pondération utilisée au recensement de 2001 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, Édith Hovington 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 (par exemple, l'évaluation de la durée de vie des ampoules électriques) et par conséquent, l'échantillonnage s'avère essentiel. Le désavantage de cette méthode 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 peut être 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 2001, on a appliqué les méthodes d'échantillonnage de plusieurs façons. On y a eu recours pour s'assurer que la qualité du travail des 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 2001. 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 les études effectuées dans le cadre du programme d'évaluation des méthodes d'échantillonnage et de pondération utilisées au recensement de 2001. Les chapitres 6 à 9 donnent les résultats de ces études. Le chapitre 10 présente certaines conclusions sur les méthodes de pondération utilisées en 2001.

Les utilisateurs trouveront des renseignements supplémentaires sur les concepts, les variables et la géographie du recensement dans le *Dictionnaire du recensement de 2001* (n° 92-378-XIF au catalogue), de même qu'un aperçu des différentes étapes du recensement dans *Le recensement de 2001 en bref* (n° 92-379-XIF au catalogue).

# 1. Collecte des données du recensement

## 1.1 Généralités

L'étape de la collecte des données du processus du recensement vise à faire en sorte que les quelque 11,8 millions de ménages du Canada soient dénombrés le jour du recensement (le mardi 15 mai 2001). Le recensement vise l'ensemble de la population du Canada, qui comprend les citoyens canadiens (par naissance ou par naturalisation), les immigrants reçus et les résidents non permanents. Les résidents non permanents sont les personnes au Canada qui sont titulaires d'un permis de travail, d'un permis de séjour pour étudiants ou d'un permis ministériel, ou qui revendiquent le statut de réfugié, ainsi que les membres de leur famille qui vivent avec eux.

Le recensement dénombre également les citoyens canadiens et les immigrants reçus qui sont 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 armées canadiennes en poste à l'étranger et les Canadiens qui se trouvent à bord de navires marchands. Comme le Recensement du Canada vise à dénombrer les personnes à l'endroit où elles résident habituellement plutôt qu'à l'endroit où elles se trouvent le jour du recensement, il est considéré comme un recensement *de jure*, ce qui signifie que les personnes se trouvant à l'étranger le jour du recensement ont été dénombrées si leur lieu habituel de résidence se trouvait au Canada. Certains pays mènent un recensement *de facto*, qui consiste à dénombrer les personnes à l'endroit où elles se trouvent le jour du recensement, même s'il ne s'agit pas de leur lieu habituel de résidence.

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 dans le secteur de dénombrement. Le RV est utilisé pour établir une liste d'adresses qui servira au cours des opérations de collecte et aux fins du contrôle.

La formule 2A est le questionnaire abrégé, qui comporte dix questions et est distribué à quatre ménages sur cinq. La formule 2B est un questionnaire complet qui renferme, outre les dix 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 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.

Une formule 4 est remplie par le personnel du recensement dans les cas où les occupants du ménage sont absents ou refusent de remplir un questionnaire. On consigne les renseignements sur les logements privés qui sont inoccupés le jour du recensement sur un questionnaire 2A ou 2B.

La formule 3 (A et B) 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.

Les Canadiens en poste à l'étranger (notamment le personnel des ambassades canadiennes ou les membres des Forces armées) sont dénombrés à l'aide d'une formule 2C, laquelle comporte les mêmes questions que le questionnaire 2B sans toutefois compter les questions relatives au logement. D'autres questions visant à recueillir des données sur le lieu habituel de résidence de ces personnes au Canada sont également posées.

## 1.2 Méthodes de collecte

Pour assurer la meilleure couverture possible aux fins de la collecte, le Canada est divisé en petits secteurs géographiques appelés secteurs de dénombrement (SD). Chaque recenseur est affecté à un SD et remplit les tâches liées aux cartes, au listage, à la distribution et à la vérification; il s'assure en outre que tous les questionnaires sont retournés aux centres de dépouillement. Le nombre de ménages dans un SD varie de 175 en région rurale à 600 en région urbaine. Au recensement de 2001, on comptait 42 851 secteurs de dénombrement au Canada. Les recenseurs travaillent sous la surveillance de commissaires au recensement (CR). En 2001, les 2 917 CR affectés au recensement étaient chargés d'embaucher les recenseurs et de veiller à la planification et à la gestion des activités de collecte sur le terrain dans leur zone d'affectation.

En 2001, environ 98 % des ménages ont été recensés selon la méthode de l'autodénombrement. Aux termes de cette méthode, un recenseur livre un questionnaire à chaque ménage au cours des deux semaines précédant le jour du recensement et un adulte ou un autre membre responsable du ménage est invité à remplir le questionnaire pour tous les membres du ménage, puis à le retourner dans une enveloppe-réponse.

Par ailleurs, environ 2 % des ménages ont été dénombrés par interview lors du recensement de 2001. Selon cette méthode, un recenseur visite le ménage et remplit lui-même par interview un questionnaire pour le ménage. Cette méthode est 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 certaines régions urbaines où l'on estime qu'il existe une forte probabilité que les répondants ne retournent pas leur questionnaire.

Les recenseurs et les CR prennent part à un certain nombre d'activités de collecte sur le terrain. Ils doivent notamment communiquer avec des ménages afin de régler des problèmes qui ont souvent trait à l'exhaustivité ou à la cohérence de l'information fournie. Ils doivent aussi traiter les cas où les répondants n'ont pas retourné leur questionnaire.

Au cours des opérations de collecte sur le terrain, les recenseurs devaient livrer un questionnaire à chaque logement situé dans leur SD et inscrire l'adresse et le nom de la personne (si possible) dans leur Registre des visites (RV). Ils devaient en outre copier les numéros d'identification uniques devant être saisis ultérieurement et utilisés pour associer chaque ménage et chaque logement à la région géographique appropriée. Ensuite, ils devaient déterminer le numéro d'îlot correspondant au logement sur leur carte de SD et l'inscrire dans le RV et sur le questionnaire. Ces numéros d'îlot ont ensuite été saisis, de façon à permettre d'associer tous les logements du Canada à un îlot donné.

## 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 d'extraction exacte et complète. Il s'agit ici de la saisie manuelle et automatisée des données, de l'imagerie des questionnaires, du contrôle, de la correction des erreurs, du codage, de l'imputation et de la pondération. La base de données finale a été transférée au Projet d'évaluation de la qualité des données, en vue de la détermination de la qualité globale des données, et au Projet de la diffusion, en vue de la production et de la commercialisation des produits et services du recensement de 2001. Dans le reste du présent chapitre, chaque opération de traitement des données est passée en revue.

Une innovation importante a été apportée au recensement de 2001; il s'agit de la création d'un système d'extraction d'images donnant accès aux images de tous les questionnaires et de tous les registres des visites (voir la section 2.3), de façon à éliminer la manutention de milliers de boîtes et de documents papier lors des processus subséquents nécessitant la consultation de l'original des formules du recensement, comme lors des recensements antérieurs.

### 2.2 Dépouillement régional

L'équipe du dépouillement régional était responsable de la saisie de l'information figurant dans les questionnaires en format électronique pour les systèmes de traitement ultérieurs ainsi que des recherches manuelles et du codage manuel des réponses aux questions sur l'industrie et la profession tirées des questionnaires 2B. Étant donné l'énorme volume de questionnaires et d'information à saisir (plus de 4 milliards de frappes), la Tâche du dépouillement régional, depuis 1981, confie le travail en sous-traitance à l'Agence des douanes et du revenu du Canada (ADRC), anciennement Revenu Canada. Celle-ci utilise son réseau de systèmes, de ressources et de personnel pour saisir et coder les données du recensement. Ce partenariat permet à l'équipe du recensement de réaliser des économies en mettant à profit le personnel qualifié et l'infrastructure déjà en place à l'ADRC. En ce qui a trait à la qualité des données, il lui permet en outre de bénéficier de l'expérience acquise par l'ADRC dans le cadre du dépouillement des questionnaires des recensements antérieurs. Lors du recensement de 2001, environ 2 800 employés de l'ADRC ont prêté le serment d'office et de discrétion prévu par la *Loi sur la statistique* en vue d'effectuer les tâches du recensement. Aux termes de cette loi, les employés de l'ADRC sont soumis aux mêmes règles et règlements que les employés de Statistique Canada.

Une fois les activités de collecte relatives à un secteur de dénombrement (SD) terminées, les questionnaires, les cartes géographiques et les registres des visites connexes étaient acheminés dans des boîtes de SD des sous-sections de la collecte sur le terrain vers un des huit centres fiscaux de l'ADRC désignés du pays.

La première étape consistait à préparer les questionnaires remplis en vue de la saisie des données. Traditionnellement, elle avait pour objet d'attribuer manuellement des codes aux réponses en lettres fournies par les répondants. En 2001, la plupart des réponses en lettres étaient converties en codes par des systèmes automatisés (voir la section 2.5). Les seules réponses en lettres devant être codées manuellement lors du recensement de 2001 étaient celles du questionnaire complet (2B) portant sur les questions relatives à l'industrie et la profession. Des travaux de recherche ont été entrepris en vue d'automatiser le codage de ces réponses et on s'attend à disposer d'un système automatisé opérationnel pour le recensement de 2006.

Les réponses aux questions sur l'industrie ont été codées à l'ADRC conformément au Système de classification des industries de l'Amérique du Nord (SCIAN), que Statistique Canada utilise comme document normatif depuis quelques années. Le SCIAN a été élaboré afin de fournir au Canada, aux États-Unis et au Mexique un cadre de référence commun pour la production de statistiques sur l'industrie

aux termes de l'Accord de libre-échange nord-américain (ALENA). Cela représente une modification au codage des réponses relatives à l'industrie qui, lors du recensement de 1996, avaient été codées à l'aide de la Classification type des industries (CTI) de 1980. Afin de permettre l'établissement de comparaisons longitudinales, les réponses recueillies sur les questionnaires complets (2B) à la question sur l'industrie du recensement de 2001 sont également codées à l'aide de la CTI de 1980 dans le cadre du codage automatisé (voir la section 2.5).

Une fois que les questionnaires avaient été reçus et enregistrés à un des centres fiscaux de l'ADRC et que les codes d'industrie et de profession avaient été attribués, l'étape suivante consistait à trier, étiqueter et mettre en lots les questionnaires en vue de la saisie des données. L'étiquette apposée sur chacun des questionnaires portait un numéro de séquence unique destiné à permettre de contrôler le cheminement du questionnaire tout au long des opérations à l'ADRC. Pour la première fois, l'étiquette comportait aussi un code à barres ayant pour objet de faciliter le balayage du questionnaire lors de la Tâche de l'imagerie (voir la section 2.3).

La saisie des données se faisait ensuite selon la méthode traditionnelle de l'entrée au clavier. Afin de vérifier la précision de l'opération de saisie, on saisissait de nouveau les données d'un échantillon de questionnaires dont les données avaient déjà été saisies. Des statistiques sur le contrôle de la qualité étaient produites en comparant entre eux les deux ensembles de données saisies. Comme on peut s'y attendre, l'entrée des données au clavier à partir des questionnaires peut se traduire par des erreurs, attribuables à diverses raisons : erreur de saisie, réponses difficiles à lire ou mal indiquées sur les questionnaires ou réponses omises par l'opérateur de saisie des données. La vérification des entrées permet de réduire les erreurs le plus possible.

Au fur et à mesure de leur saisie, les données étaient transmises en temps réel sur des lignes de communication dédiées à l'ordinateur de l'ADRC à Ottawa. À l'intérieur d'un délai de 24 heures, les données étaient ensuite transférées sur des cartouches magnétiques qui étaient livrées par un transporteur cautionné à Statistique Canada, où elles étaient chargées dans l'ordinateur central. Les questionnaires étaient enfin rassemblés dans leurs boîtes de SD en vue de leur expédition au centre de dépouillement de Statistique Canada à Ottawa. Une fois que les données avaient été saisies et transférées à Statistique Canada et que celui-ci avait accusé réception de toutes les données, l'ADRC n'avait plus aucune donnée du recensement en sa possession.

## 2.3 Imagerie

Lors des recensements antérieurs, les étapes du dépouillement qui nécessitaient la consultation des questionnaires et des registres des visites (RV) se faisaient à l'aide des documents papier. En 2001, on a éliminé la nécessité de manipuler les documents papier en créant (par balayage) une image électronique de tous les questionnaires et RV dès leur réception au centre de dépouillement du recensement en provenance des centres fiscaux de l'Agence des douanes et du revenu du Canada (ADRC). Le personnel des opérations ultérieures pouvait ensuite accéder aux images des questionnaires et des RV à l'aide d'un système d'extraction d'images, plutôt que d'avoir à utiliser les documents papier.

À leur arrivée au centre de dépouillement du recensement de 2001, les boîtes de secteurs de dénombrement (SD) étaient enregistrées et les documents étaient préparés en vue du balayage.

Les images électroniques des 13 millions de documents (en majorité des questionnaires) ont été créées à l'aide de 15 scanners à haute vitesse fonctionnant cinq jours par semaine, pendant deux quarts de travail par jour. L'identificateur géographique nécessaire pour identifier chacune des images de document était automatiquement attribué à partir du code à barres figurant sur l'étiquette apposée lors des opérations de saisie des données à l'ADRC (voir la section 2.2). Un contrôle qualitatif était effectué afin de vérifier que chaque document comportait le nombre approprié de pages et que le nombre de questionnaires de chaque type était exact pour chacun des SD. Une opération spéciale était prévue afin de résoudre les problèmes qui se posaient. Les images étaient alors enregistrées sur des disques optiques à des fins de consultation ultérieure et d'archivage. Elles étaient aussi conservées sur une

mémoire cache magnétique afin qu'on puisse y avoir accès immédiatement pour les besoins des activités de la tâche de la Vérification interactive.

## 2.4 Vérification interactive

L'objectif principal de la Vérification interactive était d'identifier et de corriger les erreurs dans les données dont la résolution exigeait de se référer aux images des questionnaires et/ou des registres des visites. Les données saisies étaient soumises à un ensemble détaillé de règles de vérification pour identifier des erreurs telles que des personnes manquantes ou comptées plus d'une fois dans un ménage; un dénombrement incorrect des résidents étrangers ou temporaires; des ménages associés aux mauvais questionnaires; ou une mauvaise classification des logements (occupés ou inoccupés). Un examen exhaustif de l'information des formules associées à chacun des problèmes était effectué pour déterminer la correction appropriée. Dans certains cas il fallait ajouter et/ou éliminer des personnes ou des logements.

Lorsque les cartouches renfermant les données étaient reçues de l'Agence des douanes et du revenu du Canada (ADRC), elles étaient chargées sur les ordinateurs de Statistique Canada en vue de la Vérification interactive. Une série de contrôles « structurels » automatisés étaient alors exécutés, surtout pour vérifier les renseignements inscrits par le recenseur sur la page couverture des questionnaires. Ces contrôles avaient entre autres pour objet d'apparier les types de questionnaire et les types de ménage, de contre-vérifier le nombre de questionnaires, de contre-vérifier le nombre de personnes dénombrées et de vérifier si les identificateurs géographiques étaient bien uniques. Les données sur le revenu tirées des questionnaires 2B étaient aussi soumises à certains contrôles visant à déceler les anomalies et à les soumettre aux spécialistes des données sur le revenu.

Tous les contrôles étaient effectués par secteur de dénombrement (SD). Les erreurs étaient identifiées, puis corrigées en se reportant aux images des questionnaires et des registres des visites (RV) du SD visé. Les corrections apportées aux données électroniques étaient faites à l'aide d'un système interactif sur micro-ordinateur. Certaines corrections étaient également notées sur les images de questionnaire ou sur les registres de visite à l'aide d'un processus électronique.

Une fois que la boîte de SD avait été soumise à tous ces contrôles, on procédait à une vérification manuelle et automatique des identificateurs géographiques transcrits par le recenseur de la carte de SD au questionnaire et au RV.

La Vérification interactive prévoyait aussi un dépouillement spécial afin d'assurer que les Canadiens qui résidaient à l'extérieur du Canada le jour du recensement (personnes à bord des navires de la garde côtière, des navires de guerre et des navires marchands battant pavillon canadien, membres du personnel diplomatique et membres des Forces armées) avaient été dénombrés correctement.

La dernière étape de la Vérification interactive consistait à reformater les données et à les acheminer aux dernières étapes du dépouillement, c.-à-d. le Codage automatisé ainsi que le Contrôle et imputation.

## 2.5 Codage automatisé et codage interactif

Le codage automatisé consistait à convertir en codes les réponses en lettres saisies sur les questionnaires complets lors du dépouillement régional (voir la section 2.2), à l'aide d'un fichier de référence ou d'une structure de classification automatisé contenant des séries de mots ou d'expressions et leur code numérique correspondant. Bien que la majorité des réponses en lettres puissent être codées de façon entièrement automatique, certaines réponses ne peuvent être appariées. Des systèmes de codage interactifs pour PC sont prévus pour aider des codeurs ayant reçu une formation spéciale et des spécialistes à attribuer un code approprié à toutes les réponses non appariées en examinant les réponses données à d'autres questions, lesquelles se rapportaient parfois à d'autres membres du

ménage. Ont été soumises au codage automatisé les réponses en lettres aux questions suivantes du questionnaire complet (2B) :

- lien avec la Personne 1;
- langue parlée à la maison;
- langues non officielles;
- première langue apprise dans l'enfance (langue maternelle);
- langue utilisée au travail;
- lieu de naissance;
- lieu de naissance des parents;
- citoyenneté;
- origine ethnique (ascendance);
- groupe de population;
- bande indienne/Première nation;
- lieu de résidence 1 an auparavant;
- lieu de résidence 5 ans auparavant;
- principal domaine d'études;
- religion;
- lieu de travail;
- industrie selon la Classification type des industries (CTI) de 1980.

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

## 2.6 Contrôle et imputation

Les données recueillies lors de toute enquête ou recensement comportent des omissions ou des incohérences. Ces erreurs peuvent survenir lorsque les répondants fournissent des réponses incorrectes ou incomplètes aux questions ou encore lors du traitement. 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. De son côté, le personnel du recensement peut coder les réponses incorrectement ou peut faire d'autres types d'erreurs au cours du traitement.

Une des premières tâches de l'opération de contrôle et d'imputation consiste à vérifier si la taille du ménage est indiquée pour tous les logements classés comme étant « occupés ». Pour les logements occupés associés non pas à un questionnaire ordinaire (formule 2A ou 2B) mais à une formule 4 établie en cas de non-réponse, la première mesure prise à l'opération du contrôle et de l'imputation était de s'assurer que la taille du ménage était valide. Si la taille du ménage était « inconnue », la procédure consistait à imputer la taille du ménage à partir du plus proche voisin. En outre, une nouvelle procédure a été introduite pour 2001 afin d'imputer à nouveau la taille du ménage dans certains des cas de logements associés à une formule 4, en fonction de l'Étude sur la classification des logements (voir la section 2.7).

L'épuration définitive des données, faite à l'étape du contrôle et de l'imputation, a été presque entièrement automatisée. Elle consistait à appliquer une série de règles de contrôle détaillées, qui permettaient de repérer les réponses manquantes ou incohérentes. Ces dernières étaient la plupart du temps corrigées en modifiant par imputation la valeur du plus petit nombre de variables possible. L'imputation était faite à l'aide des méthodes dites **déterministe** ou **hot deck à changements minimaux**. L'imputation déterministe consiste à corriger les erreurs en inférant la valeur appropriée à partir des réponses à d'autres questions. Quant à l'imputation « hot deck 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 rejets au contrôle.

Deux systèmes automatisés différents ont été utilisés pour mener à bien ce traitement.

Le système NIM (**N**earest-**n**eighbour **I**mputation **M**ethod), qui a été élaboré en vue du recensement de 1996 pour faire le contrôle et l'imputation des caractéristiques démographiques de base telles que l'âge, le sexe, l'état matrimonial, l'union libre et le lien avec la Personne 1, a été amélioré en vue du recensement de 2001 et implanté dans un système appelé SCANCIR (**S**ystème **can**adien de **co**ntrôle et d'**i**mputation du recensement). Il permet maintenant d'effectuer le contrôle et l'imputation d'autres variables telles que l'industrie, le lieu de travail, le mode de transport et la mobilité. Tout comme en 1996, le système SCANCIR permettait de soumettre les réponses à des contrôles plus détaillés et exacts, tout en préservant les réponses grâce à la méthode d'imputation « hot deck à changements minimaux ».

Le système SPIDER (**S**ystem for **P**rocessing **I**nstructions from **D**irectly **E**ntered **R**equirements) a été utilisé pour traiter les autres variables du recensement telles que la langue maternelle, le logement ou le revenu. Le système SPIDER traduit les exigences des agents spécialisés, déterminées au moyen de tables de décision logique, en modules exécutables. SPIDER fait appel aux imputations déterministe et hot deck.

## 2.7 Ajustements de la couverture en fonction des logements inoccupés et des logements non répondants

L'Étude sur la classification des logements (ÉCL) consiste à prendre un échantillon de logements ayant été inscrits comme étant soit inoccupés ou occupés lors de la collecte. Plus tard, l'ÉCL retourne à ces logements pour déterminer s'ils étaient occupés ou non le jour du recensement, ou s'ils n'auraient pas dû être inscrits parce qu'ils ne correspondaient pas à la définition d'un logement aux fins du recensement.

S'il s'avère qu'un logement était occupé, on effectue un de deux ajustements possibles dans la base de données du recensement. Pour les logements inscrits comme étant inoccupés le jour du recensement, on applique une technique d'estimation faisant appel à des **additions aléatoires**, afin d'ajouter des logements et des personnes à la base de données du recensement. Au recensement de 2001, 111 628 ménages et 222 720 personnes ont ainsi été ajoutés à la base de données pour tenir compte du nombre estimatif de personnes vivant dans des logements « inoccupés ». Le deuxième ajustement possible a trait aux logements occupés pour lesquels aucun questionnaire du recensement rempli n'a été reçu (c'est-à-dire les logements non répondants) et consiste à créer, dans la base de données du recensement, une nouvelle taille de ménage pour tous ces logements. Au total, 143 681 ménages comptant 317 587 personnes ont été ajoutés à la base de données par suite de cet ajustement.

## 2.8 Pondération

On a recueilli des renseignements sur l'âge, le sexe, l'état matrimonial, l'union libre, la langue maternelle et le lien avec la Personne 1 auprès de presque toute la population. Cependant, la majeure partie des renseignements proviennent d'un échantillon de 20 % de la population (un ménage sur cinq) à qui on a livré un questionnaire complet (formule 2B) (voir la section 1.1). On a employé la pondération pour reporter, après le contrôle et l'imputation, l'information fournie par l'échantillon à l'ensemble de la population.

La méthode de pondération a fourni des poids qui, à partir des données-échantillon, servaient à produire des estimations. Au recensement de 2001, on a fait appel à une technique que l'on appelle estimation par calibrage ou par régression. Cette procédure consiste à ajuster des poids initiaux égalant environ 5 par le plus petit nombre qui permette de faire correspondre d'aussi près que possible les estimations échantillonnelles (p. ex., nombre d'hommes ou de personnes âgées de 15 à 19 ans) aux chiffres de

population pour l'âge, le sexe, l'état matrimonial, l'union libre et la taille du ménage. 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 » du recensement.

#### 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 1966.

Plusieurs innovations importantes ont été apportées aux méthodes de 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 représentant. Ce changement découlait des résultats de plusieurs études effectuées au Canada et ailleurs (Fellegi, 1964; Hansen et coll., 1959) selon lesquels la présence d'un représentant avait un effet important sur la variance<sup>1</sup> de chiffres du recensement. On prévoyait que l'autodénombrement permettrait de réduire la variance des chiffres du recensement attribuable à la présence du représentant et, en même temps, donnerait au répondant le temps voulu pour remplir le questionnaire à son aise; par conséquent, on s'attendait à ce que les réponses soient plus exactes que par le passé.

Un autre changement survenu dans le 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 de base, soit le nom, le lien de parenté 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

<sup>1</sup> La « variance » d'une estimation est une mesure de sa précision. La variance est expliquée plus en détail au chapitre 9.

du recensement. Tous les ménages vivant dans des régions éloignées prédésignées et dans des logements collectifs<sup>2</sup> 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 de 1971. En outre, le questionnaire de 1976 ne contenait pas les questions qui posent le plus de difficulté de collecte, par exemple les questions sur le revenu, ni celles pour lesquelles les frais de codage sont le plus élevés, par exemple celles sur la profession, l'activité économique 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 et du fardeau de réponse et diminuerait les délais de production (voir Royce, 1983). L'échantillonnage du cinquième des ménages a été maintenu pour les recensements de 1986, 1991, 1996 et 2001.

### **3.2 Plan d'échantillonnage utilisé lors du recensement de 2001**

Le 15 mai 2001, jour du recensement, une multitude de renseignements ont été recueillis auprès de chaque Canadien. La majeure partie de ces informations ont été obtenues à partir d'un échantillon. Dans chaque secteur d'autodénombrement, un échantillon d'un ménage privé occupé sur cinq a été choisi pour recevoir un questionnaire complet (formule 2B) tandis qu'aux ménages 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, la langue maternelle et le lien avec la personne repère du ménage (Personne 1) ont été posées à tous les répondants. Des questions additionnelles portant sur le logement et des questions socioéconomiques ont été posées à un échantillon de la population.

Tous les logements se trouvant dans les secteurs dénombrés par un recenseur (généralement les régions éloignées ou les réserves indiennes) ont reçu le questionnaire 2B. Tous les logements collectifs ont également reçu la formule 2B. Par contre, les personnes suivantes dans les logements collectifs n'ont pas eu à répondre aux questions posées à l'échantillon de la population :

- a) les détenus dans les établissements de correction et les établissements pénitentiaires ou les prisons;
- b) les patients dans les hôpitaux généraux, les établissements de soins spéciaux, les établissements pour les personnes âgées ainsi que les établissements pour personnes souffrant de maladies chroniques et les hôpitaux psychiatriques;
- c) les foyers collectifs pour enfants (orphelinats) et les foyers pour jeunes contrevenants.

D'après les procédures de livraison de base, le recenseur devait planifier un itinéraire couvrant tous les logements de son secteur de dénombrement (SD) puis se rendre à chaque logement et y laisser un questionnaire. Le choix de l'échantillon, c'est-à-dire la décision quant au genre de questionnaire à livrer à chaque logement occupé, a été fait à partir du Registre des visites (RV) dans lequel le recenseur

---

<sup>2</sup> Un logement collectif est une institution ou un établissement commercial ou communautaire, par exemple un hôtel, un hôpital, une résidence de personnel ou un camp de chantier.

inscrivait tous les logements de son secteur. Chaque cinquième ligne du registre était ombrée, indiquant qu'un questionnaire 2B devait être laissé. Les logements correspondant à une ligne non ombrée recevaient un questionnaire abrégé (formule 2A). Un point de départ aléatoire était déterminé à partir de la première ligne du RV, et le recenseur sautait zéro, une, deux, trois ou quatre lignes selon que le cinquième, quatrième, troisième, deuxième ou premier logement du SD devait être le premier à recevoir le questionnaire complet. Par la suite, chaque logement inscrit sur une ligne ombrée recevait automatiquement le questionnaire complet. Ces procédures étaient expliquées dans le Manuel du recenseur et une attention particulière y était portée pendant la formation du recenseur afin de réduire au minimum le risque d'erreur dans la sélection de l'échantillon.

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 (SD). En tant qu'échantillon de personnes, il peut être considéré comme un échantillon stratifié avec sélection systématique de grappes, un logement constituant une grappe. Cochran (1977) ou Sarndal, Swensson et Wretman (1992) donnent une description détaillée des concepts et de la terminologie se rapportant à l'échantillonnage.

## 4. Estimations basées 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 deux sections qui suivent décrivent les considérations opérationnelles et théoriques qui régissent le choix des méthodes d'estimation à partir de l'échantillon du recensement.

### 4.1 Considérations d'ordre opérationnel

Une méthode d'estimation peut être décrite mathématiquement à l'aide d'une formule algébrique qui indique comment la valeur de l'estimateur 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. Dans le cas du recensement, par exemple, chaque cellule de chaque totalisation, fondée sur des données-échantillon recueillies à chaque niveau géographique, représente une estimation-échantillon qui nécessiterait, sous cette approche, une application distincte de la formule d'estimation. 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 d'échantillonnage, c'est-à-dire la pondération elle-même et b) la sommation des poids d'échantillonnage 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 d'échantillonnage au moment où une totalisation est extraite. Il faut noter que, comme le poids d'échantillonnage 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. de la formule 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 la population. Notons cependant qu'il peut arriver que le fait de forcer les estimations-

échantillon de caractéristiques de base à être compatibles avec les chiffres correspondants de la 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 d'échantillonnage 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 la population d'une autre totalisation basée sur les données intégrales. Il se peut que cette comparaison ne donne pas une concordance exacte.

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

Étant donné qu'il faut attribuer un poids d'échantillonnage 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 biais<sup>3</sup>, 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 ménages (à l'exclusion des ménages 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 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'éléments de la population appartenant au sous-groupe (N) et le nombre d'éléments appartenant à l'échantillon (n), et attribuer à chaque unité d'échantillonnage appartenant au sous-groupe un poids d'échantillonnage égal à  $N/n$ . Ces sous-groupes sont souvent appelés **strates a posteriori**.

Par exemple, si l'on recensait 5 000 hommes de 20 à 24 ans à l'Île-du-Prince-Édouard, et si 1 020 de ces hommes faisaient partie des ménages inclus dans l'échantillon, un poids d'échantillonnage de  $5\,000/1\,020 = 4,90$  serait attribué à chaque homme 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 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 d'échantillonnage de 5 donnerait une estimation-échantillon de 5 100 ( $1\,020 \times 5$ ).

<sup>3</sup> « 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.

On appelle **calibrage** l'exercice qui consiste à ajuster les poids d'échantillonnage 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 d'échantillonnage calibré au niveau des ménages, alors que les estimations se rapportant aux personnes l'étaient à partir d'un poids d'échantillonnage calibré au niveau des personnes.

Pour le recensement de 1991, on a adopté la méthode de la régression généralisée (« GREG ») en deux étapes, laquelle a permis d'obtenir une meilleure concordance entre les chiffres de population et les estimations correspondantes au niveau des SD qu'avec la méthode itérative du quotient. En outre, un seul poids d'échantillonnage 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.

Avec la régression généralisée (« GREG ») en deux étapes, les poids d'échantillonnage d'environ 5 ont été ajustés aussi peu que possible pour chaque ménage, 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). Chaque RP contenait en moyenne sept SD échantillonnés. On retrouve plus d'information sur les RP à la section 7.1 du présent rapport.

Au recensement de 1996, chaque SD correspondait à la tâche d'un recenseur. Les SD complets ont été regroupés pour former des RP. Au recensement de 2001, les tâches des recenseurs étaient toujours définies en fonction du SD, mais il arrivait parfois que celui-ci soit agrandi dans les secteurs urbains. En 2001, on a quand même prélevé un échantillon systématique du cinquième des ménages dans chaque SD. Cependant, un nouvel échelon géographique, à savoir l'aire de diffusion (AD), a vu le jour. Les AD devaient ressembler par leur taille aux SD de 1996 et les AD complètes ont été regroupées pour former des RP (d'environ huit AD échantillonnées chacune).

#### 4.4 Estimateur par régression généralisée en deux étapes

En ce qui a trait aux variables portant sur les intervalles d'âge quinquennal, l'état matrimonial, l'union libre, le sexe et la taille du ménage (les 32 variables auxiliaires sont énoncées à l'annexe B), la méthode de pondération du recensement de 2001 visait les objectifs suivants :

- 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 32 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 32 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 d'échantillonnage définitifs doivent se situer dans l'intervalle de 1 à 25 inclusivement. Au recensement de 1996, les poids définitifs pouvaient se situer dans l'intervalle de 0,01 à 25 inclusivement; la **limite** inférieure a été fixée à 1 en 2001 parce qu'on a estimé que chaque personne échantillonnée devait au moins se représenter elle-même.

- e) La méthode de calcul des poids d'échantillonnage doit être hautement automatisée, puisque les 6 141 RP comptant des ménages susceptibles d'être échantillonnés doivent être traitées au cours d'une courte période. Cette méthode doit également garantir un ajustement automatique en fonction des différents schémas de réponse dans les RP du pays.

Les poids d'échantillonnage sont calculés séparément dans chaque RP. Au recensement de 2001, deux ou trois facteurs d'ajustement sont appliqués aux poids d'échantillonnage **initiaux** au niveau des SD (qui correspondent au nombre de ménages privés dans la population divisé par leur nombre dans l'échantillon). D'abord, les ménages sont parfois stratifiés *a posteriori* au niveau des RP en fonction de leur taille, étant donné que les ménages de petite taille et de grande taille sont sous-représentés dans l'échantillon. Les poids d'échantillonnage sont ensuite soumis à un deuxième ajustement visant à 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 de ces trois ajustements avant d'aborder la question du retranchement de contraintes et des diverses raisons pour lesquelles des contraintes peuvent être retranchées.

Dans un premier temps, 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 d'échantillonnage initiaux sont alors multipliés par un facteur de façon à obtenir les poids stratifiés *a posteriori*. Par exemple, selon les poids d'échantillonnage stratifiés *a posteriori*, le nombre estimatif de ménages d'une personne dans une RP donnée correspondrait au nombre de ménages d'une personne dans la population de cette RP. En de très rares occasions, il peut arriver qu'un poids stratifié *a posteriori* soit tronqué 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.

Dans un deuxième temps, un facteur d'ajustement de la pondération par régression à la **première étape** est calculé au niveau des AD. Les 32 variables auxiliaires (âge, sexe, état matrimonial, taille du ménage) qui doivent être appliquées au niveau des RP à la deuxième étape 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 16 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 d'échantillonnage stratifiés *a posteriori* (ou aux poids initiaux en l'absence de stratification *a posteriori*). Les poids d'échantillonnage de la première étape permettent généralement de réduire, mais non d'éliminer, les écarts population/estimation au niveau des AD pour les 32 contraintes.

Enfin, un facteur d'ajustement de la pondération par régression est calculé à la **deuxième étape** au niveau des RP. Les 32 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 d'échantillonnage définitifs. Les poids d'échantillonnage définitifs permettent d'éliminer ou de réduire considérablement les écarts population/estimation au niveau des RP pour les 32 variables auxiliaires.

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

- elles sont **faibles** (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 d'échantillonnage 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 faible (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 faibles, LD et QLD au niveau des RP :

- (i) La taille d'une contrainte est déterminée en fonction du nombre de ménages dans la population auxquels cette contrainte s'applique. Une contrainte dite FAIBLE (le paramètre FAIBLE correspondait à 20, 30 ou 40 ménages en 2001) est retranchée parce que les estimations produites à partir de contraintes faibles ont tendance à être très instables.
- (ii) Les contraintes LD sont ensuite retranchées.
- (iii) 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 2001, mais toujours à 1 000 en 1996), 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 2001, mais toujours à 10 000 en 1996), 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 2001, la valeur du paramètre MAXC était toujours dix fois plus élevée que la valeur du paramètre COND.
- (iv) 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 faibles 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 d'échantillonnage 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 d'échantillonnage aberrants. Afin d'accroître l'efficacité du calcul, on utilise la méthode de fractionnement 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 faibles, linéairement dépendantes ou quasi linéairement dépendantes selon l'analyse initiale de la matrice à inverser. Si un des poids

d'échantillonnage 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 d'échantillonnage du recensement sont calculés indépendamment pour chaque RP, de sorte qu'il est possible d'utiliser des ensembles de paramètres de pondération différents (p. ex. stratification *a posteriori* ou non, FAIBLE, COND, MAXC, intervalle de poids permis). En 1996, un ensemble de paramètres identique était utilisé pour toutes les RP du pays. En 2001, 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 d'échantillonnage de chaque RP au moyen de dix ensembles de paramètres différents. Dans chaque cas, une statistique était calculée afin de déterminer quel ensemble de paramètres réduisait les différences entre les chiffres de population et les estimations-échantillon pour les contraintes. Les poids d'échantillonnage obtenus avec cet ensemble de paramètres étaient utilisés pour la RP correspondante. Afin de conserver certaines contraintes importantes, deux RP ont été pondérées au moyen de paramètres « personnalisés » qui étaient différents des dix autres ensembles. Ce processus de sélection des meilleurs poids d'échantillonnage 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 d'échantillonnage sont établis à l'aide de la régression généralisée (« GREG ») uniquement pour les ménages privés ayant reçu le questionnaire complet du recensement parmi un SD échantillonné (1 sur 5). Un poids d'échantillonnage de 0 est attribué aux ménages privés du SD échantillonné ayant reçu une formule abrégée. On attribue un poids d'échantillonnage de 1 à tous les ménages privés des SD non échantillonnés, car la totalité des répondants de ces secteurs fournissent des renseignements sur la formule 2B. Les ménages collectifs reçoivent aussi un poids d'échantillonnage de 1. **Dans le présent rapport, et à moins d'avis contraire, on emploie le terme « ménage » pour désigner un ménage privé.**

## 4.5 Traitement à deux passages

Aux recensements de 1996 et de 2001, les réponses en lettres à la question sur le lien qui étaient fournies sur les questionnaires abrégés (formules 2A) n'ont pas été saisies en raison de contraintes budgétaires. On leur a plutôt attribué la valeur « Autre ». En revanche, les réponses en lettres fournies sur les questionnaires complets (formules 2B) 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 du premier passage, tous les questionnaires (complets et abrégés) sont traités ensemble, ce qui représente 100 % des données. Les réponses en lettres 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 en lettres fournies sur les questionnaires abrégés. Le processus de contrôle et d'imputation est le même pour les deux types de questionnaire. Lors du second passage, seuls les questionnaires complets sont traités; les questionnaires abrégés ne sont pas utilisés pour l'imputation. On utilise les réponses en lettres à la question sur le lien qui sont saisies sur les questionnaires complets plutôt que les réponses « Autre ». Comme les réponses en lettres sont accessibles, on suppose que la qualité des résultats est plus élevée lors du second 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 d'échantillonnage. 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 en lettres 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

d'échantillonnage du premier passage aux résultats imputés des questionnaires complets du deuxième passage. L'écart entre les chiffres de population (selon les résultats du premier passage) et les estimations du deuxième passage est faible pour la plupart des contraintes. Se reporter au tableau 7.2.2.2 et au graphique 7.2.2.3 à la section 7.2.2, où sont comparés les résultats des premier et deuxième passages.

## 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 évaluation de la variance d'échantillonnage pour diverses caractéristiques-échantillon (20 %).

Chacune de ces études est brièvement décrite ci-après tandis que les résultats obtenus sont présentés aux chapitres 6 à 9.

Trois facteurs servent à expliquer pourquoi les chiffres présentés dans les chapitres suivants ne concordent pas parfaitement avec les chiffres publiés. Dans un premier temps, seuls les ménages sujets à l'échantillonnage font partie de ces études. Dans un deuxième temps, on a plutôt utilisé les données du premier passage que celles du deuxième passage (voir la section 4.5). Enfin, cette situation s'explique du fait qu'aucune correction n'a été faite pour les « additions aléatoires » (voir la section 2.7).

### 5.1 Biais d'échantillonnage

L'objectif de cette étude était de relever les caractéristiques qui affichaient des écarts importants entre les estimations fondées sur les poids d'échantillonnage initiaux et les chiffres de population connus. Ces écarts présentent de l'intérêt pour deux raisons. D'une part, ils peuvent être utiles pour identifier les biais dans l'échantillon de ménages sélectionnés sur le terrain, et d'autre part, ils peuvent montrer l'incidence de la non-réponse sur les questions-échantillon du recensement (les questionnaires complets sur lesquels les questions-échantillon sont sans réponse sont convertis en questionnaires abrégés au cours du dépouillement). Ces biais dans les questionnaires abrégés préviennent les biais possibles dans les estimations tirées des questionnaires complets. Les biais dans les caractéristiques des questionnaires abrégés sont corrigés par calibrage. Si les caractéristiques des questionnaires complets sont corrélés avec celles des questionnaires abrégés, le calibrage devrait aussi réduire leur biais.

### 5.2 Évaluation des méthodes de pondération

Le but de cette étude était d'évaluer l'efficacité de l'estimateur par régression généralisée. À cette fin, on a examiné le niveau de concordance entre les estimations-échantillon et les chiffres de population à l'égard de l'ensemble des contraintes au niveau des RP, pour le Canada. 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 d'échantillonnage 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 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 d'échantillonnage 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 pour le plan d'échantillonnage et les 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 valeur Z

$$Z^{(0)} = \frac{\hat{X}^{(0)} - X}{\sqrt{V(\hat{X}^{(0)})}}$$

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 d'échantillonnage initiaux. Pour calculer la valeur 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, on peut voir que la valeur  $Z^{(0)}$  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 1996 et de 2001 (ainsi que les différences  $\hat{X}^{(0)} - X$ ) pour 32 caractéristiques très semblables aux contraintes ayant été appliquées pour le calcul des poids d'échantillonnage définitifs (voir l'annexe B). Si la valeur  $Z^{(0)}$  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,  $|Z^{(0)}| > 3$  est de  $0,0026 \times 32 = 0,0832$  pour les 32 caractéristiques présentées au tableau 6.1. Cependant, pour le recensement de 2001, 25 des 32 caractéristiques 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 2001 est biaisé. Les valeurs Z associées à la population totale, aux femmes, aux femmes de 15 ans et plus, aux personnes de 5 à 14 ans, aux personnes de 55 ans et plus, aux personnes mariées, aux ménages de deux personnes et aux ménages de quatre personnes sont positives et élevées, ce qui signifie que ces caractéristiques sont surreprésentées dans l'échantillon. En revanche, les valeurs Z associées aux hommes de 15 ans et plus, aux personnes de 20 à 34 ans, aux célibataires, aux personnes séparées et aux ménages d'une personne sont négatives et élevées, ce qui révèle que ces caractéristiques sont sous-représentées dans l'échantillon. Le tableau 6.1 et le graphique 6.1 montrent également que la valeur absolue de Z est souvent beaucoup plus élevée en 2001 qu'en 1996.

Le biais peut provenir de plusieurs sources : erreurs commises par les recenseurs (p. ex., s'ils ne choisissent pas l'échantillon selon les exigences établies), biais de non-réponse (p. ex., les jeunes hommes sont moins susceptibles de remplir un questionnaire complet qu'un questionnaire abrégé), biais de réponse (p. ex., si les recensés ne répondent pas de la même façon sur la formule 2B et la formule 2A), erreurs de traitement, et ainsi de suite. En ce qui touche le biais de non-réponse, 1,3 % des ménages (inclus ou non dans l'échantillon) n'ont pas rempli le questionnaire de 2001 (parce qu'ils ont refusé de le faire ou qu'on n'a pu les joindre), comparativement à 0,8 % en 1996. Ces ménages sont appelés « ménages oubliés ou ayant opposé un refus ». De plus, 0,7 % des ménages de l'échantillon en 2001 ont fourni certaines réponses aux questions de base, mais n'ont fourni aucune réponse aux questions-échantillon. En 1996, le pourcentage correspondant était de 0,2 %. Au cours du traitement des données, les ménages de l'échantillon qui ont carrément refusé de répondre, soit à toutes les questions, soit aux questions-échantillons seulement, ont été convertis de ménages 2B en ménages 2A, lesquels ne font pas partie de l'échantillon. Par conséquent, seules les réponses aux questions de base ont été imputées au besoin. Ce processus de conversion des ménages échantillonnés en ménages non échantillonnés est appelé conversion de document 2A/2B. Il est possible que les ménages oubliés ou ayant opposé un refus, ainsi que les ménages n'ayant pas fourni de réponses aux questions-échantillon aient des caractéristiques différentes de celles des autres ménages. Ainsi, la conversion de ménages 2B en ménages 2A pourrait biaiser l'échantillon. Par exemple, on sait que le pourcentage de ménages occupant une maison individuelle non attenante qui ont été oubliés ou ont opposé un refus est moitié moins grand que le pourcentage de ménages oubliés ou ayant opposé un refus pour l'ensemble de la population.

Le graphique 6.1 révèle que la valeur Z est plus élevée en 2001 qu'en 1996 pour de nombreuses caractéristiques. La valeur Z étant une variable aléatoire, il est possible que ces différences ne soient pas toutes statistiquement significatives. Les 12 caractéristiques dont la valeur Z affiche une différence statistiquement significative sont signalées par un astérisque (voir le graphique 6.1). Elles ont été identifiées au moyen d'une valeur W, qui est définie à l'annexe C.

La variation géographique du biais a aussi été étudiée. Les valeurs Z correspondant aux 32 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) comme pour l'ensemble du Canada. Le biais observable pour ces quatre régions est présenté pour les recensements de 2001 et de 1996 dans les graphiques 6.2 et 6.3 respectivement. Là encore, on s'est servi de la valeur W pour relever les différences statistiquement significatives entre les régions, qui sont indiquées par la présence des initiales des régions visées au bas du graphique. Par exemple, la mention « QOn QOu » indique qu'il y a une différence significative entre le Québec et l'Ontario et entre le Québec et la région de l'Ouest.

Le graphique 6.2 montre que les seules régions à afficher une variation du biais en 2001 sont le Québec par rapport à l'Ontario et le Québec par rapport à la région de l'Ouest. Il est intéressant de noter que cette observation tient pour sept caractéristiques. La plupart des caractéristiques se rapportant à l'âge sont représentées de la même façon dans toutes les régions, mais la différence qui ressort le plus est une surreprésentation au Québec en comparaison avec une sous-représentation en Ontario et dans la région de l'Ouest des personnes âgées de 15 à 19 ans. D'autres différences ressortent entre les régions lorsqu'on compare les caractéristiques non liées à l'âge, la plupart se manifestant à l'égard des caractéristiques de la population. Quant aux caractéristiques se rapportant aux ménages, elles tendent à concorder d'une région à l'autre, exception faite des ménages de trois personnes, qui présentent une variation entre le Québec et l'Ontario.

Si l'on compare les biais affichés par les régions au recensement de 2001 avec ceux du recensement de 1996 (voir les graphiques 6.2 et 6.3), on remarque le maintien de certaines tendances (p. ex., dans le cas des hommes, des hommes de 15 ans et plus, des femmes de 15 ans et plus, des célibataires, des personnes mariées).

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 d'échantillonnage. Par conséquent, les inférences fondées sur des estimations calibrées devraient être plus justes.

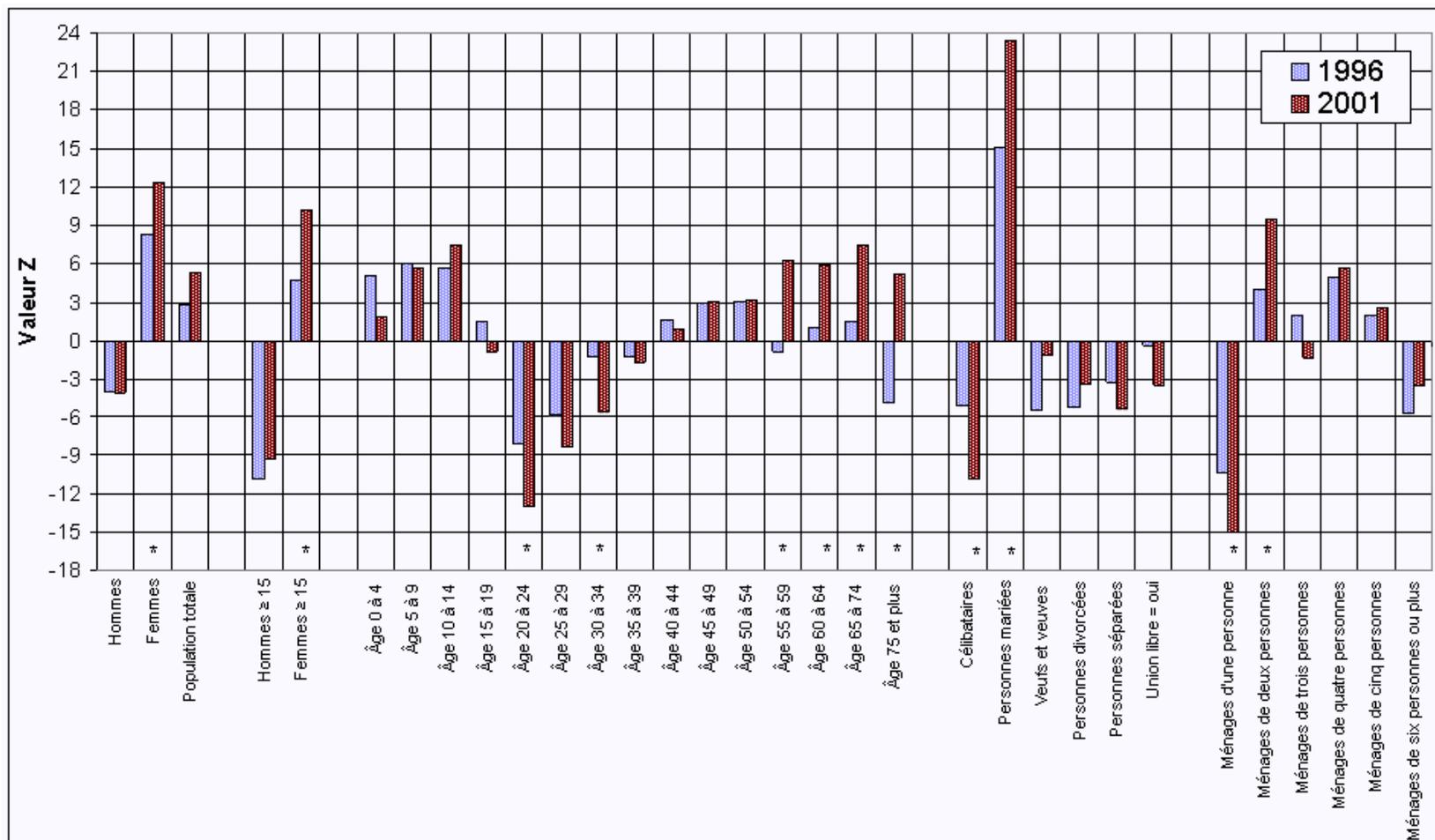
**Tableau 6.1 : Différences population/estimation fondées sur les poids d'échantillonnage initiaux, recensements de 2001 et de 1996**

Caractéristique	Recensement de 2001						Recensement de 1996					
	Chiffre	Estimation <sup>1</sup>	Différence <sup>2</sup>	Écart <sup>3</sup>	E.T. <sup>4</sup>	Valeur Z <sup>5</sup>	Chiffre	Estimation <sup>1</sup>	Différence <sup>2</sup>	Écart <sup>3</sup>	E.T. <sup>4</sup>	Valeur Z <sup>5</sup>
Hommes	14 171 941	14 146 867	-25 074	-0,18	6 139	-4,08	13 717 654	13 694 786	-22 868	-0,17	5 752	-3,98
Femmes	14 699 518	14 772 915	73 397	0,50	5 940	12,36	14 176 680	14 222 665	45 985	0,32	5 552	8,28
Population totale	28 871 459	28 919 783	48 324	0,17	8 991	5,37	27 894 334	27 917 451	23 117	0,08	8 227	2,81
Hommes ≥ 15 ans	11 340 286	11 295 995	-44 291	-0,39	4 747	-9,33	10 781 073	10 732 804	-48 269	-0,45	4 449	-10,85
Femmes ≥ 15 ans	11 998 509	12 042 929	44 420	0,37	4 342	10,23	11 383 130	11 402 113	18 983	0,17	4 006	4,74
Âge 0 à 4	1 636 092	1 641 720	5 628	0,34	2 986	1,88	1 858 332	1 874 111	15 779	0,85	3 073	5,14
Âge 5 à 9	1 910 359	1 928 604	18 245	0,96	3 213	5,68	1 932 023	1 950 728	18 705	0,97	3 120	6,00
Âge 10 à 14	1 986 213	2 010 534	24 321	1,22	3 271	7,44	1 939 776	1 957 694	17 918	0,92	3 125	5,73
Âge 15 à 19	1 986 163	1 983 519	-2 644	-0,13	3 245	-0,81	1 903 023	1 907 732	4 709	0,25	3 074	1,53
Âge 20 à 24	1 892 572	1 851 491	-41 081	-2,17	3 168	-12,97	1 840 654	1 816 301	-24 353	-1,32	3 013	-8,08
Âge 25 à 29	1 835 744	1 810 124	-25 620	-1,40	3 077	-8,33	1 971 123	1 953 292	-17 831	-0,90	3 053	-5,84
Âge 30 à 34	2 031 513	2 013 625	-17 888	-0,88	3 173	-5,64	2 405 559	2 401 580	-3 979	-0,17	3 317	-1,20
Âge 35 à 39	2 452 299	2 446 624	-5 675	-0,23	3 427	-1,66	2 486 060	2 482 136	-3 924	-0,16	3 339	-1,18
Âge 40 à 44	2 510 847	2 513 920	3 073	0,12	3 439	0,89	2 268 423	2 273 674	5 251	0,23	3 177	1,65
Âge 45 à 49	2 273 676	2 283 700	10 024	0,44	3 286	3,05	2 050 229	2 059 233	9 004	0,44	3 040	2,96
Âge 50 à 54	2 031 050	2 041 054	10 004	0,49	3 137	3,19	1 581 484	1 589 751	8 267	0,52	2 707	3,05
Âge 55 à 59	1 549 675	1 567 071	17 396	1,12	2 758	6,31	1 271 221	1 269 086	-2 135	-0,17	2 448	-0,87
Âge 60 à 64	1 234 930	1 249 389	14 459	1,17	2 469	5,86	1 157 926	1 160 459	2 533	0,22	2 338	1,08
Âge 65 à 74	2 059 079	2 083 362	24 283	1,18	3 256	7,46	1 991 721	1 996 303	4 582	0,23	3 068	1,49
Âge 75 et plus	1 481 247	1 495 045	13 798	0,93	2 676	5,16	1 236 780	1 225 372	-11 408	-0,92	2 332	-4,89
Célibataires	13 282 845	13 196 174	-86 671	-0,65	8 018	-10,81	12 779 218	12 741 878	-37 340	-0,29	7 320	-5,10
Personnes mariées	11 750 092	11 906 204	156 112	1,33	6 678	23,38	11 537 475	11 628 813	91 338	0,79	6 076	15,03
Veufs et veuves	1 341 497	1 339 109	-2 388	-0,18	2 254	-1,06	1 303 304	1 291 501	-11 803	-0,91	2 130	-5,54

Caractéristique	Recensement de 2001						Recensement de 1996					
	Chiffre	Estimation <sup>1</sup>	Différence <sup>2</sup>	Écart <sup>3</sup>	E.T. <sup>4</sup>	Valeur Z <sup>5</sup>	Chiffre	Estimation <sup>1</sup>	Différence <sup>2</sup>	Écart <sup>3</sup>	E.T. <sup>4</sup>	Valeur Z <sup>5</sup>
Personnes divorcées	1 794 079	1 784 704	-9 375	-0,52	2 824	-3,32	1 605 136	1 591 530	-13 606	-0,85	2 612	-5,21
Personnes séparées	702 946	693 591	-9 355	-1,33	1 749	-5,35	669 201	663 729	-5 472	-0,82	1 675	-3,27
Union libre = oui	2 267 634	2 253 253	-14 381	-0,63	4 090	-3,52	1 770 338	1 768 774	-1 564	-0,09	3 568	-0,44
Ménages d'une personne	2 908 857	2 866 182	-42 675	-1,47	2 847	-14,99	2 584 348	2 558 041	-26 307	-1,02	2 524	-10,42
Ménages de deux personnes	3 709 282	3 739 781	30 499	0,82	3 224	9,46	3 385 597	3 397 657	12 060	0,36	3 011	4,00
Ménages de trois personnes	1 848 476	1 845 071	-3 405	-0,18	2 541	-1,34	1 804 304	1 809 076	4 772	0,26	2 435	1,96
Ménages de quatre personnes	1 812 783	1 826 921	14 138	0,78	2 481	5,70	1 813 493	1 825 159	11 666	0,64	2 378	4,91
Ménages de cinq personnes	714 618	719 013	4 395	0,61	1 664	2,64	737 751	740 921	3 170	0,43	1 640	1,93
Ménages de six personnes ou plus	332 959	328 968	-3 991	-1,20	1 155	-3,46	334 207	327 786	-6 421	-1,92	1 124	-5,71

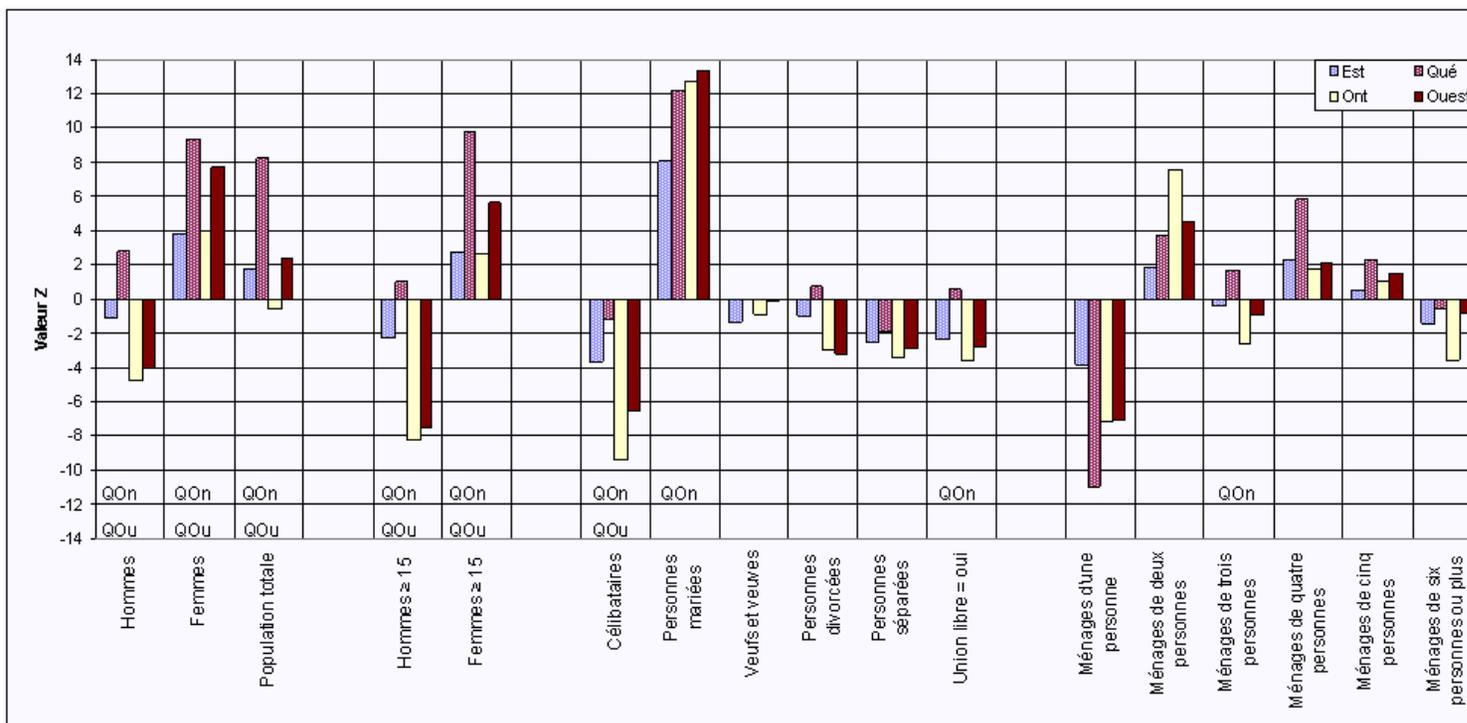
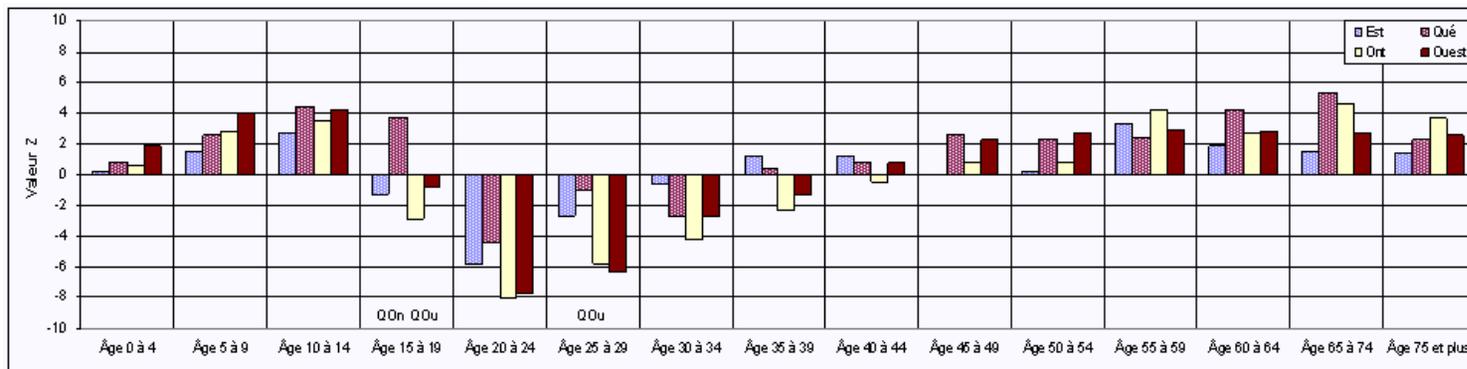
- 1 Fondée sur les poids d'échantillonnage initiaux
- 2 Différence : estimation-chiffre
- 3 Écart : écart (100\* [estimation-chiffre]/chiffre)
- 4 E.T. : erreur-type de l'estimation fondée sur le poids d'échantillonnage initial
- 5 Valeur Z : (estimation-chiffre)/E.T.

**Graphique 6.1 :** Valeurs Z des différences population/estimation fondées sur les poids d'échantillonnage initiaux, Canada, recensements de 2001 et de 1996

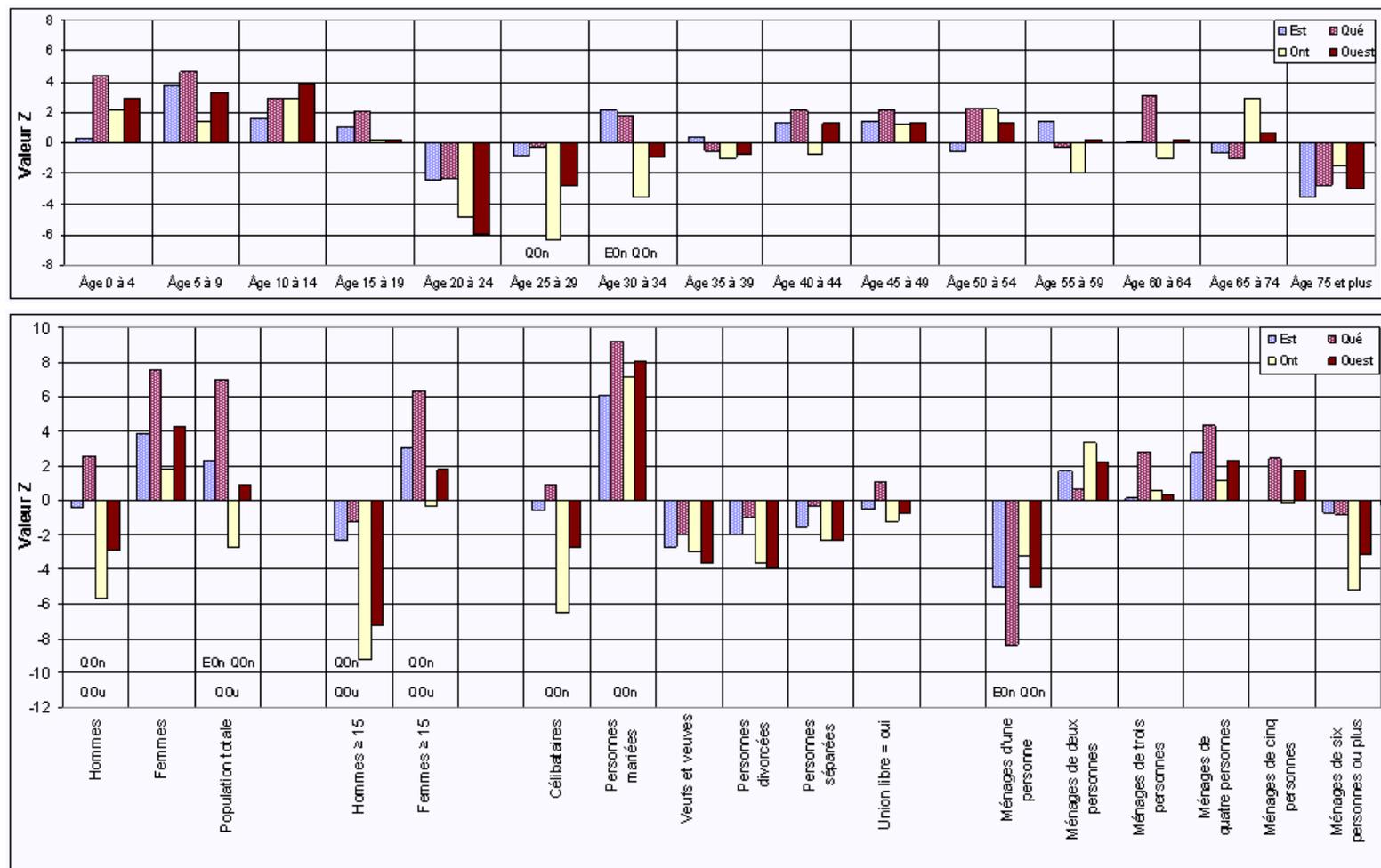


\* indique une différence significative entre les biais des deux recensements

Graphique 6.2 : Valeurs Z selon les régions en 2001



Graphique 6.3 : Valeurs Z selon les régions en 1996



## 7. Évaluation des méthodes de pondération

Le présent chapitre a pour but de présenter, puis d'évaluer, certains éléments du processus de pondération du recensement tels 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. 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 2001, le pays était divisé en 6 148 RP, chacune comptant en moyenne huit 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), aux secteurs de recensement (SR) et, enfin, aux circonscriptions électorales fédérales (CÉF).
- 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 ci-dessous montre que 5 784 RP, ou 94,2 % de l'ensemble des RP, se situent à l'intérieur de l'intervalle des 1 000 à 3 000 ménages au recensement de 2001. Un pourcentage légèrement plus élevé de RP se situait à l'intérieur de cet intervalle en 1996. Le nombre moyen de logements par RP était de 2 047. Plusieurs RP ont affiché un compte supérieur à la moyenne dont le plus élevé s'établit à 17 043 logements. Il convient de noter que les sept RP qui ont enregistré une population de zéro personnes au recensement de 2001 ne sont pas représentées au tableau 7.1.1. Y sont également exclues les RP dont toutes les AD n'ont pas été sujettes à l'échantillonnage, y compris, par exemple, la totalité des RP des Territoires du Nord-Ouest et du Nunavut.

On obtiendra une concordance entre les estimations-échantillon et les chiffres de population uniquement pour les régions géographiques formées exclusivement de RP complètes. Le tableau 7.1.2 permet d'examiner le lien entre les limites des SDR et des SR du recensement de 2001 et les limites des RP. Pour une SDR donnée, par exemple, la catégorie « Régions géographiques formées seulement d'une partie de RP alors que le reste de la RP renferme uniquement des régions géographiques complètes du même genre » indique que la SDR est comprise entièrement dans une RP (c.-à-d. qu'elle ne chevauche pas deux RP) et que la RP renferme uniquement des SDR complètes. Ces SDR peuvent représenter un village ou une petite ville. La catégorie « Régions géographiques formées seulement d'une partie de RP alors que le reste de la RP ne renferme pas uniquement des régions géographiques complètes du même genre » est similaire à la catégorie précédente, sauf que la RP ne renferme pas uniquement des SDR complètes (c.-à-d. qu'au moins une SDR dans la RP chevauche deux RP ou plus). Une SDR appartenant à la catégorie « Régions géographiques renfermant au moins une RP complète » est une SDR (souvent une ville de taille moyenne ou une grande ville) qui comprend au moins une RP complète et dont chaque RP renferme une seule SDR ou une partie d'une seule SDR. Si la SDR est incluse dans la catégorie « Régions géographiques chevauchant au moins une limite de RP », elle chevauche deux RP ou plus.

Les quatre catégories de régions présentées ici s'excluent mutuellement et couvrent toutes les régions. Ces définitions s'appliquent aussi aux SR.

D'après les chiffres présentés au tableau 7.1.2, on constate que 12,8 % des SDR et 65,4 % 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.

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 2001		Recensement de 1996	
	N <sup>bre</sup> de RP	Pourcentage	N <sup>bre</sup> de RP	Pourcentage
1 - 999	1	0,0	4	0,1
1 000 - 1 499	1 132	18,4	1 686	28,4
1 500 - 1 999	2 248	36,6	2 213	37,2
2 000 - 2 499	1 622	26,4	1 417	23,9
2 500 - 3 000	786	12,8	560	9,4
3 001+	352	5,8	61	1,0
Total	6 141	100,0	5 941	100,0

**Tableau 7.1.2 : Nombre de SDR et de SR qui respectent les limites des RP au recensement de 2001**

Description	SDR		SR	
	Nombre	%	Nombre	%
Régions géographiques formées seulement d'une partie de RP alors que le reste de la RP renferme uniquement des régions géographiques complètes du même genre	4 165	74,4	1 563	30,8
Régions géographiques formées seulement d'une partie de RP alors que le reste de la RP ne renferme pas uniquement des régions géographiques complètes du même genre	567	10,1	106	2,1
Régions géographiques renfermant au moins une RP complète	717	12,8	3 313	65,4
Régions géographiques chevauchant au moins une limite de RP	151	2,7	87	1,7
<b>Total</b>	<b>5 600</b>		<b>5 069</b>	

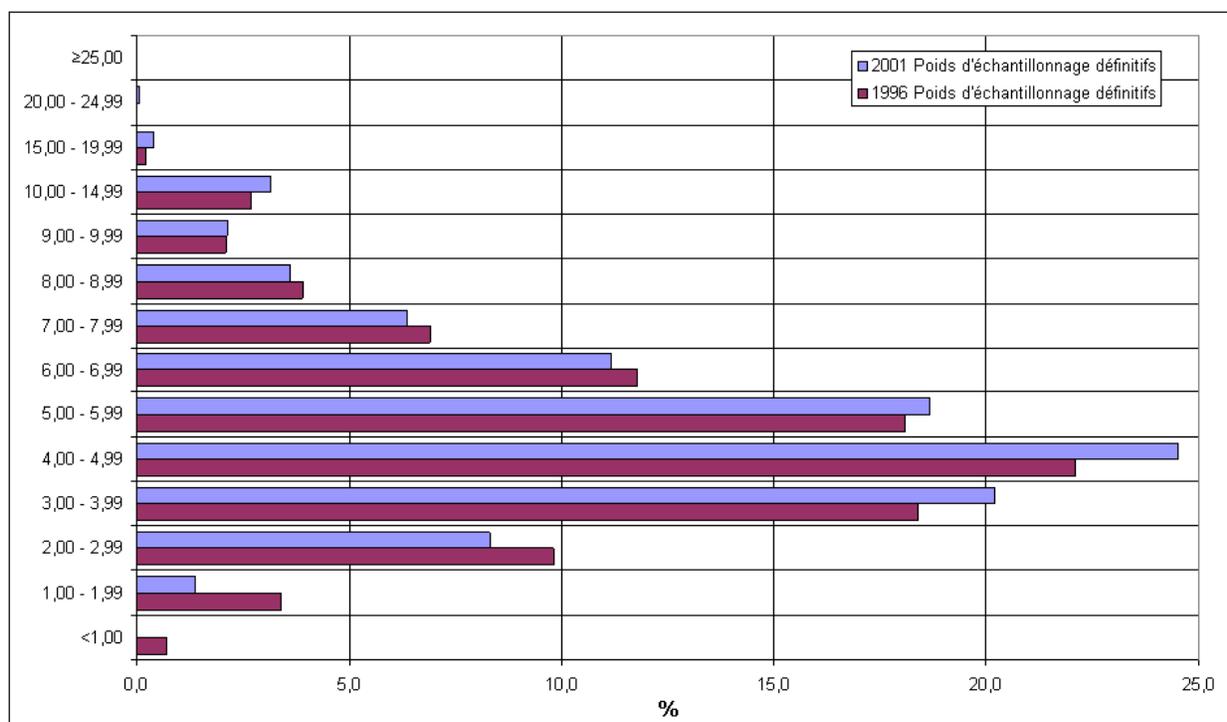
## 7.2 Évaluation de la méthode de pondération du recensement

### 7.2.1 Répartition des poids d'échantillonnage

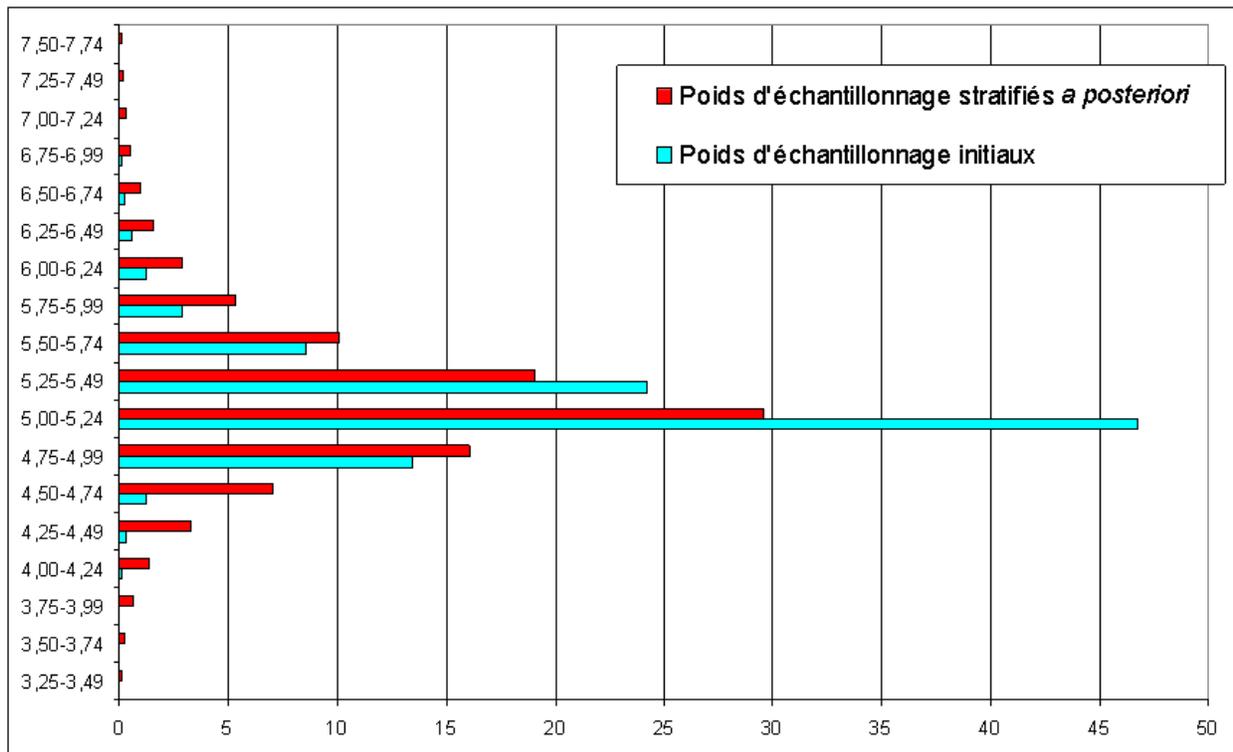
Le graphique 7.2.1.1 permet de comparer la répartition des poids d'échantillonnage définitifs du recensement de 2001 avec celle de 1996. Bien que les répartitions soient très semblables, il faut noter que les poids d'échantillonnage ne pouvaient pas être inférieurs à 1 en 2001. Si l'on compare les données des deux recensements, on constate que le pourcentage des ménages auxquels correspondent des petits poids d'échantillonnage (inférieurs à 2,99, voire à 1,00 dans 0,7 % des cas) était plus élevé pour 1996 que pour 2001, alors qu'en 2001, un pourcentage plus grand de ménages qu'en 1996 avait des poids d'échantillonnage entre 3,00 et 5,99. Les répartitions des poids d'échantillonnage supérieurs à 6,00 ne présentent que de légers écarts.

Les graphiques 7.2.1.2 à 7.2.1.4 permettent de comparer les distributions des poids d'échantillonnage initiaux, stratifiés *a posteriori*, obtenus à la première étape et définitifs du recensement de 2001. Les poids initiaux sont 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 définitifs deviennent progressivement plus étendues, à mesure que les contraintes deviennent plus restrictives.

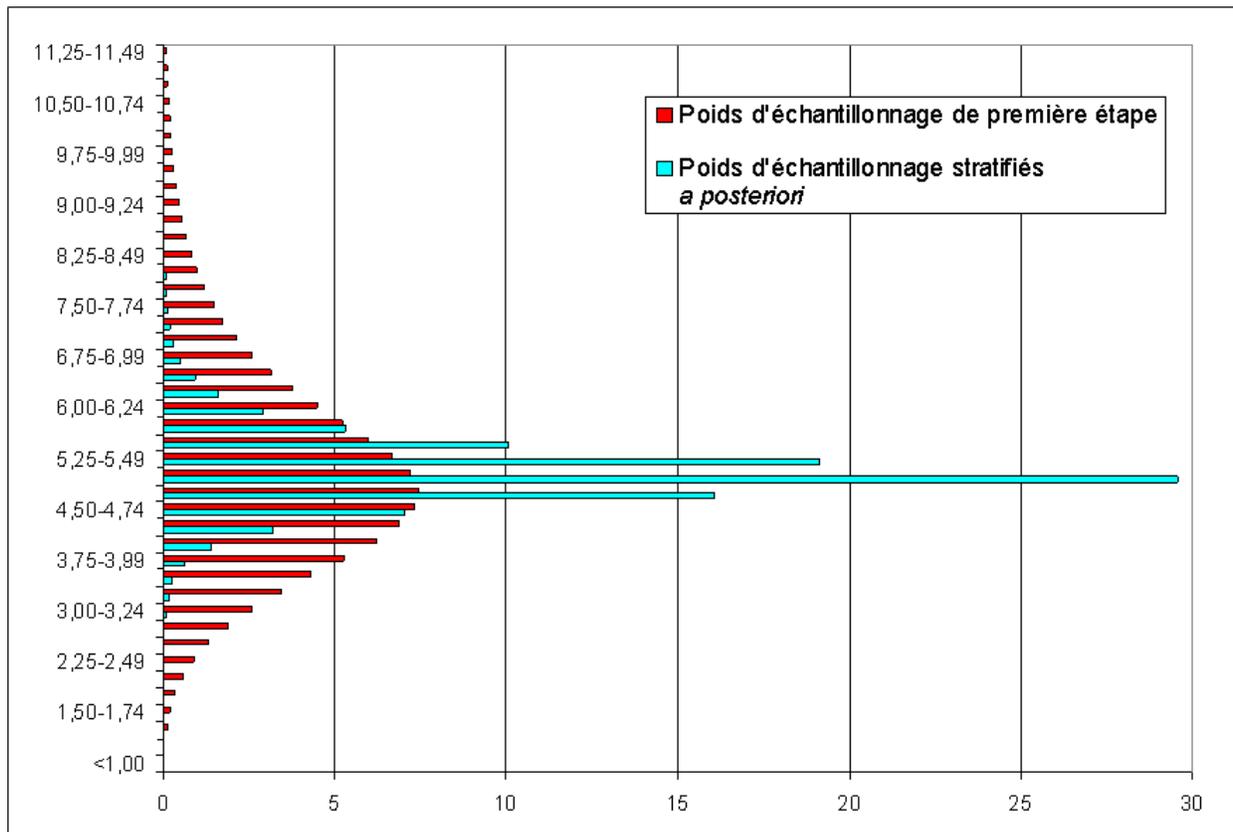
**Graphique 7.2.1.1 : Comparaison des poids d'échantillonnage définitifs associés aux ménages des recensements de 2001 et de 1996**



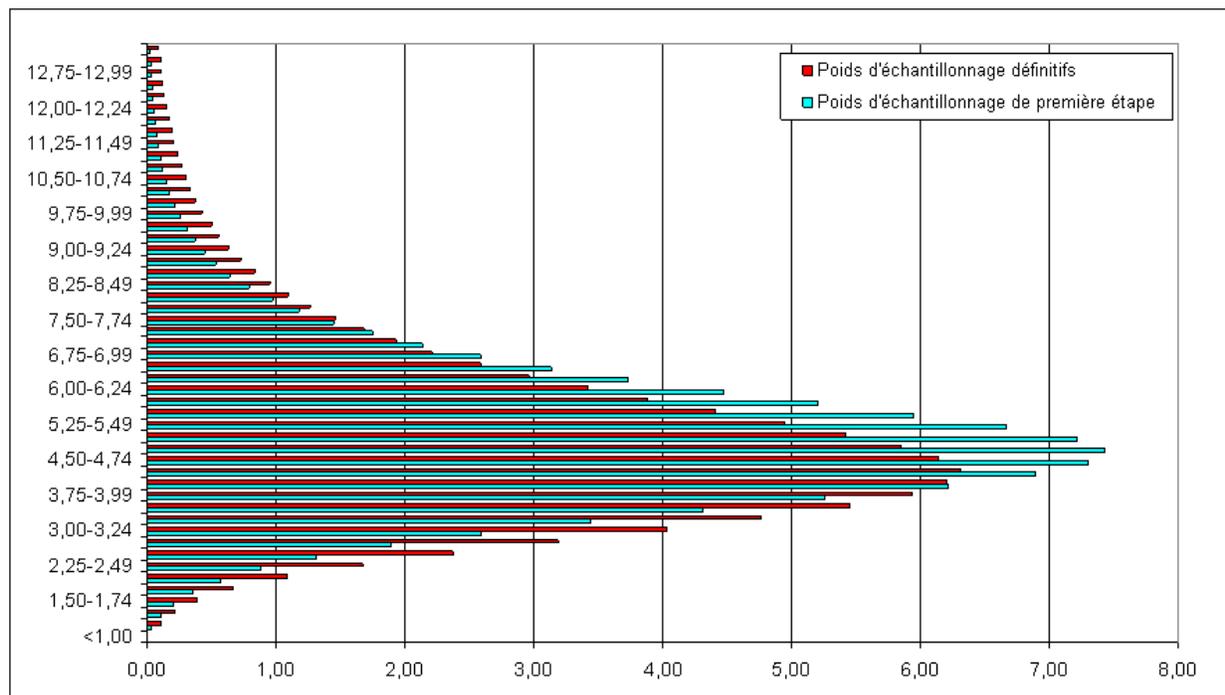
**Graphique 7.2.1.2 : Comparaison des poids d'échantillonnage initiaux et des poids stratifiés *a posteriori*, recensement de 2001**



**Graphique 7.2.1.3 : Comparaison des poids d'échantillonnage stratifiés *a posteriori* et des poids obtenus à la première étape, recensement de 2001**



**Graphique 7.2.1.4 : Comparaison des poids d'échantillonnage obtenus à la première étape et des poids d'échantillonnage définitifs, recensement de 2001**



## 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 d'échantillonnage définitifs sont choisis de manière à réduire, ou à éliminer, les écarts entre les chiffres de population et les estimations-échantillon correspondantes pour 32 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** ». Les échantillons-estimation et les chiffres de population sont fondés sur les logements occupés dans les SD échantillonnés.

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 2001 et en 1996 pour les 32 contraintes au niveau des RP ainsi que pour les poids d'échantillonnage initiaux et définitifs. Le graphique 7.2.2.1 est semblable au graphique 6.1 sauf qu'au lieu de présenter des statistiques Z fondées sur les poids d'échantillonnage initiaux, il en présente des écarts population/estimation. En raison de ces similitudes, et étant donné que le chapitre 6 renferme d'autres précisions à cet égard, le graphique 7.2.2.1 ne sera pas examiné en détail. De façon générale, la tendance qui en ressort, c'est que les écarts enregistrés en 2001 sont généralement beaucoup plus importants qu'en 1996. Le tableau 7.2.2.1 montre que les valeurs absolues des écarts population/estimation d'après les poids d'échantillonnage définitifs en 2001 sont généralement plus petites qu'en 1996 pour les intervalles d'âge quinquennal et pour la plupart des

réponses associées à l'état matrimonial. En revanche, l'inverse est généralement vrai pour la caractéristique « Union libre = oui » et certaines tailles de ménage. Comme on le verra à la section 7.2.3, les variations observées dans les écarts d'un recensement à l'autre sont le plus souvent attribuables au fait que le nombre de contraintes retranchées a changé. Lorsqu'on compare les graphiques 7.2.2.1 et 7.2.2.2, il est clair que pour 2001 les écarts population/estimation fondés sur les poids d'échantillonnage définitifs sont nettement plus faibles que ceux fondés sur les poids d'échantillonnage initiaux, sauf dans le cas des ménages de cinq personnes. Comme on le verra à la section 7.2.3, cela s'explique sans doute par la fréquence de retranchement élevée de cette contrainte parce qu'elle causait des poids d'échantillonnage aberrants et, dans une moindre mesure, parce qu'elle était quasi linéairement dépendante.

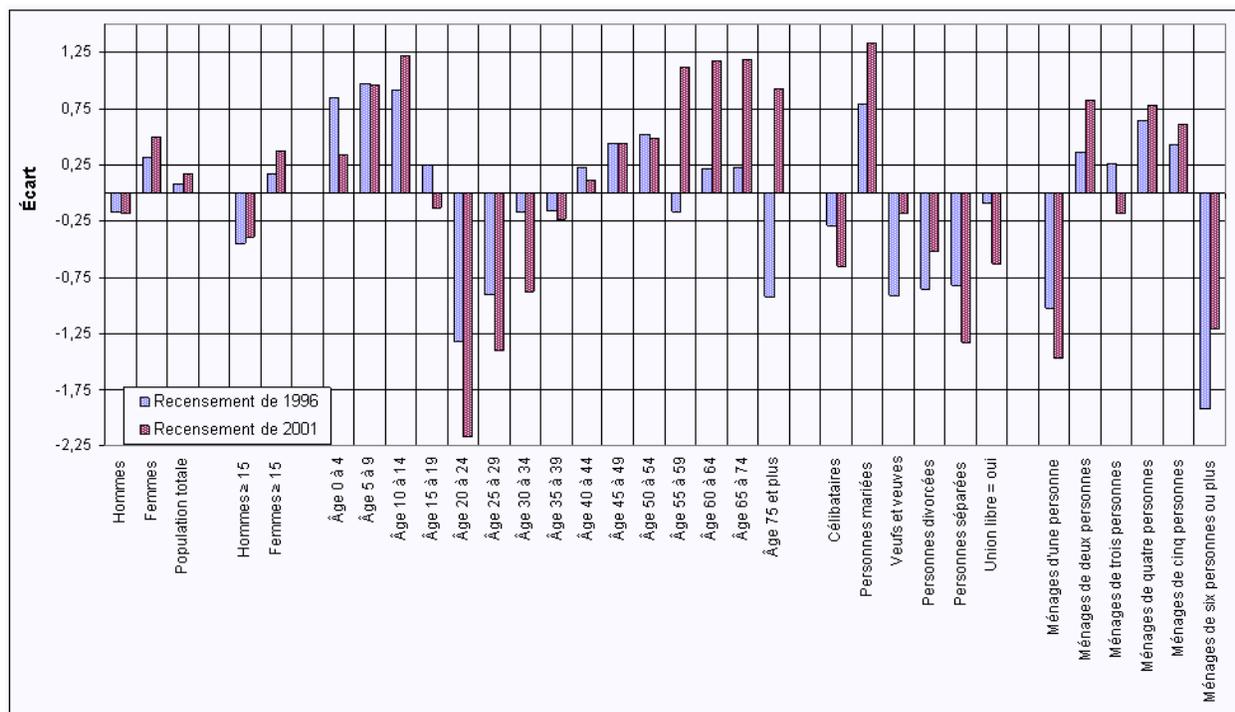
Le tableau 7.2.2.2 et le graphique 7.2.2.3 présentent les résultats des premier et deuxième passages quant aux différences et aux écarts population/estimation fondés sur les poids d'échantillonnage définitifs pour les 32 contraintes au niveau des RP, pour le Canada. Il en ressort que les écarts du premier passage sont plus faibles, ce qui s'explique par le fait que les poids d'échantillonnage du recensement ont été calculés à partir des résultats du premier passage. Se reporter à la section 4.5 pour obtenir de plus amples renseignements sur le traitement à deux passages.

**Tableau 7.2.2.1 : Comparaison des écarts population/estimation des recensements de 1996 et de 2001, pour le Canada**

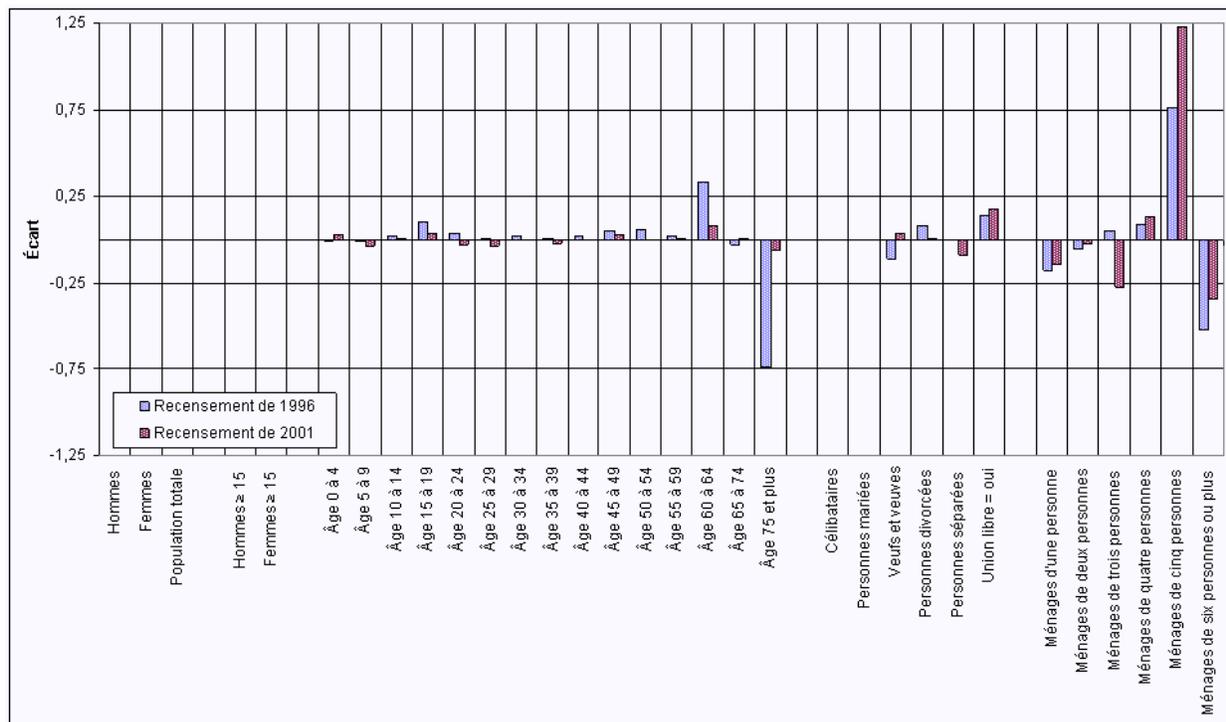
Caractéristique	Recensement de 2001			Recensement de 1996		
	Poids initiaux	Poids définitifs		Poids initiaux	Poids définitifs	
	Différence	Différence	Écart	Différence	Différence	Écart
Hommes	-25 074	-	0,00	-22 868	15	0,00
Hommes ≥ 15 ans	-44 291	51	0,00	-48 269	-276	0,00
Personnes ≥ 15 ans	130	-	0,00	-29 285	3	0,00
Total des ménages	-1 040	-	0,00	1 060	-	0,00
Population totale	48 324	-	0,00	23 117	-	0,00
Âge 0 à 4	5 628	559	0,03	15 779	-208	-0,01
Âge 5 à 9	18 245	-792	-0,04	18 705	-258	-0,01
Âge 10 à 14	24 321	234	0,01	17 918	462	0,02
Âge 15 à 19	-2 644	779	0,04	4 709	1 853	0,10
Âge 20 à 24	-41 081	-504	-0,03	-24 353	803	0,04
Âge 25 à 29	-25 620	-785	-0,04	-17 381	105	0,01
Âge 30 à 34	-17 888	7	0,00	-3 979	361	0,02
Âge 35 à 39	-5 675	-556	-0,02	-3 924	320	0,01
Âge 40 à 44	3 073	100	0,00	5 251	366	0,02
Âge 45 à 49	10 024	687	0,03	9 004	971	0,05
Âge 50 à 54	10 004	-87	0,00	8 267	993	0,06
Âge 55 à 59	17 396	81	0,01	-2 135	254	0,02
Âge 60 à 64	14 459	933	0,08	2 533	3 847	0,33
Âge 65 à 74	24 283	271	0,01	4 582	-662	-0,03
Âge 75 et plus	13 798	-926	-0,06	-11 408	-9 207	-0,74

Caractéristique	Recensement de 2001			Recensement de 1996		
	Poids initiaux	Poids définitifs	Écart	Poids initiaux	Poids définitifs	Écart
	Différence	Différence		Différence	Différence	
Célibataires	-86 671	-53	0,00	-37 340	115	0,00
Personnes mariées	156 112	-57	0,00	91 338	73	0,00
Veufs et veuves	-2 388	557	0,04	-11 803	-1 387	-0,11
Personnes divorcées	-9 375	206	0,01	-13 606	1 209	0,08
Personnes séparées	-9 355	-653	-0,09	-5 472	-10	0,00
Union libre = oui	-14 381	4 115	0,18	-1 404	2 415	0,14
Ménages d'une personne	-42 675	-4 175	-0,14	-	-4 750	-0,18
Ménages de deux personnes	30 499	-906	-0,02	12 060	-1 666	-0,05
Ménages de trois personnes	-3 405	-5 010	-0,27	4 772	871	0,05
Ménages de quatre personnes	14 138	2 414	0,13	11 666	1 694	0,09
Ménages de cinq personnes	4 395	8 818	1,23	3 170	5 576	0,76
Ménages de six personnes ou plus	-3 991	-1 142	-0,34	-	-1 725	-0,52

Graphique 7.2.2.1 : Écarts population/estimation fondés sur les poids d'échantillonnage initiaux, recensements de 1996 et de 2001



**Graphique 7.2.2.2 : Écarts population/estimation fondés sur les poids d'échantillonnage définitifs, recensements de 1996 et de 2001**



**Tableau 7.2.2.2 : Écarts population/estimation fondés sur les poids d'échantillonnage définitifs pour le Canada, comparaison entre le premier passage et le deuxième passage, recensement de 2001**

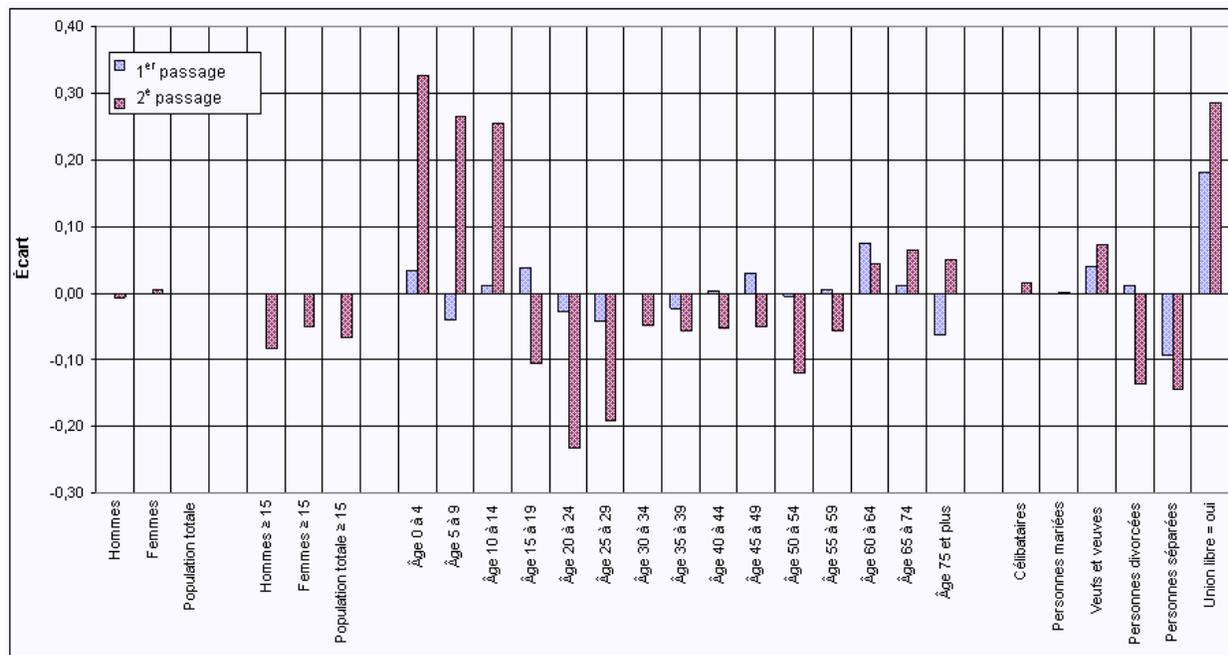
Caractéristique	Recensement de 2001 – 1 <sup>er</sup> passage				Recensement de 2001 – 2 <sup>e</sup> passage				passage 2 – passage 1	
	Chiffre	Estimation	Différence	Écart	Chiffre	Estimation	Différence	Écart	Différence	Écart
Hommes	14 171 941	14 171 941	0	0,00	14 393 344	14 392 459	-885	-0,01	-885	-0,01
Femmes	14 699 518	14 699 518	0	0,00	14 911 511	14 912 396	885	0,01	885	0,01
Population totale	28 871 459	28 871 459	0	0,00	29 304 855	29 304 855	0	0,00	0	0,00
Hommes ≥ 15 ans	11 340 286	11 340 337	51	0,00	11 487 144	11 477 463	-9 681	-0,08	-9 732	-0,08
Femmes ≥ 15 ans	11 998 509	11 998 458	-51	0,00	12 139 636	12 133 442	-6 194	-0,05	-6 144	-0,05
Population totale ≥ 15 ans	23 338 795	23 338 795	0	0,00	23 626 780	23 610 904	-15 876	-0,07	-15 876	-0,07
Âge 0 à 4	1 636 092	1 636 651	559	0,03	1 682 077	1 687 571	5 494	0,33	4 935	0,29
Âge 5 à 9	1 910 359	1 909 567	-792	-0,04	1 960 872	1 966 069	5 197	0,27	5 990	0,31
Âge 10 à 14	1 986 213	1 986 447	234	0,01	2 035 126	2 040 311	5 185	0,25	4 951	0,24
Âge 15 à 19	1 986 163	1 986 942	779	0,04	2 026 860	2 024 694	-2 166	-0,11	-2 945	-0,15
Âge 20 à 24	1 892 572	1 892 068	-504	-0,03	1 922 977	1 918 522	-4 455	-0,23	-3 951	-0,21
Âge 25 à 29	1 835 744	1 834 959	-785	-0,04	1 866 784	1 863 210	-3 574	-0,19	-2 789	-0,15
Âge 30 à 34	2 031 513	2 031 520	7	0,00	2 063 738	2 062 711	-1 027	-0,05	-1 034	-0,05
Âge 35 à 39	2 452 299	2 451 743	-556	-0,02	2 484 983	2 483 560	-1 423	-0,06	-867	-0,03
Âge 40 à 44	2 510 847	2 510 947	100	0,00	2 540 694	2 539 345	-1 349	-0,05	-1 449	-0,06
Âge 45 à 49	2 273 676	2 274 363	687	0,03	2 297 674	2 296 514	-1 160	-0,05	-1 847	-0,08
Âge 50 à 54	2 031 050	2 030 963	-87	0,00	2 051 231	2 048 768	-2 463	-0,12	-2 376	-0,12
Âge 55 à 59	1 549 675	1 549 756	81	0,01	1 564 428	1 563 521	-907	-0,06	-988	-0,06
Âge 60 à 64	1 234 930	1 235 863	933	0,08	1 246 010	1 246 568	558	0,04	-375	-0,03
Âge 65 à 74	2 059 079	2 059 350	271	0,01	2 073 468	2 074 803	1 335	0,06	1 065	0,05
Âge 75 et plus	1 481 247	1 480 321	-926	-0,06	1 487 933	1 488 687	754	0,05	1 680	0,11

Caractéristique	Recensement de 2001 – 1 <sup>er</sup> passage				Recensement de 2001 – 2 <sup>e</sup> passage				passage 2 – passage 1	
	Chiffre	Estimation	Différence	Écart	Chiffre	Estimation	Différence	Écart	Différence	Écart
Célibataires	13 282 845	13 282 792	-53	0,00	13 576 338	13 578 613	2 275	0,02	2 328	0,02
Personnes mariées	11 750 092	11 750 035	-57	0,00	11 853 964	11 854 210	246	0,00	303	0,00
Veufs et veuves	1 341 497	1 342 054	557	0,04	1 353 562	1 354 561	999	0,07	442	0,03
Personnes divorcées	1 794 079	1 794 285	206	0,01	1 807 982	1 805 493	-2 489	-0,14	-2 695	-0,15
Personnes séparées	702 946	702 293	-653	-0,09	713 009	711 977	-1 032	-0,14	-379	-0,05
Union libre = oui	2 267 634	2 271 749	4 115	0,18	2 322 437	2 329 084	6 647	0,29	2 532	0,10
Ménages d'une personne	2 908 857	2 904 682	-4 175	-0,14	2 932 655	*	*	*	*	*
Ménages de deux personnes	3 709 282	3 708 376	-906	-0,02	3 736 957	*	*	*	*	*
Ménages de trois personnes	1 848 476	1 843 466	-5 010	-0,27	1 868 996	*	*	*	*	*
Ménages de quatre personnes	1 812 783	1 815 197	2 414	0,13	1 833 471	*	*	*	*	*
Ménages de cinq personnes	714 618	723 436	8 818	1,23	729 190	*	*	*	*	*
Ménages de six personnes ou plus	332 959	331 817	-1 142	-0,34	352 349	*	*	*	*	*

\* Données non disponibles

**Nota :** Les personnes dénombrées à l'aide d'une formule 2C (personnes dénombrées à l'extérieur du Canada) sont comprises dans les chiffres et estimations du deuxième passage, mais exclues des chiffres et estimations du premier passage.

**Graphique 7.2.2.3 : Écarts population/estimation fondés sur les poids d'échantillonnage définitifs pour le Canada, comparaison entre le premier passage et le deuxième passage, recensement de 2001**



## 7.2.3 Retraitement des contraintes

Pour le recensement de 2001, on a ajusté les paramètres du système de pondération (voir la section 4.4), de sorte que le nombre de contraintes qui ont été retranchées a diminué par rapport à 1996. C'est ce que l'on observe dans la présente section. Ainsi, comme on l'a vu à la section 7.2.2, les écarts population/estimation du recensement de 2001 sont plus faibles que ceux du recensement de 1996.

Le tableau 7.2.3.1 montre la fréquence à laquelle chacune des 32 contraintes a été retranchée dans les 6 141 RP échantillonnées en 2001 et les 5 941 RP échantillonnées en 1996. Parfois la raison du retraitement d'une contrainte (c.-à-d. parce qu'elle est faible, linéairement dépendante ou quasi linéairement dépendante, ou parce qu'elle cause des poids d'échantillonnage aberrants [se reporter à la section 4.4]) permet de mieux comprendre l'ampleur de certains écarts population/estimation fournis dans le graphique 7.2.2.2. L'analyse qui suit portera surtout sur les résultats du recensement de 2001. 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 un faible écart population/estimation. De plus, une contrainte comme « Union libre = oui » ou « Ménages de 5 personnes » qui est retranchée fréquemment parce qu'elle cause des poids d'échantillonnage aberrants ou parce qu'elle est quasi linéairement dépendante (comme c'est le cas pour « Ménages d'une personne », « ...de trois personnes » et « ...de quatre personnes ») peut causer de grands écarts population/estimation, comme on peut le voir dans le graphique 7.2.2.2.

Le tableau 7.2.3.2 présente un résumé de l'information fournie au tableau 7.2.3.1. Dans le premier, on note que le nombre de contraintes linéairement dépendantes retranchées en 1996 a été majoré de 2, afin de tenir compte du fait que les contraintes « Personnes séparées » et « Ménages de six personnes ou plus » n'ont pas été utilisées en 1996 parce qu'elles étaient linéairement dépendantes d'autres contraintes (voir l'annexe B). En 2001, le paramètre FAIBLE a été augmenté pour certaines RP. Par conséquent, on note au tableau 7.2.3.2 que le nombre de contraintes retranchées parce qu'elles étaient faibles est passé de 0,1 en 1996 à 0,4 en 2001. En outre, les contraintes COND et MAXC ont été accrues

pour certaines RP en 2001, de sorte que le tableau 7.2.3.2 montre que le nombre de contraintes retranchées parce qu'elles étaient linéairement dépendantes a chuté de 1,6 en 1996 à 1,0 en 2001.

Le tableau 7.2.3.3 présente 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 huit AD. À ce moment, elle compterait 16 contraintes au niveau des AD. Le tableau 7.2.3.3 montre qu'en 2001, le nombre de ces contraintes à avoir été retranchées parce qu'elles étaient linéairement dépendantes était de 0,7 contrairement à 2,2 en 1996. Cette différence est attribuable au fait que pour certaines RP la valeur des paramètres COND et MAXC s'est trouvée être augmentée pour 2001. En l'absence de chiffres portant sur le nombre de contraintes retranchées au niveau des AD au recensement de 1996, des chiffres approximatifs ont été produits (voir le tableau 7.2.3.3) à partir des données de 2001 avec les paramètres du système de pondération de 1996.

**Tableau 7.2.3.1 : Fréquence à laquelle des contraintes au niveau des RP ont été retranchées en 1996 et en 2001 lors de l'ajustement des poids d'échantillonnage définitifs**

Caractéristique	Recensement de 2001					Recensement de 1996				
	Faible	LD	QLD	Aber.	Total	Faible	LD	QLD	Aber.	Total
Hommes	0	0	0	0	0	0	0	0	1	1
Femmes**	-	-	-	-	-	-	-	-	-	-
Population totale	0	0	0	0	0	0	0	0	0	0
Hommes ≥ 15	0	4	24	27	55	0	1	136	3	140
Personnes ≥ 15	0	0	0	0	0	0	0	1	0	1
Total des ménages	0	0	0	0	0	0	0	0	0	0
Âge 0 à 4	29	4 286	2	124	4 441	6	3 071	20	57	3 154
Âge 5 à 9	68	406	4	251	729	30	709	77	135	951
Âge 10 à 14	79	1 359	2	141	1 581	35	2 110	33	61	2 239
Âge 15 à 19	18	492	6	131	647	6	514	27	96	643
Âge 20 à 24	2	243	15	125	385	1	216	133	119	469
Âge 25 à 29	3	877	9	94	983	1	347	108	82	538
Âge 30 à 34	3	158	5	83	249	1	29	23	42	95
Âge 35 à 39	3	6	1	35	45	1	0	6	31	38
Âge 40 à 44	2	0	0	19	21	1	3	13	45	62
Âge 45 à 49	2	2	3	41	48	1	4	9	50	64
Âge 50 à 54	2	7	1	38	48	2	157	67	83	309
Âge 55 à 59	3	238	7	79	327	2	636	213	147	998
Âge 60 à 64	5	1 751	65	130	1 951	3	1 122	973	128	2 226
Âge 65 à 74	5	2	32	49	88	4	3	214	81	302
Âge 75 et plus	42	2 308	8	38	2 396	36	2 864	100	60	3 060
Célibataires	1	0	0	2	3	0	0	1	3	4
Personnes mariées	1	1	0	2	4	0	0	0	4	4
Veufs et veuves	6	593	15	128	742	2	0	174	345	521
Personnes divorcées	3	15	11	94	123	1	1	213	252	467

Caractéristique	Recensement de 2001					Recensement de 1996				
	Faible	LD	QLD	Aber.	Total	Faible	LD	QLD	Aber.	Total
Personnes séparées*	20	5 510	3	34	5 567	-	-	-	-	-
Union libre = oui	16	0	0	278	294	23	0	1	272	296
Ménages d'une personne	2	194	1 869	22	2 087	1	12	4 583	4	4 600
Ménages de deux personnes	1	2	310	15	328	0	0	1 154	12	1 166
Ménages de trois personnes	7	40	2 537	42	2 626	2	22	189	47	260
Ménages de quatre personnes	50	187	1 102	98	1 437	23	145	52	37	257
Ménages de cinq personnes	401	1 206	143	281	2 031	193	997	865	92	2 147
Ménages de six personnes ou plus*	1 941	3 960	121	9	6 031	-	-	-	-	-

\* Indique que la caractéristique n'a pas été utilisée comme contrainte en 1996 parce qu'il s'agissait d'une contrainte redondante.

\*\* Indique que la caractéristique n'a pas été utilisée comme contrainte en 1996 ou 2001 parce qu'il s'agissait d'une contrainte redondante.

Faible = contrainte faible

LD = contrainte linéairement dépendante

QLD = contrainte quasi linéairement dépendante

Aber. = contrainte qui cause des poids d'échantillonnage aberrants

**Tableau 7.2.3.2 : Fréquence à laquelle des contraintes au niveau des RP ont été retranchées en 1996 et en 2001 lors de l'ajustement des poids d'échantillonnage définitifs – Statistiques sommaires**

	Recensement de 2001					Recensement de 1996				
	Faible	LD	QLD	Aber.	Total	Faible	LD	QLD	Aber.	Total
Total des contraintes retranchées	2 715	23 847	6 295	2 410	35 267	375	12 963	9 385	2 289	25 012
N <sup>bre</sup> de contraintes retranchées par RP	0,4	3,9	1,0	0,4	5,7	0,1	2,2	1,6	0,4	4,2
Total corrigé pour tenir compte des deux contraintes non utilisées en 1996 (LD)						375	24 845	9 385	2 289	36 894
N <sup>bre</sup> de contraintes retranchées par RP						0,1	4,2	1,6	0,4	6,2
Totaux combinés	Faible + LD	26 562	QLD + Aber.	8 705	35 267	Faible + LD	25 220	QLD + Aber.	11 674	36 894
N <sup>bre</sup> de contraintes retranchées par RP		4,3		1,4	5,7		4,2		2,0	6,2

Faible = contrainte faible

LD = contrainte linéairement dépendante

QLD = contrainte quasi linéairement dépendante

Aber. = contrainte qui cause des poids d'échantillonnage aberrants

**Tableau 7.2.3.3 : Fréquence à laquelle des contraintes au niveau des AD ont été retranchées en 1996 et en 2001 lors de l'ajustement des poids d'échantillonnage définitifs – Statistiques sommaires**

	Recensement de 2001					Recensement de 1996**				
	Faible	LD	QLD	Aber.	Total	Faible	LD	QLD	Aber.	Total
Total des contraintes retranchées	1 354	357	4 191	917	6 819	1 082	393	12 973	1 069	15 517
N <sup>bre</sup> de contraintes retranchées par RP	0,2	0,1	0,7	0,1	1,1	0,2	0,1	2,2	0,2	2,6
Totaux combinés	Faible + LD	1 711	QLD + Aber.	5 108	6 819	Faible + LD	1 475	QLD + Aber.	14 042	15 517
N <sup>bre</sup> de contraintes retranchées par RP		0,3		0,8	1,1		0,2		2,4	2,6

\*\* Les chiffres du recensement de 1996 sont calculés à partir des données de 2001 avec les paramètres du système de 1996.

Faible = contrainte faible

LD = contrainte linéairement dépendante

QLD = contrainte quasi linéairement dépendante

Aber. = contrainte qui cause des poids d'échantillonnage aberrants

## 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 d'échantillonnage définitifs pour le Canada, calculés selon la méthode suivante :

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

Les estimations-échantillon et les chiffres de population portent sur les logements occupés dans les SD échantillonnés.

Dans le présent chapitre, les écarts population/échantillon des recensements de 1996 et de 2001 seront étudiés en fonction des régions géographiques suivantes :

- 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 12,8 % des SDR et 65,4 % 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. Dans le cas des régions géographiques de plus petite taille que les RP (comme les AD), les écarts population/estimation sont habituellement plus marqués.

Les graphiques et tableaux de ce chapitre présentent les centiles des écarts population/estimation pour 31 caractéristiques qui, à quelques exceptions près, sont identiques aux 32 contraintes au niveau des RP appliquées aux poids d'échantillonnage (voir l'annexe B). Définissons le terme **centile** au moyen d'un exemple. Ainsi, le tableau 8.2.1 affiche un centile de -6,07 % pour les ménages de six personnes ou plus au recensement de 2001. Ce chiffre signifie que 10 % des RP ont un écart de -6,07 % ou moins. Un 90<sup>e</sup> centile de 7,98 % signifie que 10 % des RP ont un écart de 7,98 % 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.

Les centiles au niveau des RP pour toutes les caractéristiques et les centiles pour la contrainte « Nombre total de ménages » ont été difficiles à obtenir pour le recensement de 1996. On a produit des estimations très approximatives des résultats de 1996 en soumettant les données du recensement de 2001 au

système de pondération pour les contraintes de 2001 énoncées à l'annexe B et en appliquant tous les autres paramètres comme en 1996.

Les sections qui suivent permettent de constater que, de façon générale, les écarts population/estimation du recensement de 2001 pour le Canada, les DR et les RP étaient plus faibles que ceux du recensement de 1996 tandis qu'au niveau des AD et des SR, ils étaient quelque peu plus élevés. Cela concorde avec l'objectif établi en vue de 2001 qui était de réduire l'importance des écarts pour les régions de niveau supérieur, toujours à partir de poids d'échantillonnage d'au moins 1.

## 8.1 Aires de diffusion

Le Canada est divisé en 52 993 AD, dont 47 933 étaient sujettes à l'échantillonnage. La population d'une AD peut varier entre 400 et 700 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.

L'aire de diffusion (AD) a été introduite au recensement de 2001 (voir la section 4.2). En 1996, son rôle était joué par le secteur de dénombrement (SD), ce qui explique pourquoi les centiles de 1996 dans les graphiques 8.1.1 et 8.1.2 sont présentés au niveau des SD, alors que ceux de 2001 sont présentés au niveau des AD. Pour presque toutes les caractéristiques, les intervalles des AD de 2001 sont un peu plus grands que ceux des SD de 1996 entre les 10<sup>e</sup> et 90<sup>e</sup> centiles et entre les 25<sup>e</sup> et 75<sup>e</sup> centiles, probablement parce que le paramètre FAIBLE (voir la section 4.4) était fixé à 20 en 1996, alors qu'en 2001 il était établi à 30 ou à 40 pour un bon nombre de RP. Le fait d'attribuer une valeur plus élevée au paramètre FAIBLE en 2001 a fait en sorte qu'un plus grand nombre de contraintes ont été retranchées; les écarts ont donc été plus grands lors de l'ajustement de la première étape au niveau des AD, ce qui a eu tendance à augmenter l'importance des écarts obtenus à la suite de l'ajustement de la deuxième étape pour les 32 contraintes au niveau des AD, alors que ce ne fut pas le cas en 1996.

Trois caractéristiques présentées dans le graphique 8.1.2 méritent un examen plus approfondi. Les intervalles entre les 10<sup>e</sup> et 90<sup>e</sup> centiles et entre les 25<sup>e</sup> et 75<sup>e</sup> centiles pour la caractéristique « État matrimonial = séparé » sont plus petits en 2001 qu'en 1996, parce que cette caractéristique a été utilisée comme contrainte de pondération pour 2001, mais non pour 1996 (voir l'annexe B). L'intervalle entre les 10<sup>e</sup> et 90<sup>e</sup> centiles était nul en 2001 pour la « population totale », alors qu'il ne l'était pas en 1996, ce qui peut s'expliquer par le fait que beaucoup moins de contraintes au niveau des AD ont été retranchées à la deuxième étape en 2001 parce qu'elles étaient quasi linéairement dépendantes (voir le tableau 7.2.3.3.). En outre, le paramètre MAXC de 1996 (voir la section 4.4) était établi à 10 000, alors qu'en 2001 il se situait généralement entre 20 000 et 160 000, ce qui permettait de conserver un plus grand nombre de contraintes. Enfin, les intervalles entre les 10<sup>e</sup> et 90<sup>e</sup> centiles et entre les 25<sup>e</sup> et 75<sup>e</sup> centiles pour la caractéristique « Union libre = oui » sont beaucoup plus grands en 2001. Le tableau 7.2.2.1 montre que les écarts population/estimation fondés sur les poids d'échantillonnage initiaux au niveau du Canada pour 2001 et 1996 se chiffraient à -14 381 et -1 404 respectivement pour cette caractéristique. On ignore la raison de cette augmentation de l'ampleur de l'écart en 2001. L'écart population/estimation fondé sur les poids d'échantillonnage définitifs au niveau du Canada était descendu à 4 115 en 2001 et à 2 415 en 1996. Étant donné ces tendances au niveau du Canada, il n'est pas surprenant que les intervalles pour cette contrainte soient plus grands au niveau des AD en 2001 qu'en 1996. Il n'en reste pas moins que l'ampleur de l'augmentation de ces intervalles est surprenante.

## 8.2 Régions de pondération

Le Canada (à l'exclusion des Territoires du Nord-Ouest et du Nunavut) est divisé en 6 148 RP, dont 6 141 sont échantillonnées. Les RP ont une population moyenne de 4 701 personnes et sont

habituellement constituées de huit AD entières. Bien qu'on s'en serve pour calculer les poids d'échantillonnage du recensement, on ne publie pas de données au niveau des RP.

Le tableau 8.2.1 montre que les 10<sup>e</sup>, 25<sup>e</sup>, 50<sup>e</sup>, 75<sup>e</sup> et 90<sup>e</sup> centiles correspondent à zéro pour presque toutes les caractéristiques de la population, tant en 2001 qu'en 1996. Dans le cas des caractéristiques des ménages, la plupart des 25<sup>e</sup>, 50<sup>e</sup> et 75<sup>e</sup> centiles correspondent également à zéro, alors que certains des 10<sup>e</sup> et 90<sup>e</sup> 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. Il convient de prendre note que les chiffres de 1996 sont estimés à partir des données de 2001 et des paramètres de pondération utilisés en 1996.

### 8.3 Subdivisions de recensement

Le Canada est divisé en 5 600 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 5 400 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 12,8 % 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. Au recensement de 2001, les intervalles au niveau des SDR entre les 10<sup>e</sup> et 90<sup>e</sup> centiles sont plus petits pour la plupart des contraintes, mais d'une ampleur similaire à celle des intervalles observés pour les 10<sup>e</sup> et 90<sup>e</sup> centiles au niveau des AD. On suppose que cela s'explique par le fait que 84,5 % des SDR sont formées uniquement d'une partie de RP (voir le tableau 7.1.2); on ne peut donc pas s'attendre à une cohérence parfaite entre les chiffres de population et les estimations pour la plupart des contraintes. En revanche, les intervalles observés pour les 25<sup>e</sup> et 75<sup>e</sup> centiles au niveau des SDR sont beaucoup plus petits que ceux observés au niveau des AD, ce qui résulte probablement du fait que certaines des contraintes sont appliquées à de plus grandes municipalités, qui peuvent être des regroupement de RP complètes.

Certains écarts population/estimation au recensement de 2001 ont diminué par rapport à 1996, et pour les caractéristiques « Âge 75 et plus », « Veufs et veuves », « Personnes séparées » et « Personnes divorcées », la réduction a été importante. Par contre, d'autres caractéristiques comme « Ménages de trois personnes » et « Ménages de six personnes ou plus » ont enregistré une moins bonne concordance en comparaison avec 1996.

### 8.4 Secteurs de recensement

Les secteurs de recensement (SR) se retrouvent uniquement dans les grands centres urbains dont le noyau urbain compte 50 000 habitants ou plus. Il y a 4 798 SR au Canada. Les SR comptent habituellement entre 2 500 et 8 000 habitants, la moyenne étant d'environ 4 400 personnes. Le tableau 7.1.2 montre que 65,4 % 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. Étant donné que 32,9 % des SR sont formés uniquement d'une partie de RP (voir le tableau 7.1.2), il n'est pas surprenant que les 10<sup>e</sup> et 90<sup>e</sup> centiles soient relativement élevés en 2001. Il est toutefois plus surprenant de constater l'ampleur de la différence entre les centiles de 2001 et ceux de 1996, qui peut être attribuable en partie au fait que les écarts au niveau des AD de 2001 sont un peu plus grands que ceux de 1996 (voir les graphiques 8.1.1 et 8.1.2). Les 25<sup>e</sup> et 75<sup>e</sup> centiles pour les écarts correspondent généralement à zéro (probablement parce que 65,4 % des SR sont formés de RP complètes). Par conséquent, ils ne sont pas inclus dans les graphiques.

## 8.5 Divisions de recensement

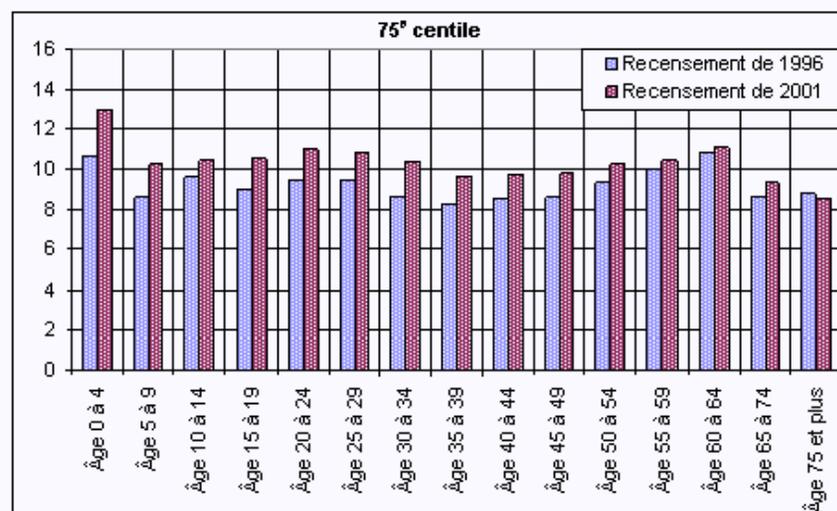
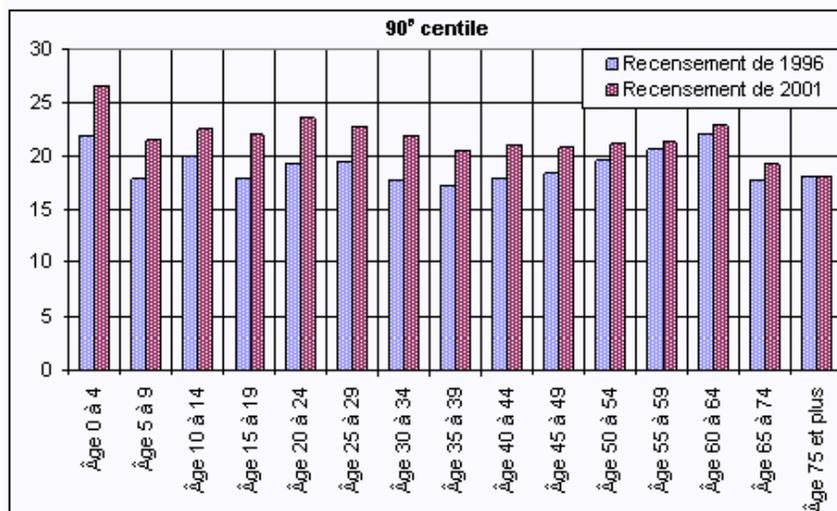
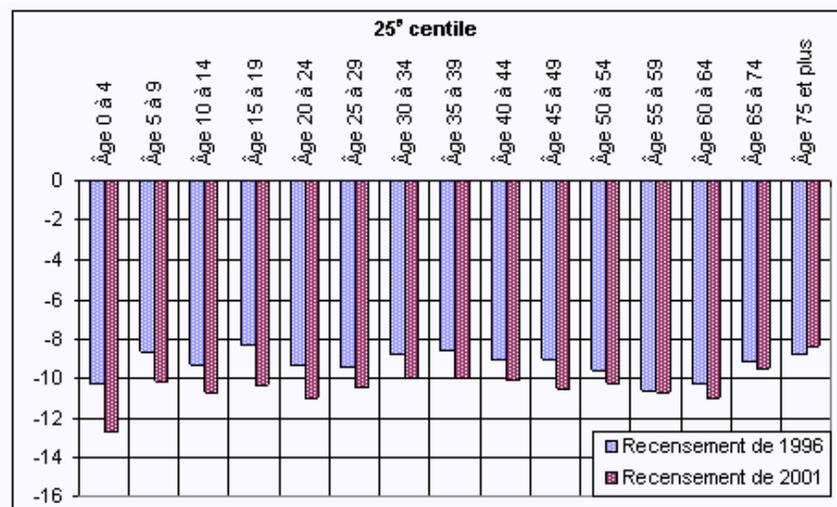
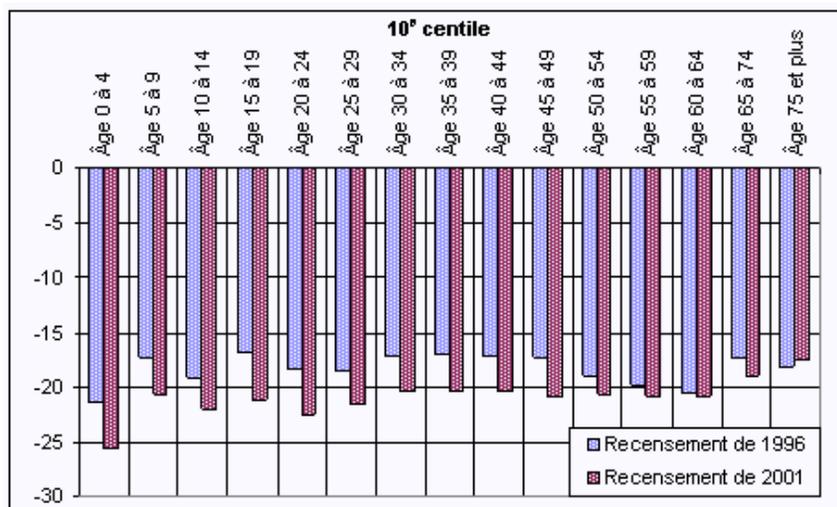
Le Canada est divisé en 288 DR, qui comptent en moyenne 104 000 habitants. Une DR peut correspondre à un comté, à une municipalité régionale, à un district régional ou à toute autre région créée en vertu d'une loi provinciale ou territoriale.

Le tableau 8.5.1 présente un résumé des résultats relatifs au niveau de cohérence des recensements de 2001 et de 1996 pour les DR échantillonnées. Toutes les DR sont formées de RP complètes. Ainsi, les caractéristiques qui constituent des contraintes de pondération et qui étaient rarement retranchées affichent une cohérence parfaite ou quasi parfaite au niveau des DR<sup>4</sup>. Pour ce qui est des autres caractéristiques, on constate de façon générale que les centiles de 2001 sont inférieurs à ceux de 1996 dans le cas des caractéristiques de la population, alors que dans le cas des ménages, c'est l'inverse qui se produit. Cette observation concorde avec ce que l'on a observé au tableau 7.2.2.1 au sujet des écarts population/estimation pour le Canada.

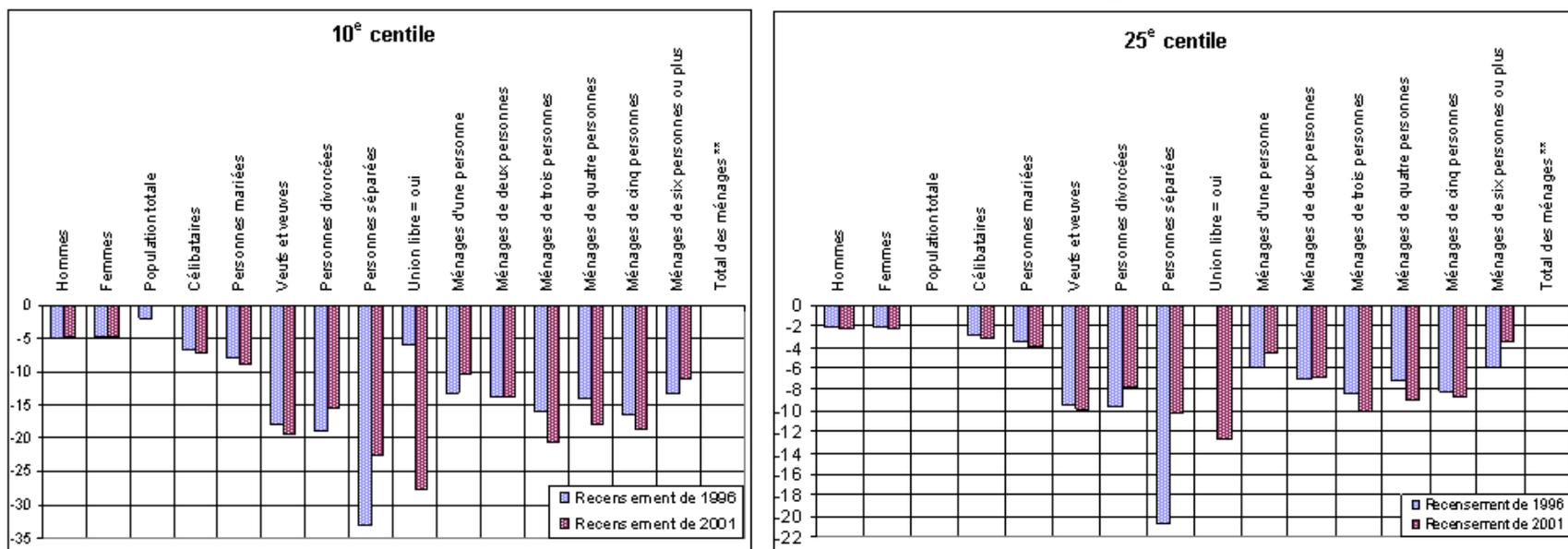
---

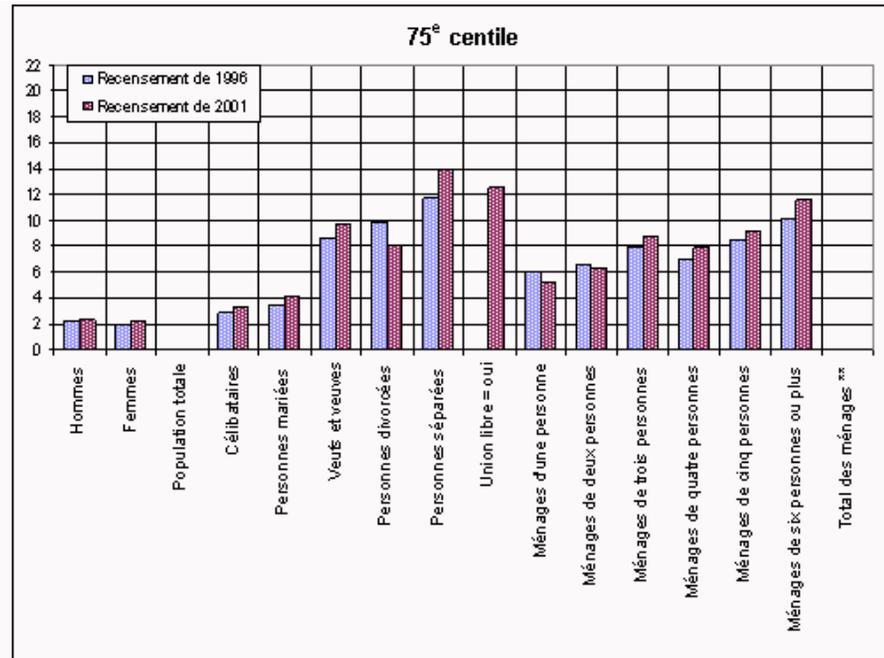
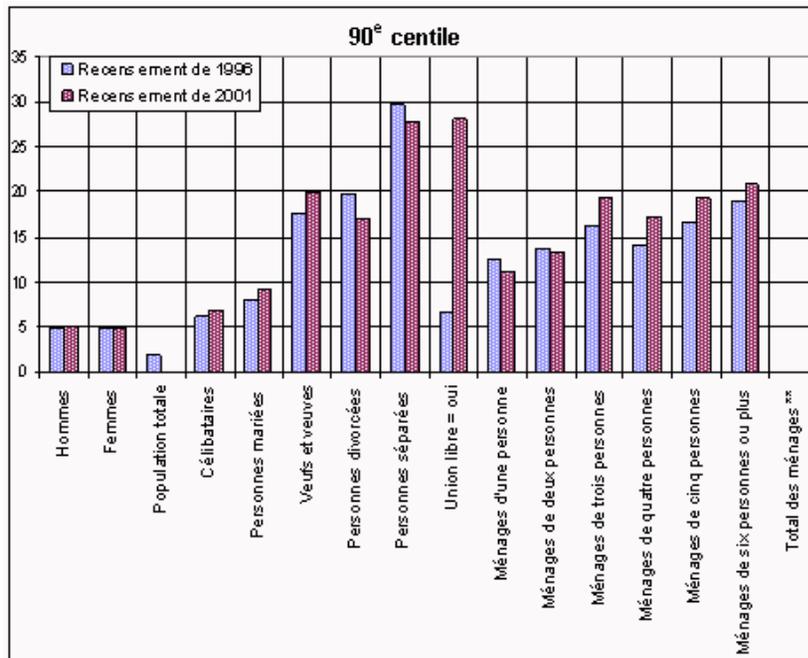
<sup>4</sup> Même lorsque les caractéristiques présentent une cohérence parfaite, les totalisations publiées des caractéristiques de base établies à partir des données-échantillon ne concordent pas parfaitement avec les totalisations des mêmes caractéristiques basées sur les données intégrales. Cette différence est attribuable au fait que l'on utilise les résultats du deuxième passage pour traiter les données des questionnaires abrégés (données-échantillon) et ceux du premier passage pour traiter les données intégrales (100 %) (voir la section 4.5). En outre, les résidents d'établissements institutionnels (voir la section 3.2) sont inclus dans les totalisations faites à partir des données intégrales, mais exclus des totalisations fondées sur les données-échantillon.

**Graphique 8.1.1 : Centiles des écarts population/estimation pour les AD (recensement de 2001) et les SD (recensement de 1996) pour les groupes d'âge**



Graphique 8.1.2 : Centiles des écarts population/estimation pour les AD (recensement de 2001) et les SD (recensement de 1996) pour les autres caractéristiques de la population et les caractéristiques des ménages





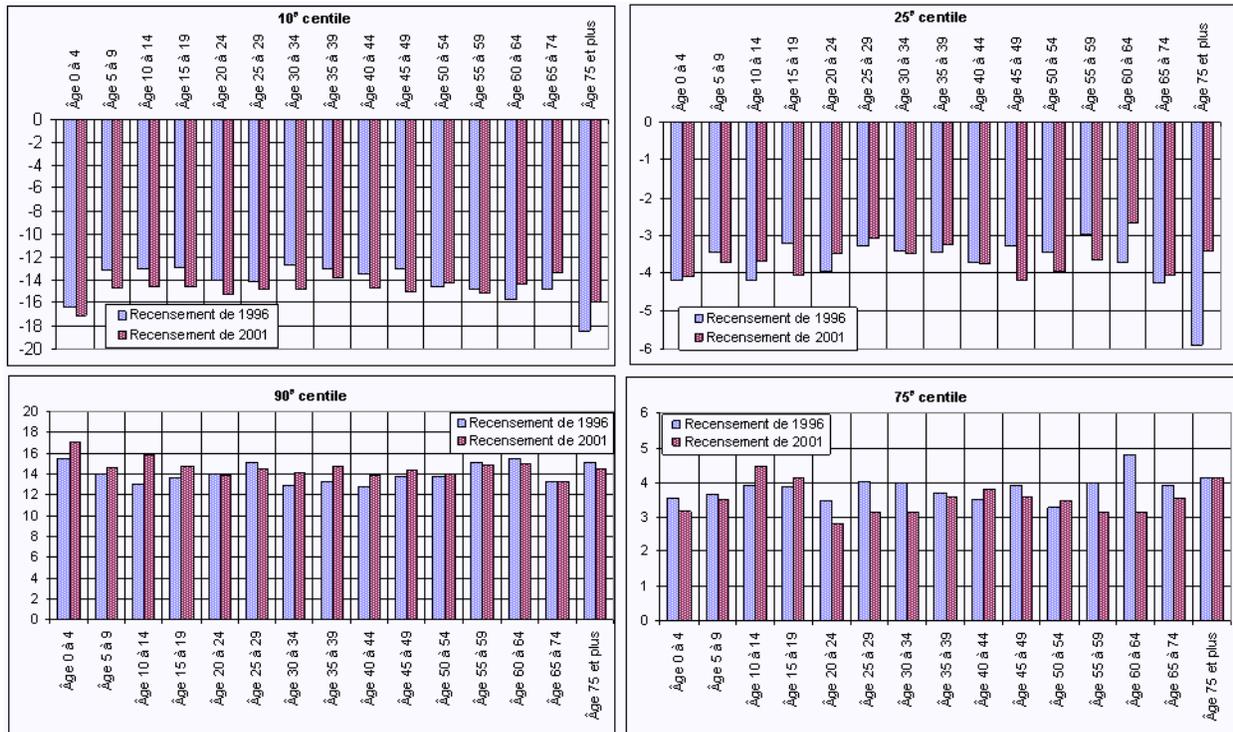
\*\* Les centiles du total des ménages pour 1996 sont estimés à partir des données de 2001.

**Tableau 8.2.1 : Centiles des écarts population/estimation pour les RP**

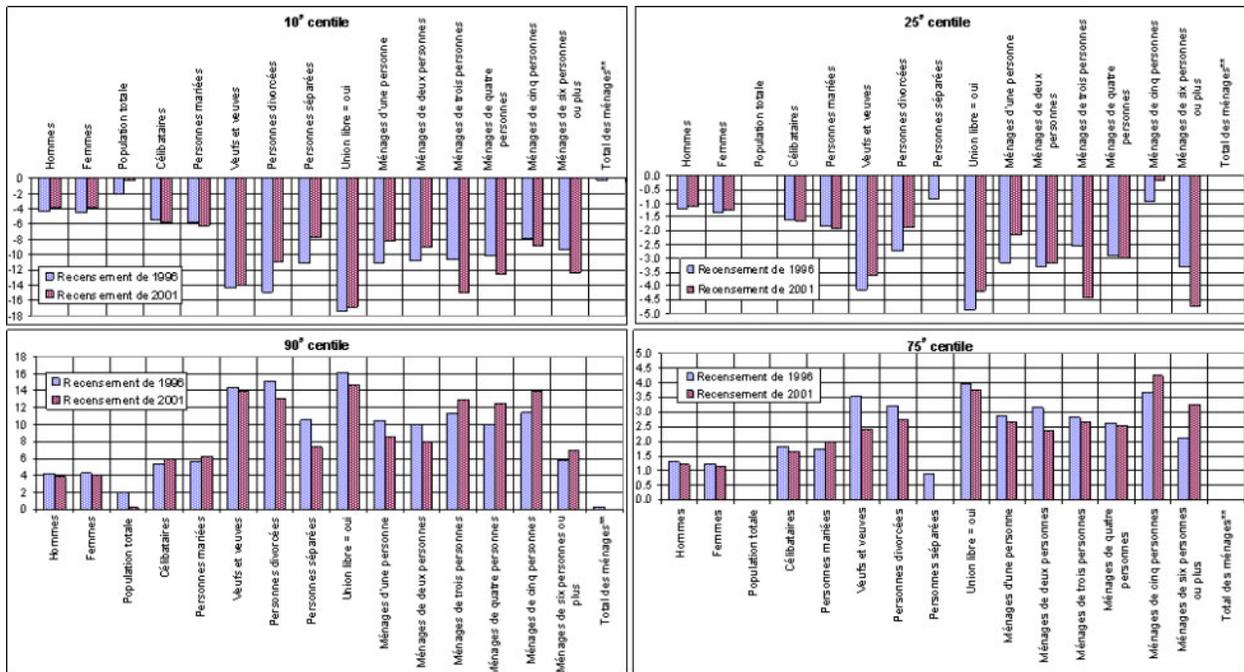
Caractéristiques	Centiles de 2001					Centiles de 1996**				
	10 <sup>e</sup>	25 <sup>e</sup>	50 <sup>e</sup>	75 <sup>e</sup>	90 <sup>e</sup>	10 <sup>e</sup>	25 <sup>e</sup>	50 <sup>e</sup>	75 <sup>e</sup>	90 <sup>e</sup>
<b>Caractéristiques de la population</b>										
Hommes	0	0	0	0	0	0	0	0	0	0
Femmes	0	0	0	0	0	0	0	0	0	0
Total de la population	0	0	0	0	0	0	0	0	0	0
Âge 0 à 4	0	0	0	0	0	0	0	0	0	0
Âge 5 à 9	0	0	0	0	0	0	0	0	0	0
Âge 10 à 14	0	0	0	0	0	0	0	0	0	0
Âge 15 à 19	0	0	0	0	0	0	0	0	0	0
Âge 20 à 24	0	0	0	0	0	0	0	0	0	0
Âge 25 à 29	0	0	0	0	0	0	0	0	0	0
Âge 30 à 34	0	0	0	0	0	0	0	0	0	0
Âge 35 à 39	0	0	0	0	0	0	0	0	0	0
Âge 40 à 44	0	0	0	0	0	0	0	0	0	0
Âge 45 à 49	0	0	0	0	0	0	0	0	0	0
Âge 50 à 54	0	0	0	0	0	0	0	0	0	0
Âge 55 à 59	0	0	0	0	0	0	0	0	0	0
Âge 60 à 64	0	0	0	0	0	0	0	0	0	1,31
Âge 65 à 74	0	0	0	0	0	0	0	0	0	0
Âge 75 et plus	0	0	0	0	0	-1,98	0	0	0	0
Célibataires	0	0	0	0	0	0	0	0	0	0
Personnes mariées	0	0	0	0	0	0	0	0	0	0
Veufs et veuves	0	0	0	0	0	0	0	0	0	0
Personnes divorcées	0	0	0	0	0	0	0	0	0	0
Personnes séparées	0	0	0	0	0	0	0	0	0	0
Union libre = oui	0	0	0	0	0	0	0	0	0	0
<b>Caractéristiques des ménages</b>										
Ménages d'une personne	-1,11	0	0	0	0,03	-1,33	-0,53	0	0,04	0,57
Ménages de deux personnes	0	0	0	0	0	-0,24	0	0	0	0
Ménages de trois personnes	-1,80	-0,18	0	0	0,40	0	0	0	0	0
Ménages de quatre personnes	-0,06	0	0	0	0,16	0	0	0	0	0
Ménages de cinq personnes	0	0	0	0	7,89	-0,57	0	0	0	7,40
Ménages de six personnes ou plus	-6,07	-1,57	1,16	4,63	7,98	-4,89	-1,59	0,91	3,62	6,18
Total des ménages	0	0	0	0	0	0	0	0	0	0

\*\* Les centiles pour 1996 sont estimés à partir des données de 2001.

**Graphique 8.3.1 : Centiles des écarts population/estimation pour les SDR pour les groupes d'âge**

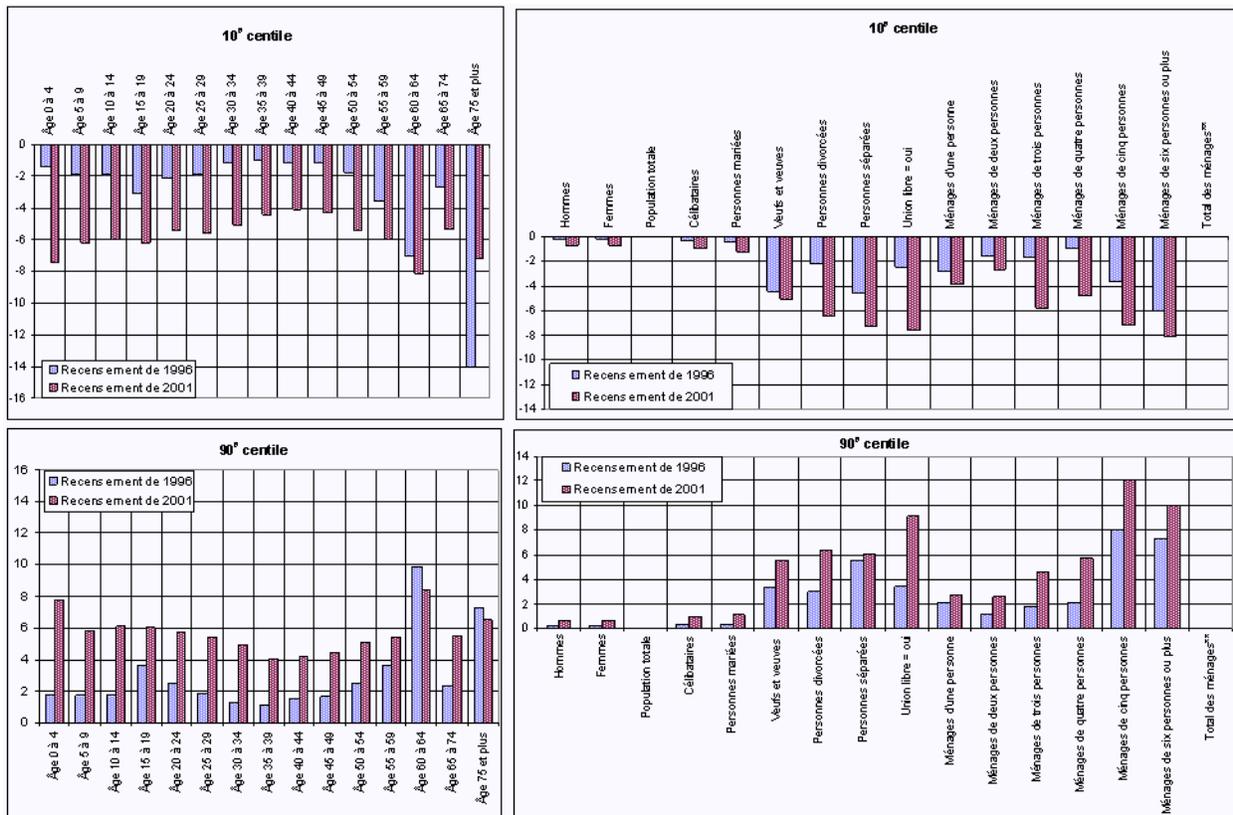


**Graphique 8.3.2 : Centiles des écarts population/estimation pour les SDR pour les autres caractéristiques de la population et les caractéristiques des ménages**



\*\* Les centiles du total des ménages pour 1996 sont estimés à partir des données de 2001.

**Graphique 8.4.1 : Centiles des écarts population/estimation pour les SR**



\*\* Les centiles du total des ménages pour 1996 sont estimés à partir des données de 2001.

**Tableau 8.5.1 : Centiles des écarts population/estimation pour les DR**

Caractéristiques	Centiles de 2001					Centiles de 1996				
	10 <sup>e</sup>	25 <sup>e</sup>	50 <sup>e</sup>	75 <sup>e</sup>	90 <sup>e</sup>	10 <sup>e</sup>	25 <sup>e</sup>	50 <sup>e</sup>	75 <sup>e</sup>	90 <sup>e</sup>
<b>Caractéristiques de la population</b>										
Hommes	0	0	0	0	0	0	0	0	0	0
Femmes	0	0	0	0	0	0	0	0	0	0
Total de la population	0	0	0	0	0	0	0	0	0	0
Âge 0 à 4	-0,81	0	0	0	0,72	-0,44	0	0	0	0,05
Âge 5 à 9	-0,51	0	0	0	0,23	-0,10	0	0	0	0,21
Âge 10 à 14	-0,20	0	0	0	0,36	0	0	0	0	0,04
Âge 15 à 19	-0,05	0	0	0	0,06	0	0	0	0	0,43
Âge 20 à 24	-0,41	0	0	0	0,29	-0,89	0	0	0,09	1,17
Âge 25 à 29	-0,54	0	0	0	0,44	-0,95	0	0	0	1,32
Âge 30 à 34	-0,04	0	0	0	0,08	0	0	0	0	0
Âge 35 à 39	0	0	0	0	0	0	0	0	0	0
Âge 40 à 44	0	0	0	0	0	0	0	0	0	0
Âge 45 à 49	0	0	0	0	0	0	0	0	0	0
Âge 50 à 54	0	0	0	0	0	-0,17	0	0	0	0,21
Âge 55 à 59	0	0	0	0	0	-1,05	0	0	0	0,85
Âge 60 à 64	-0,18	0	0	0	0,49	-1,51	-0,01	0	0,92	2,52
Âge 65 à 74	0	0	0	0	0	-0,25	0	0	0	0,06
Âge 75 et plus	-0,65	0	0	0	0,28	-3,70	-1,94	-0,17	0	0,81
Célibataires	0	0	0	0	0	0	0	0	0	0
Personnes mariées	0	0	0	0	0	0	0	0	0	0
Veufs et veuves	-0,14	0	0	0	0,04	-1,53	-0,33	0	0,09	1,60
Personnes divorcées	-0,08	0	0	0	0,31	-1,00	0	0	0,34	2,04
Personnes séparées	-0,96	0	0	0	0,94	-7,29	-2,25	0	1,15	4,30
Union libre = oui	-0,21	0	0	0	0,84	-0,86	0	0	0	1,36
<b>Caractéristiques des ménages</b>										
Ménages d'une personne	-0,34	-0,15	-0,04	0	0,01	-0,57	-0,36	-0,20	-0,10	0,10
Ménages de deux personnes	-0,04	0	0	0	0	-0,10	0	0	0	0,01
Ménages de trois personnes	-1,05	-0,59	-0,23	0	0,29	0	0	0	0	0,10
Ménages de quatre personnes	-0,27	-0,07	0	0,16	0,67	0	0	0	0	0,16
Ménages de cinq personnes	-0,79	0	0,99	2,47	5,16	-0,59	0	0,60	1,81	3,09
Ménages de six personnes ou plus	-9,17	-3,86	-0,76	1,65	3,87	-6,55	-3,39	-0,84	0,99	2,65
Total des ménages**	0	0	0	0	0	0	0	0	0	0

\*\* Les centiles du total des ménages pour 1996 sont estimés à partir des données de 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 vraie de la variable estimée. Le chapitre 6 présente les résultats de l'Étude 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-échantillon<sup>5</sup>. L'Étude de la variance d'échantillonnage a été effectuée 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 sur la formule 2B). Théoriquement, on pourrait fournir une mesure de précision (à savoir l'estimation de la variance d'échantillonnage) pour 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 comptant 1 000 personnes, l'erreur-type non ajustée serait de 25.

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 :

$$ETNA = \sqrt{\frac{4E(N - E)}{N}}$$

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 :

$$\sqrt{\frac{4(750)(9\,000 - 750)}{9\,000}} = 52$$

<sup>5</sup> Malheureusement, la variance d'échantillonnage ne nous fournit aucune indication de l'importance de l'erreur non due à l'échantillonnage.

Le tableau 9.2 fournit des facteurs d'ajustement<sup>6</sup> 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 caractéristiques<sup>7</sup> énumérées au tableau 9.2. Ce calcul a été fait pour chaque RP dans l'échantillon. 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 les catégories. 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, 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 caractéristiques... ». 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 infra-provinciale, il faut choisir la colonne « Facteur national ou provincial ». Au tableau 9.2, les facteurs d'ajustement associés à certaines 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 qu'étant donné qu'aucun échantillonnage n'a été fait dans les Territoires du Nord-Ouest et au Nunavut, les facteurs d'ajustement de toutes les caractéristiques dans ces territoires doivent correspondre à zéro. Comme on a eu recours à l'échantillonnage au Yukon, 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 plus petite qu'une province, 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 caractéristiques. N % des facteurs d'ajustement au niveau des RP sont inférieurs au N<sup>e</sup> centile et 100 – N % d'entre eux sont supérieurs au N<sup>e</sup> centile. Par exemple, 90 % des facteurs d'ajustement au niveau des RP sont inférieurs au 90<sup>e</sup> 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<sup>e</sup> centile fournirait une estimation très prudente, et le 75<sup>e</sup> 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.

<sup>6</sup> On appelle communément « effet du plan » le carré des facteurs d'ajustement.

<sup>7</sup> Par exemple, « 10 000 \$ – 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 ».

- d) Les facteurs d'ajustement de l'erreur-type ne s'appliquent pas aux montants d'argent, par exemple, mais aux estimations du nombre de personnes, de ménages, de logements ou de familles. Ainsi, les facteurs d'ajustement du revenu du ménage s'appliquent aux estimations du nombre de ménages dont le revenu se situe dans un intervalle donné, 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 1991 et 2001. L'estimation du recensement de 2001 pour cette caractéristique était 1 830 680. Le chiffre de population obtenu pour le Canada au recensement de 2001 était 29 639 030. 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 2 621. D'après le tableau 9.2, le facteur d'ajustement au niveau national pour la caractéristique « période d'immigration » après 1980 est 1,88. Il s'ensuit que l'erreur-type ajustée pour cette estimation est  $2\,621 \times 1,88 = 4\,928$ .

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 tiré pour une enquête donnée 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 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 de façon à produire des intervalles de confiance de 95 % d'une assez bonne précision. Au niveau infra-provincial, 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 de 95 % pour le nombre de personnes ayant immigré au Canada entre 1991 et 2001 serait  $1\,830\,680 \pm 2(4\,928)$  ou  $1\,830\,680 \pm 9\,856$ .

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

Total estimé de personnes, de ménages ou de logements dans la région pour la caractéristique étudiée	Nombre total 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

**Tableau 9.2 : Facteurs d'ajustement de l'erreur-type au niveau national ou provincial et pour les RP**

Caractéristiques	Facteurs au niveau national ou provincial	Centiles des facteurs au niveau de la RP						
		1 <sup>er</sup>	25 <sup>e</sup>	50 <sup>e</sup>	75 <sup>e</sup>	90 <sup>e</sup>	95 <sup>e</sup>	99 <sup>e</sup>
<b>Caractéristiques de la population</b>								
<b>Âge</b>								
Groupes d'âge 0 à 4, 5 à 9, 10 à 14, 15 à 19, 20 à 24, 25 à 29	0,15	0,00	0,00	0,04	0,17	0,32	0,53	1,35
Groupes d'âge 30 à 34, 35 à 44, 45 à 54, 55 à 59, 60 à 64	0,10	0,00	0,00	0,00	0,10	0,16	0,20	0,40
Groupes d'âge 65+	0,11	0,00	0,00	0,00	0,18	0,34	0,60	1,28
<b>Sexe</b>	0,07	0,00	0,01	0,05	0,09	0,12	0,14	0,19
<b>Union libre</b>								
Vit en union libre	0,46	0,00	0,00	0,21	0,47	0,77	1,36	2,17
Ne vit pas en union libre	0,36	0,00	0,00	0,09	0,34	0,65	1,00	1,91
<b>État matrimonial</b>								
Célibataires, personnes mariées (sauf les personnes séparées)	0,07	0,00	0,00	0,04	0,07	0,11	0,14	0,19
Personnes séparées, divorcées, veufs et veuves	0,14	0,00	0,00	0,00	0,22	0,41	0,57	1,31
<b>Plus haut niveau de scolarité</b>	1,20	0,64	1,09	1,20	1,30	1,41	1,49	1,73
<b>Plus haut grade, certificat ou diplôme</b>	1,18	0,63	1,08	1,19	1,29	1,39	1,48	1,75
<b>Principal domaine d'études</b>	1,18	0,83	1,11	1,19	1,28	1,37	1,44	1,63
<b>Lieu de naissance</b>								
Né(e) au Canada	1,29	0,29	1,23	1,40	1,54	1,68	1,76	1,98
Né(e) à l'extérieur du Canada	1,00	0,59	1,00	1,11	1,21	1,31	1,39	1,60
<b>Citoyenneté</b>								
Canada, par naissance	1,21	0,00	1,14	1,37	1,61	1,90	2,15	2,78
Autre	1,48	0,60	1,27	1,47	1,71	1,96	2,13	2,60
<b>Nombre de citoyennetés</b>								
Citoyenneté canadienne seulement	1,23	0,29	1,17	1,35	1,51	1,66	1,76	2,00
Une ou deux autres citoyennetés	1,68	0,02	1,27	1,53	1,81	2,07	2,28	2,86

Caractéristiques	Facteurs au niveau national ou provincial	Centiles des facteurs au niveau de la RP						
		1 <sup>er</sup>	25 <sup>e</sup>	50 <sup>e</sup>	75 <sup>e</sup>	90 <sup>e</sup>	95 <sup>e</sup>	99 <sup>e</sup>
<b>Période d'immigration</b>								
Avant 1950, 1951-1960, 1961-1970, 1971-1980	1,36	0,73	1,14	1,28	1,43	1,58	1,69	2,00
1981-1990, 1991-1995, 1996-2001	1,88	0,71	1,42	1,74	2,02	2,30	2,51	3,09
<b>Âge à l'immigration</b>								
	1,24	0,81	1,19	1,30	1,42	1,54	1,64	1,96
<b>Mobilité (1 an auparavant)</b>								
Personnes n'ayant pas déménagé	1,68	0,56	1,43	1,69	1,92	2,11	2,21	2,44
Personnes ayant déménagé (migrants, non-migrants)	1,79	0,44	1,33	1,66	1,94	2,17	2,32	2,66
<b>Mobilité (5 ans auparavant)</b>								
Personnes n'ayant pas déménagé	1,66	0,67	1,47	1,69	1,87	2,03	2,13	2,31
Personnes ayant déménagé (migrants, non-migrants)	1,78	0,60	1,54	1,79	2,01	2,23	2,38	2,90
<b>Population immigrante/non immigrante</b>								
Population immigrante	1,29	0,31	1,24	1,42	1,63	1,96	2,25	2,98
Population non immigrante	1,17	0,00	1,01	1,26	1,44	1,59	1,68	1,92
<b>Minorité visible</b>								
Chinois, Asiatique, Noir, Philippin, Latino-Américain, Arabe, Coréen, Japonais, Minorités visibles n.i.a., Minorités visibles multiples	2,16	0,00	1,38	1,87	2,34	2,77	3,09	4,02
Autochtone	1,41	0,37	1,30	1,60	1,91	2,19	2,39	3,01
Autre	1,52	0,00	1,20	1,63	1,97	2,22	2,35	2,65
<b>Origine ethnique</b>								
Français, anglais	1,51	0,00	1,20	1,62	1,96	2,20	2,34	2,62
Autre	1,99	0,00	1,23	1,64	2,17	2,63	2,95	3,85
<b>Confession religieuse</b>								
	1,69	0,00	1,19	1,59	1,97	2,37	2,67	3,45
<b>Langue parlée à la maison – Anglais</b>								
Nouveau-Brunswick, Colombie-Britannique, Ontario, Alberta	1,63	0,15	1,33	1,69	1,96	2,17	2,30	2,62
Québec	1,63	0,73	1,53	1,83	2,06	2,26	2,44	2,86
Autres provinces	1,16	0,00	0,71	1,27	1,73	2,06	2,21	2,65
Canada	1,15	-	-	-	-	-	-	-
<b>Langue parlée à la maison – Français</b>								
Nouvelle-Écosse, Québec	1,26	0,00	0,90	1,41	1,77	2,02	2,17	2,68

Caractéristiques	Facteurs au niveau national ou provincial	Centiles des facteurs au niveau de la RP						
		1 <sup>er</sup>	25 <sup>e</sup>	50 <sup>e</sup>	75 <sup>e</sup>	90 <sup>e</sup>	95 <sup>e</sup>	99 <sup>e</sup>
Nouveau-Brunswick	0,94	0,38	1,28	1,61	1,84	2,11	2,35	2,47
Autres provinces	1,59	0,07	1,31	1,64	1,96	2,28	2,55	3,42
Canada	0,74	-	-	-	-	-	-	-
<b>Première langue officielle parlée – Anglais</b>								
Québec	1,48	0,71	1,32	1,60	1,84	2,03	2,13	2,53
Terre-Neuve	0,69	0,00	0,23	0,54	0,94	1,27	1,53	2,22
Autres provinces	1,24	0,12	0,94	1,24	1,50	1,72	1,85	2,18
Canada	0,82	-	-	-	-	-	-	-
<b>Première langue officielle parlée – Français</b>								
Nouveau-Brunswick	0,93	0,40	1,15	1,41	1,61	1,84	1,97	2,15
Autres provinces	1,27	0,17	1,17	1,39	1,65	1,90	2,05	2,53
Canada	0,79	-	-	-	-	-	-	-
<b>Première langue officielle parlée – Autre</b>								
Français et anglais	1,69	0,00	1,23	1,51	1,79	2,07	2,30	2,87
Ni français ni anglais	1,49	0,00	1,21	1,42	1,65	1,90	2,14	2,77
<b>Langue officielle parlée – Anglais</b>								
Québec	1,40	0,00	1,26	1,48	1,67	1,89	2,10	2,56
Autres provinces	1,30	0,34	1,12	1,36	1,54	1,69	1,78	1,97
Canada	0,86	-	-	-	-	-	-	-
<b>Langue officielle parlée – Français</b>								
Nouveau-Brunswick, Québec	1,16	0,49	1,15	1,34	1,47	1,59	1,67	1,89
Autres provinces	1,46	0,00	1,05	1,30	1,57	1,93	2,26	3,09
Canada	0,87	-	-	-	-	-	-	-
<b>Langue officielle parlée – Autre</b>								
Français et anglais	1,29	0,62	1,26	1,42	1,57	1,71	1,80	1,99
Ni français ni anglais	1,49	0,00	1,20	1,42	1,64	1,89	2,12	2,75
<b>Langue maternelle – Anglais</b>								
Ontario, Alberta, Colombie-Britannique	1,37	0,00	1,12	1,40	1,65	1,84	1,95	2,12
Québec	1,45	0,60	1,26	1,51	1,77	1,95	2,07	2,44
Autres provinces	1,04	0,00	0,80	1,09	1,37	1,63	1,77	2,09
Canada	1,08	-	-	-	-	-	-	-
<b>Langue maternelle – Français</b>								

Caractéristiques	Facteurs au niveau national ou provincial	Centiles des facteurs au niveau de la RP						
		1 <sup>er</sup>	25 <sup>e</sup>	50 <sup>e</sup>	75 <sup>e</sup>	90 <sup>e</sup>	95 <sup>e</sup>	99 <sup>e</sup>
Nouveau-Brunswick	0,83	0,12	1,11	1,28	1,55	1,82	1,96	2,24
Québec	1,09	0,00	0,81	1,19	1,52	1,76	1,90	2,21
Autres provinces	1,35	0,63	1,22	1,40	1,62	1,86	2,02	2,45
Canada	0,72	-	-	-	-	-	-	-
<b>Langue maternelle – Autre</b>	1,84	0,00	1,08	1,40	1,90	2,41	2,75	3,62
<b>Langue de travail – Anglais</b>								
Québec	1,16	0,71	1,14	1,26	1,37	1,48	1,55	1,73
Autres provinces	0,78	0,19	0,67	0,77	0,88	1,01	1,11	1,32
Canada	0,76	-	-	-	-	-	-	-
<b>Langue de travail – Français</b>								
Nouveau-Brunswick, Québec	0,88	0,38	0,76	0,88	1,06	1,27	1,36	1,54
Autres provinces	1,25	0,00	1,07	1,23	1,41	1,64	1,84	2,44
Canada	0,73	-	-	-	-	-	-	-
<b>Langue de travail – Autre</b>	1,14	0,00	0,85	1,10	1,43	1,76	2,02	2,67
<b>Industrie</b>	1,44	0,65	1,08	1,21	1,32	1,43	1,52	1,75
<b>Profession</b>	1,07	0,74	0,97	1,14	1,26	1,38	1,46	1,70
<b>Travail en 2000</b>	1,10	0,68	1,10	1,21	1,31	1,41	1,48	1,65
<b>Semaines travaillées en 2000</b>	1,04	0,63	1,07	1,19	1,29	1,38	1,44	1,58
<b>Heures travaillées pendant la semaine de référence</b>	1,39	0,75	1,11	1,19	1,27	1,35	1,40	1,54
<b>Travail à plein temps / à temps partiel</b>								
Travail à plein temps	0,82	0,71	0,56	0,92	1,00	1,08	1,13	1,26
Travail à temps partiel	1,12	0,96	1,11	1,18	1,25	1,32	1,38	1,51
<b>Année du dernier travail</b>								
En 2001, en 2000, avant 2000	0,86	0,39	0,81	0,97	1,15	1,28	1,35	1,49
N'a jamais travaillé	1,24	0,69	1,13	1,25	1,36	1,46	1,54	1,75
<b>Catégorie de travailleurs</b>								
Travailleurs rémunérés	0,87	0,39	0,85	1,04	1,29	1,47	1,59	1,86

Caractéristiques	Facteurs au niveau national ou provincial	Centiles des facteurs au niveau de la RP						
		1 <sup>er</sup>	25 <sup>e</sup>	50 <sup>e</sup>	75 <sup>e</sup>	90 <sup>e</sup>	95 <sup>e</sup>	99 <sup>e</sup>
Travailleurs autonomes, familiaux non rémunérés, ou ayant une entreprise non constituée en société	1,25	0,70	1,11	1,23	1,35	1,50	1,63	2,00
<b>Travaux ménagers non rémunérés</b>	1,19	0,63	1,11	1,21	1,31	1,42	1,50	1,69
<b>Activité</b>								
Personnes occupées	0,90	0,00	0,87	1,05	1,22	1,36	1,45	1,74
Chômeurs	1,24	0,00	1,04	1,19	1,36	1,56	1,72	2,18
Inactifs	1,04	0,58	1,01	1,16	1,29	1,42	1,50	1,71
<b>Mode de transport pour se rendre au travail</b>								
En tant que conducteur, À pied, Transport en commun	1,01	0,58	1,00	1,16	1,29	1,41	1,49	1,74
Bicyclette, Motocyclette, En tant que passager, Taxi	1,23	0,45	1,06	1,19	1,32	1,46	1,59	1,96
Autre	1,21	0,36	1,04	1,18	1,35	1,54	1,71	2,21
<b>Lieu de travail – Provinces</b>	0,71	0,00	0,87	1,04	1,26	1,51	1,72	2,26
<b>Lieu de travail – Classification des secteurs statistiques (zone d'influence des RMR et des AR [ZIM])</b>								
ZIM forte ou modérée	1,00	0,13	1,04	1,18	1,31	1,50	1,66	2,14
ZIM faible ou sans influence	0,80	0,00	0,90	1,11	1,31	1,53	1,69	2,14
À l'intérieur d'une AR ou d'une RMR	0,84	0,37	0,93	1,08	1,24	1,41	1,55	1,98
Dans les territoires	0,31	0,00	0,72	0,87	1,02	1,23	1,37	1,86
<b>Lieu de travail – Catégorie de navettage</b>								
Dans la SDR de résidence	1,01	0,45	0,95	1,09	1,21	1,32	1,38	1,49
Dans une autre SDR de résidence	1,08	0,64	1,09	1,19	1,30	1,44	1,56	1,97
<b>Catégorie de lieu de travail</b>								
À domicile, sans adresse de travail fixe	1,25	0,69	1,16	1,26	1,36	1,47	1,54	1,72
En dehors du Canada	1,26	0,00	0,96	1,16	1,35	1,57	1,74	2,25
Lieu habituel de travail	0,94	0,41	0,85	0,94	1,03	1,12	1,17	1,28
<b>Nombre de personnes dans les intervalles de revenu total (\$)</b>								
0-9 999	0,70	0,00	0,55	0,70	0,82	0,94	1,01	1,17
10 000-19 999, 20 000-29 999, 30 000-39 999, 40 000-49 999, 50 000-59 999, 60 000-69 999, 70 000-79 999, 75 000 ou plus	1,14	0,89	1,09	1,15	1,22	1,28	1,33	1,45

Caractéristiques	Facteurs au niveau national ou provincial	Centiles des facteurs au niveau de la RP						
		1 <sup>er</sup>	25 <sup>e</sup>	50 <sup>e</sup>	75 <sup>e</sup>	90 <sup>e</sup>	95 <sup>e</sup>	99 <sup>e</sup>
<b>Situation des particuliers dans la famille de recensement</b>								
Conjoint, conjointe	0,11	0,00	0,06	0,09	0,13	0,18	0,21	0,28
Enfant	0,49	0,24	0,40	0,48	0,58	0,68	0,74	0,89
Parent seul de sexe féminin, parent seul de sexe masculin, personne ne faisant pas partie d'une famille de recensement	0,85	0,52	0,80	0,93	1,10	1,25	1,33	1,55
Autre	0,34	0,00	0,18	0,90	1,35	1,74	2,02	2,74
<b>Présence dans la famille de recensement</b>								
<b>Conjoint, conjointe, partenaire en union libre</b>								
Oui, conjoint ou conjointe	0,19	0,00	0,10	0,16	0,23	0,30	0,36	0,47
Oui, partenaire de même sexe	1,76	0,00	1,49	1,71	2,00	2,38	2,68	3,38
Oui, partenaire de sexe opposé	0,59	0,00	0,19	0,45	0,72	1,06	1,47	2,09
Non	1,28	0,65	1,07	1,22	1,38	1,59	1,75	2,13
<b>Situation des particuliers dans la famille économique</b>								
Conjoint, conjointe	0,26	0,04	0,19	0,27	0,57	1,27	1,49	2,03
Parent seul, enfant	0,63	0,33	0,65	0,84	1,04	1,18	1,28	1,51
Autres membres de la famille	1,35	0,61	1,10	1,24	1,43	1,64	1,80	2,24
Toutes les autres caractéristiques de la population	1,00	-	-	-	-	-	-	-
<b>Caractéristiques des ménages et des logements</b>								
<b>Mode d'occupation</b>	0,83	0,58	0,89	0,98	1,05	1,12	1,17	1,27
<b>Période de construction</b>	1,03	0,74	1,05	1,13	1,20	1,28	1,35	1,58
<b>Nombre de pièces</b>	1,09	0,77	1,04	1,11	1,18	1,24	1,29	1,44
<b>Nombre de chambres à coucher</b>	1,21	0,68	1,02	1,10	1,20	1,34	1,48	1,87
<b>Type de construction</b>								
Maison individuelle non attenante, maison en rangée, appartement dans un immeuble, habitation mobile, autre logement mobile	0,73	0,30	0,83	0,98	1,07	1,18	1,27	1,52
Maison jumelée, appartement ou plain-pied dans un	1,02	0,34	0,99	1,09	1,19	1,32	1,44	1,80

Caractéristiques	Facteurs au niveau national ou provincial	Centiles des facteurs au niveau de la RP						
		1 <sup>er</sup>	25 <sup>e</sup>	50 <sup>e</sup>	75 <sup>e</sup>	90 <sup>e</sup>	95 <sup>e</sup>	99 <sup>e</sup>
duplex non attenant, autre maison individuelle attenant								
<b>Taille du ménage</b>								
Ménage d'une personne	0,09	0,00	0,00	0,00	0,08	0,18	0,26	0,54
Autre	0,17	0,00	0,00	0,00	0,25	0,86	1,21	1,72
<b>Principal soutien du ménage</b>	0,00	-	-	-	-	-	-	-
<b>Âge du principal soutien du ménage</b>								
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	0,67	0,46	0,61	0,67	0,76	0,89	1,00	1,21
65+	0,47	0,31	0,42	0,48	0,55	0,65	0,74	0,93
<b>Sexe du principal soutien du ménage</b>	0,67	0,45	0,61	0,68	0,78	0,87	0,92	1,01
<b>Nombre de soutiens du ménage</b>								
Un seul soutien du ménage	1,17	0,78	0,98	1,04	1,09	1,14	1,17	1,23
Plus d'un soutien du ménage	1,18	0,00	1,01	1,11	1,28	1,51	1,71	2,36
<b>Personne repère soutien du ménage</b>	1,14	0,85	1,08	1,14	1,20	1,27	1,31	1,40
<b>Une personne qui ne vit pas ici est un soutien du ménage</b>	1,05	0,50	0,95	1,07	1,20	1,35	1,48	1,78
<b>Nombre de ménages dans les intervalles de loyer brut (intervalles de 100 \$)</b>	1,07	0,69	1,01	1,11	1,21	1,33	1,43	1,74
<b>Nombre de ménages dans les intervalles de loyer brut en pourcentage du revenu du ménage</b>								
Moins de 10 %	0,50	0,00	0,42	0,57	0,69	0,78	0,83	0,91
Entre 10-20 %, 20-30 %, 30-40 %, 40-50 %, plus de 50 %	0,87	0,59	0,83	0,88	0,94	1,00	1,05	1,21
<b>Nombre de ménages dans les intervalles de principales dépenses de propriété (intervalles de 200 \$)</b>	1,05	0,85	1,04	1,11	1,18	1,25	1,31	1,50
<b>Nombre de ménages dans les intervalles de principales dépenses de propriété en pourcentage du revenu du ménage</b>								
Entre 30-40 %	2,89	2,08	2,71	2,89	3,04	3,22	3,36	3,68

Caractéristiques	Facteurs au niveau national ou provincial	Centiles des facteurs au niveau de la RP						
		1 <sup>er</sup>	25 <sup>e</sup>	50 <sup>e</sup>	75 <sup>e</sup>	90 <sup>e</sup>	95 <sup>e</sup>	99 <sup>e</sup>
Moins de 10 %, Entre 10-20 %, 20-30 %, 40-50 %, plus de 50 %	0,88	0,65	0,85	0,89	0,94	1,00	1,05	1,19
<b>Personne responsable des dépenses du ménage</b>								
Le premier soutien	0,00	-	-	-	-	-	-	-
D'autres soutiens	0,90	0,00	0,90	1,07	1,29	1,52	1,72	2,38
<b>Nombre de ménages dans les intervalles de revenu du ménage (intervalles de 10 000 \$)</b>	1,05	0,69	1,02	1,10	1,17	1,23	1,27	1,36
<b>Nombre de ménages dans les intervalles de valeur du logement</b>	0,97	0,71	1,01	1,10	1,18	1,29	1,39	1,69
<b>Condominium enregistré</b>								
Fait partie d'un condominium enregistré	0,87	0,56	0,91	1,03	1,14	1,28	1,40	1,82
Ne fait pas partie d'un condominium enregistré	0,80	0,52	0,85	0,96	1,04	1,11	1,16	1,30
<b>État du logement</b>								
Entretien régulier, réparations majeures, et réparations mineures	0,88	0,73	0,86	0,90	0,94	0,97	1,00	1,07
<b>Toutes les autres caractéristiques des ménages et des logements</b>	1,00	-	-	-	-	-	-	-
<b>Caractéristiques de la famille de recensement</b>								
<b>Activité sur le marché du travail du conjoint, de la conjointe ou du parent seul</b>								
Conjoint occupé ou conjointe occupée	0,63	0,30	0,50	0,59	0,68	0,76	0,82	0,93
Parent seul occupé	1,29	0,72	0,97	1,07	1,19	1,32	1,42	1,69
<b>Groupes d'âge des enfants à la maison</b>	0,15	0,00	0,00	0,00	0,23	0,43	0,71	1,56
<b>Travail en 2000 du conjoint, de la conjointe ou du parent seul</b>								
A travaillé en 2000	0,85	0,48	0,70	0,79	0,88	0,97	1,04	1,16
N'a pas travaillé en 2000	0,79	0,54	0,69	0,74	0,80	0,86	0,90	1,01
<b>Toutes les autres caractéristiques de la famille de recensement</b>	1,00	-	-	-	-	-	-	-

Caractéristiques	Facteurs au niveau national ou provincial	Centiles des facteurs au niveau de la RP						
		1 <sup>er</sup>	25 <sup>e</sup>	50 <sup>e</sup>	75 <sup>e</sup>	90 <sup>e</sup>	95 <sup>e</sup>	99 <sup>e</sup>
<b>Caractéristiques de la famille économique</b>								
<b>Nombre de ménages dans les intervalles de revenu provenant d'un travail autonome (\$)</b>								
0-9 999	1,22	0,00	0,71	0,92	1,09	1,22	1,29	1,45
10 000-19 999, 20 000-29 999, 30 000-39 999, 40 000-49 999, 50 000-59 999, 60 000-69 999, 70 000-79 999, 75 000 ou plus	1,51	0,67	1,05	1,17	1,31	1,47	1,61	1,97
<b>Catégorie de faible revenu</b>								
Above line <sup>1</sup>	1,78	1,05	1,65	1,86	2,07	2,27	2,39	2,69
Below line <sup>2</sup>	1,90	1,34	1,76	1,94	2,13	2,32	2,44	2,72
<b>Taxes municipales et scolaires comprises dans les paiements hypothécaires</b>								
Taxes comprises, Taxes non comprises	1,07	0,91	1,05	1,11	1,16	1,22	1,26	1,37
<b>Salaires et traitements (\$)</b>								
0-1 999	0,75	0,00	0,61	0,77	0,91	1,02	1,08	1,18
2 000-4 999, 5 000-6 999, 7 000-9 999, 10 000-11 999, 12 000-14 999, 15 000-19 999, 20 000-24 000, 25 000-29 999, 30 000-34 999, 35 000-39 000, 40 000-44 999, 45 000-49 999, 50 000-59 999, 60 000-74 999, 75 000 ou plus	1,18	0,77	1,11	1,19	1,28	1,37	1,44	1,63
<b>Langue maternelle de la personne repère de la famille – Anglais</b>								
Terre-Neuve	0,16	0,00	0,08	0,13	0,18	0,25	0,31	0,50
Nouvelle-Écosse, Île-du-Prince-Édouard, Yukon	0,33	0,05	0,24	0,30	0,38	0,49	0,55	0,86
Québec	1,00	0,70	1,03	1,12	1,21	1,30	1,37	1,62
Autres provinces	0,61	0,29	0,50	0,61	0,74	0,87	0,94	1,08
Canada	0,59	-	-	-	-	-	-	-
<b>Langue maternelle de la personne repère de la famille – Français</b>								
Québec	0,44	0,00	0,20	0,37	0,62	0,85	0,97	1,13
Nouveau-Brunswick	0,62	0,13	0,63	1,02	1,12	1,19	1,24	1,35
Autres provinces	1,02	0,68	1,03	1,13	1,22	1,32	1,39	1,59
Canada	0,52	-	-	-	-	-	-	-

Caractéristiques	Facteurs au niveau national ou provincial	Centiles des facteurs au niveau de la RP						
		1 <sup>er</sup>	25 <sup>e</sup>	50 <sup>e</sup>	75 <sup>e</sup>	90 <sup>e</sup>	95 <sup>e</sup>	99 <sup>e</sup>
<b>Langue maternelle de la personne repère de la famille – Autre</b>	0,11	0,00	0,90	1,07	1,23	1,42	1,58	1,99
<b>Toutes les autres caractéristiques de la famille économique</b>	1,00	-	-	-	-	-	-	-

<sup>1</sup> Revenu non inférieur au seuil de faible revenu établi.

<sup>2</sup> Revenu inférieur au seuil de faible revenu établi.

## 10. Conclusion

L'échantillonnage fait maintenant partie intégrante du processus de recensement. 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 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 dépouillement 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 2001, 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 que ce biais ait été causé partiellement 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 recenseurs d'un échantillon biaisé. Les méthodes utilisées pour pondérer les données-échantillon en vue d'obtenir des estimations à l'échelle de la population totale ont été appliquées avec succès et, de façon générale, ont permis d'atteindre les niveaux prévus de cohérences des estimations-échantillon et des chiffres de population et, par le fait même, de compenser à certains biais observés dans l'échantillon. La cohérence obtenue aux niveaux provincial et national était meilleure qu'en 1996 pour la plupart des caractéristiques.

## 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 2001* (no 92-378-XIF au catalogue). L'utilisateur doit se reporter au *Dictionnaire du recensement de 2001* 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) :** L'aire de diffusion est une petite unité géographique relativement stable formée d'un ou de plusieurs îlots. 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) :** Terme général pour des régions créées en vertu des lois provinciales (comme les comtés, les municipalités régionales de comté et les *regional districts*) ou d'autres genres de régions. Les divisions de recensement sont des régions géographiques intermédiaires entre la municipalité (subdivision de recensement) et la province.

**État matrimonial :** Situation conjugale d'une personne. Voici les diverses catégories de réponse : marié et union libre; séparé, mais toujours légalement marié; divorcé; veuf ou veuve; jamais légalement marié (célibataire).

**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 hors famille de recensement, 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 (par exemple, 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 dénombrement (SD) :** Un secteur de dénombrement correspond à la région géographique dénombrée par un recenseur. Un SD est constitué d'un ou de plusieurs îlots adjacents. Tout le territoire du Canada est divisé en SD.

**Secteur de recensement (SR) :** Les secteurs de recensement 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.

**Subdivision de recensement (SDR) :** Terme générique qui désigne les municipalités (telles que définies par les lois provinciales) 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).

## **Annexe B. Contraintes au niveau des RP et des AD appliquées aux poids d'échantillonnage des recensements de 2001 et de 1996**

### **Contraintes au niveau des RP (personnes)**

- Total des personnes
- Total des personnes âgées de  $\geq 15$  ans
  
- Hommes
- Hommes âgés de  $\geq 15$  ans
  
- Personnes âgées de 0 à 4 ans
- Personnes âgées de 5 à 9 ans
- Personnes âgées de 10 à 14 ans
- Personnes âgées de 15 à 19 ans
- Personnes âgées de 20 à 24 ans
- Personnes âgées de 25 à 29 ans
- Personnes âgées de 30 à 34 ans
- Personnes âgées de 35 à 39 ans
- 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  $\geq 75$  ans
  
- Personnes mariées
- Personnes célibataires
- Personnes divorcées
- Veufs et veuves
- Personnes séparées\*\*
- Union libre = oui

### **Contraintes au niveau des RP (ménages)**

- Ménages de 1 personne
- Ménages de 2 personnes
- Ménages de 3 personnes
- Ménages de 4 personnes
- Ménages de 5 personnes
- Ménages de 6 personnes ou plus\*\*
- Total des ménages

**Contraintes au niveau des AD (constituaient des contraintes au niveau des SD en 1996)**

- Total des ménages dans l'AD
- Total des personnes dans l'AD

\*\* Caractéristiques non utilisées comme contraintes en 1996 parce qu'on les savait redondantes.

## Annexe C. Statistiques utilisées dans l'Étude du biais d'échantillonnage

On a vu au chapitre 6 que selon le mode d'échantillonnage aléatoire, la valeur

$$Z^{(0)} = \frac{\hat{X}^{(0)} - X}{\sqrt{V(\hat{X}^{(0)})}}$$

suivra une distribution à peu près normale (0,1). C'est ce que sert à démontrer la justification qui suit.

L'échantillonnage a été fait indépendamment pour chaque SD. Par conséquent,  $\hat{X}^{(0)}$  est la somme de  $H$  variables aléatoires indépendantes, où  $H$  est le nombre de SD au Canada. Comme il y a 35 883 SD échantillonnés au Canada,  $H$  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'avait pas une moyenne de 0 si les échantillons de ménages au niveau des SD présentaient un biais important, pour quelque raison que ce soit.

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 d'échantillonnage 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. On peut utiliser la statistique

$$W = \frac{rb(\hat{X}_1^{(0)}) - rb(\hat{X}_2^{(0)})}{\sqrt{\frac{1}{X_1^2} V(\hat{X}_1^{(0)}) + \frac{1}{X_2^2} V(\hat{X}_2^{(0)})}}$$

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.

## Annexe D. Produits et services du recensement de 2001

Le recensement constitue une source de données fiables pour décrire les caractéristiques des personnes et des logements au Canada. La gamme de produits et services élaborée à partir des données du recensement est conçue de façon que ces données soient utiles, compréhensibles et accessibles à tous les utilisateurs. D'autres sources, comme le *Catalogue du recensement de 2001*, le site Web de Statistique Canada ([www.statcan.ca](http://www.statcan.ca)) et, en particulier, le Catalogue en ligne, contiennent de l'information détaillée sur la gamme complète des produits et services du recensement de 2001.

Les produits et services comportent de nouvelles caractéristiques à l'occasion du recensement de 2001 :

### 1. Support

- On favorise Internet comme moyen pour diffuser les produits de données normalisés et les produits de référence.
- Davantage de données sont offertes gratuitement au grand public à partir d'Internet.

### 2. Contenu

- On diffuse les tableaux de données du recensement de 2001 selon des **thèmes**, c'est-à-dire des groupements de variables dont les sujets sont apparentés.
- Dans la mesure du possible, on a simplifié la langue et le vocabulaire dans les produits du recensement de 2001 diffusés dans Internet afin de rendre l'information accessible à un plus grand nombre de personnes.
- On a mis en place des moyens qui permettent aux utilisateurs de faire des recherches et de naviguer dans les **produits normalisés du recensement** (y compris les **produits de référence**) diffusés dans Internet.

### 3. Géographie

- Des unités géographiques, comme les aires de diffusion, les régions urbaines, les localités désignées et les zones d'influence métropolitaine, s'ajoutent à la gamme des produits normalisés. Certaines des nouvelles unités, comme les aires de diffusion, remplacent d'autres unités.

### 4. Variables

- Au recensement de 2001, on a recueilli de l'information sur de nouveaux sujets : le lieu de naissance des parents, les autres langues parlées à la maison et la langue de travail. Par ailleurs, la question sur la religion, qui est posée lors de chaque recensement décennal, figurait dans le questionnaire de 2001. En outre, on a élargi la définition de la variable portant sur la structure de la famille afin d'inclure les couples de même sexe.

## Bibliographie

- BANKIER, Michael Douglas. 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, William Gemmell. 1977. *Sampling Techniques*, 3<sup>e</sup> éd., Toronto, John Wiley and Sons.
- FELLEGI, Ivan P. 1964. « Response Variance and its Estimation », *Journal of the American Statistical Association*, vol. 59, p. 1016 à 1041.
- FULLER, Wayne A. 2002. « Estimation par régression appliquée à l'échantillonnage », *Techniques d'enquête*, vol. 28, n° 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 George, 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° 92-378-XIF au catalogue de Statistique Canada, Ottawa, « série des produits de référence du recensement de 2001 ».
- STATISTIQUE CANADA. 2002b. *Le recensement de 2001 en bref*, produit n° 92-379-XIF au catalogue de Statistique Canada, Ottawa, « série des produits de référence du recensement de 2001 ».