

N° 62F0026M au catalogue — N° 001

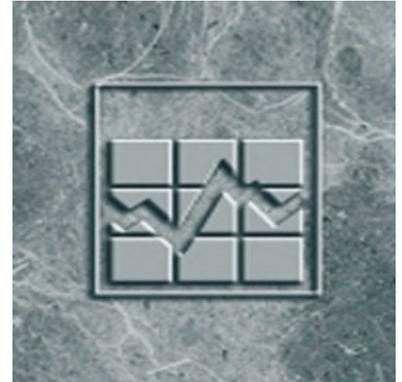
ISSN 1708-8887

ISBN 978-1-100-94412-8

Document de recherche

Série de documents de recherche sur les
dépenses des ménages

Enquête sur les dépenses des ménages de 2004 : indicateurs de qualité des données



Division des méthodes d'enquêtes auprès des ménages
Immeuble R.-H.-Coats, Ottawa, K1A 0T6

Téléphone: 613-951-7355



Statistique Canada
Statistics Canada

Canada

Comment obtenir d'autres renseignements

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

Centre de contact national de Statistique Canada

Numéros sans frais (Canada et États-Unis) :

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

Appels locaux ou internationaux :

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

Programme des services de dépôt

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

Comment accéder à ce produit

Le produit n° 62F0026M au catalogue est disponible gratuitement sous format électronique. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.gc.ca et de parcourir par « Ressource clé » > « Publications ».

Normes de service à la clientèle

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

Enquête sur les dépenses des ménages de 2004 : indicateurs de qualité des données

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2010

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

Avril 2010

N° 62F0026M au catalogue
ISSN 1708-8887
ISBN 978-1-100-94412-8

Périodicité : hors série

Ottawa

This publication is available in English upon request (catalogue no. 62F0026M).

Note de reconnaissance

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

Ce document a été préparé par :

Charles Mitchell
Sylvain Nadon

de la Division des méthodes d'enquêtes auprès des ménages (DMEM).

Les auteurs remercient l'équipe responsable de l'Enquête sur les dépenses des ménages de la Division de la statistique du revenu (DSR) pour leur collaboration dans la préparation de ce document.

Signes conventionnels

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

- . indisponible pour toute période de référence
- .. indisponible pour une période de référence précise
- ... n'ayant pas lieu de figurer
- 0 zéro absolu ou valeur arrondie à zéro
- 0^s valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
- ^P provisoire
- ^r révisé
- x confidentiel en vertu des dispositions de la *Loi sur la statistique*
- E à utiliser avec prudence
- F trop peu fiable pour être publié

Table des matières

Faits saillants	6
Introduction	9
1 Les erreurs d'échantillonnage.....	10
1.1 Les mesures de l'erreur d'échantillonnage.....	10
1.2 Les coefficients de variation.....	11
1.3 Modèle pour dériver une approximation du CV	13
1.4 Suppression des données non fiables dans les tableaux d'estimations	14
2 La non-réponse	15
2.1 Le taux de réponse, les taux de non-réponse et les taux de vacance	15
2.2 La non-réponse selon le niveau d'urbanisation	17
2.3 La non-réponse selon les strates de revenu	18
2.4 L'ajustement pour la non-réponse.....	20
3 Les erreurs de couverture	20
3.1 Le sous-dénombrement et le surdénombrement : les taux de glissement.....	21
3.2 L'ajustement au niveau de la population et des ménages	28
4 Les erreurs de réponse	29
5 Les erreurs de traitement.....	30
5.1 Proportion de ménages ou personnes nécessitant imputation, à l'échelle nationale et provinciale	32
Références.....	39
Annexe A Notation algébrique	40

Faits saillants

Les erreurs d'échantillonnage

- Les coefficients de variation (CV) de l'estimation de la moyenne des dépenses totales par ménage varient entre 1,3 % et 2,7 % selon les provinces. Le CV à l'échelle nationale est de 0,7 %.
- Les coefficients de variation (CV) de l'estimation de la moyenne pour les différentes catégories de dépenses sommaires sont pour la plupart inférieurs ou égaux à 1,9 % à l'échelle nationale, et généralement inférieurs à 5,0 % à l'échelle provinciale. Les résultats pour les différentes catégories de caractéristiques des logements et de l'équipement ménager sont similaires. La taille de l'échantillon étant moindre à l'Île-du-Prince-Édouard, les CV ont tendance à être plus élevés que ceux des autres provinces.

La non-réponse

- Le taux de réponse final est de 69,2 %. Les taux de réponse provinciaux s'échelonnent de 58,7 % en Ontario à 78,1 % en Saskatchewan.
- Le taux de non-réponse est de 30,8 %. La non-réponse est due aux refus (18,4 %), aux ménages qu'on n'a pas réussi à contacter (8,6 %) et aux ménages dont les données ont été considérées inutilisables (3,8 %).
- Le taux de non-réponse a tendance à croître avec le niveau d'urbanisation. On observe un taux de non-réponse de 24,3 % en milieu rural, et un taux de 33,7 % dans les centres urbains d'un million d'habitants et plus.
- L'analyse des taux de non-réponse dans les strates d'aires géographiques à revenu élevé provenant du plan d'échantillonnage révèle que le taux dans les strates à revenu élevé (38,9 %) est supérieur à celui observé dans les autres strates (29,9 %). Le taux de refus des strates à revenu élevé (25,3 %) est supérieur à ce qui est observé pour les autres strates (17,6 %).

Les erreurs de couverture

- On observe un sous-dénombrement des ménages de 8,3 % à l'échelle nationale. On constate un sous-dénombrement des ménages pour chacune des provinces, les taux variant de 4,1 % à 12,1 %. C'est en Colombie-Britannique que le sous-dénombrement des ménages est le plus élevé.
- Le sous-dénombrement des personnes se situe à 9,3 % à l'échelle nationale. On observe aussi un sous-dénombrement des personnes pour chacune des provinces, les taux variant de 4,1 % à 14,2 %. C'est en Colombie-Britannique que le sous-dénombrement des personnes est le plus élevé.
- On constate que le taux de glissement national pour les enfants (0 à 17 ans) est très différent de ceux des autres groupes d'âge. En effet, chez les enfants, on

observe un taux de sous-dénombrement de 0,1 %, alors que chez les adultes le sous-dénombrement est beaucoup plus important, se chiffrant à 11,9 %.

Les erreurs de réponse

- Les erreurs de réponse comprennent, entre autres, les erreurs de rappel, l'erreur télescopique et les erreurs dues aux réponses par procuration. Dans l'Enquête sur les dépenses des ménages, le fardeau imposé au répondant, étant donné la longueur de l'entrevue, peut entraîner la fatigue du répondant et avoir un impact sur la qualité des données. Le temps total d'entrevue varie selon les caractéristiques des ménages et peut être élevé pour certains ménages. Le temps moyen de l'entrevue était d'une heure et quarante-cinq minutes.

Les erreurs de traitement reliées à l'imputation

i) Des variables de dépenses

- Il a fallu imputer au moins une variable de dépenses¹ pour 31,1 % des ménages à l'échelle nationale. Ce taux plus élevé en 2004 est dû au changement apporté au questionnaire en ce qui concerne les dépenses reliées aux services de communication au foyer (téléphone, téléphone cellulaire et accès à Internet), aux services de télédistribution (câble) et aux services de diffusion par satellite. Désormais, pour ces services, on accepte qu'un ménage répondant fournisse seulement les dépenses totales d'un forfait (services groupés), tout en indiquant quels services sont inclus. Ceci a eu pour effet que 21,5 % des ménages utilisables ont nécessité l'imputation d'au moins un de ces cinq services.
- Le taux global d'imputation en excluant ces cinq services est de 13,2 % à l'échelle nationale, ce qui est comparable aux taux obtenus les années précédentes. Pour ces dépenses, la majorité des ménages ont nécessité l'imputation d'une ou deux variables parmi les 240 variables de dépenses. À l'échelle provinciale, les taux d'imputation varient entre un minimum de 6,2 % pour le Québec et un maximum de 18,8% pour la Nouvelle-Écosse.
- Les valeurs imputées contribuent à 20,4 % de l'estimation du total des dépenses de services de télédistribution et à 32,8 % de l'estimation du total des dépenses de services d'accès à Internet. Les taux d'imputation parmi les ménages déclarant des dépenses à l'un ou l'autre de ces deux services sont respectivement de 20,5 % et 40,3 %. Ces taux plus élevés sont principalement dus au fait qu'une proportion importante des forfaits incluait ces deux services parmi les ménages qui ont déclaré avoir payé pour un forfait.
- Près de 20 % des individus ont nécessité l'imputation de variables sur les dépenses vestimentaires mais pour la très grande majorité, le total était fourni par le répondant et seules les composantes ont été imputées.

1. Excluant les dépenses vestimentaires et les dépenses de la section sur les impôts personnels, sécurité et dons en argent qui sont collectées au niveau des individus et non au niveau des ménages.

- Près de 4 % des personnes âgées de 15 ans et plus ont nécessité l'imputation d'au moins une des variables de la section sur les impôts personnels, sécurité et dons en argent.

ii) Des variables de revenu

- Il a fallu imputer au moins une variable de revenu pour 5 % des individus âgés de 15 ans et plus. Pour 80 % d'entre eux, le revenu total était fourni par le répondant mais toutes les composantes (sources de revenu) ont été imputées.

iii) Des variables catégoriques

- Il a fallu imputer au moins une variable catégorique pour 7,6 % des ménages. Pour environ 72 % d'entre eux, seulement une variable a été imputée. À l'échelle provinciale, les taux d'imputation varient entre un minimum de 2,2 % pour le Nouveau-Brunswick et un maximum de 12,9 % pour l'Alberta. Les variables catégoriques qui sont imputées se trouvent aux sections suivantes du questionnaire: «Caractéristiques du logement», «Équipement associé au logement», «Mode d'occupation du logement» et «Tabac et dépenses diverses», pour les variables se rapportant aux achats par le biais de vente directe.

Introduction

L'Enquête sur les dépenses des ménages (EDM) est une enquête annuelle qui collecte les dépenses et les revenus des ménages à l'aide d'entrevues personnelles. L'échantillon de l'EDM 2004 est composé de 20 446 ménages² répartis dans les 10 provinces. La collecte est effectuée au cours des mois de janvier, février et mars et les revenus et dépenses sont obtenus pour la période du 1^{er} janvier au 31 décembre de l'année précédente. À la suite d'un remaniement en 1997, cette enquête remplace l'Enquête périodique sur les dépenses des familles et l'Enquête sur l'équipement ménager (avec ajustements aux questionnaires et aux échantillons).

L'EDM, comme toutes les enquêtes, est sujette aux erreurs et ce malgré toutes les précautions prises aux différentes étapes de l'enquête pour les contrôler. Quoiqu'il n'existe pas de mesure exhaustive de la qualité des données d'une enquête, certaines mesures de qualité produites aux différentes étapes de l'enquête peuvent fournir à l'utilisateur de l'information pertinente pour une bonne interprétation des données.

Ce document vise donc à présenter les indicateurs de qualité produits pour l'enquête sur les dépenses des ménages de 2004. On y trouve les indicateurs de qualité usuels généralement utiles aux utilisateurs pour l'interprétation des données tels que les coefficients de variation, les taux de réponse et de non-réponse, les taux de glissement et les taux d'imputation.

Les indicateurs de qualité ont été classés selon les principaux types d'erreur que l'on retrouve dans une enquête. Les erreurs d'échantillonnage, c'est-à-dire les erreurs dues au fait que les inférences tirées au sujet de l'ensemble de la population sont fondées sur des renseignements recueillis auprès d'un échantillon de la population et non pas auprès de toute la population, sont présentées dans la section 1. Les sections suivantes couvrent les erreurs non dues à l'échantillonnage. La non-réponse et les erreurs de couverture sont d'abord discutées dans les sections 2 et 3. Les erreurs de réponse et les erreurs de traitement sont traitées dans les sections 4 et 5 respectivement.

Ce document est axé sur la qualité des données. Pour une description détaillée de la méthodologie de l'enquête, on peut consulter la référence [1].

2 . Au départ, l'échantillon compte 23 897 logements desquels on doit identifier et enlever les logements inadmissibles (voir section 2.1) pour en arriver à 20 446 ménages auprès desquels on collecte les dépenses et les revenus.

1 Les erreurs d'échantillonnage

Les erreurs d'échantillonnage découlent du fait que les inférences que l'on tire au sujet de l'ensemble de la population à partir de l'enquête sont fondées sur l'information que l'on a recueillie auprès d'un échantillon de la population et non pas auprès de toute la population. Outre le plan de sondage et la méthode d'estimation qui ont été appliqués pour l'enquête sur les dépenses des ménages, la taille d'échantillon ainsi que la variabilité de chaque caractéristique sont des facteurs déterminants de l'erreur d'échantillonnage. Les caractéristiques qui sont rares ou qui sont distribuées de façon très différente dans la population auront une erreur d'échantillonnage plus grande que les caractéristiques qu'on observe plus fréquemment ou qui sont plus homogènes dans la population.

1.1 Les mesures de l'erreur d'échantillonnage

L'erreur-type est une mesure standard de l'erreur d'échantillonnage. L'erreur-type correspond au degré de variation de l'estimation considérant qu'un échantillon particulier a été choisi, plutôt qu'un autre, parmi tous les échantillons possibles de même taille sous le même plan de sondage. Comme l'EDM utilise un plan d'échantillonnage et une méthode d'estimation complexes, on estime l'erreur-type à partir d'une méthode de rééchantillonnage appelée le bootstrap. Jusqu'à l'année de référence 2002, l'EDM utilisait la méthode de rééchantillonnage du jackknife afin de produire les estimations de l'erreur type. Il a été décidé d'utiliser la méthode du bootstrap à compter de l'EDM de 2003 principalement en raison du fait que la Division de la statistique du revenu (DSR) allait diffuser les médianes des dépenses et requérait donc les coefficients de variation de celles-ci. La méthode du bootstrap permet l'estimation de la variance de statistiques non lisses comme les quantiles. Pour plus de détails sur cette méthode, voir la référence [2].

Le coefficient de variation (CV) est également une mesure de fiabilité de l'estimation fréquemment utilisée. Il exprime tout simplement l'erreur-type en pourcentage de l'estimation. Ainsi, si on obtient une estimation Y pour une certaine caractéristique et que SE correspond à l'erreur-type estimée, le CV sera $(SE/Y) \times 100$.

Finalement, l'erreur-type ou le coefficient de variation peuvent être utilisés pour dériver une autre mesure de précision des estimations, l'intervalle de confiance. Cette mesure indique le niveau de confiance selon lequel la valeur réelle dans la population pour une caractéristique qu'on a observée se trouve à l'intérieur de certaines limites. Un intervalle avec un niveau de confiance de 95 % correspond à l'estimation obtenue à partir de l'échantillon ± 2 erreur-type : $(Y \pm 2 SE)$.³ Cela signifie que si on répétait l'échantillonnage un grand nombre de fois, chaque échantillon fournirait un intervalle différent et 95 % des intervalles contiendraient la vraie valeur de la caractéristique. De façon similaire, si on répétait l'échantillonnage, l'intervalle $Y \pm SE$ contiendrait la vraie valeur dans 68 % des cas.

3. L'intervalle de confiance se calcule directement à partir du CV de façon similaire, soit $Y \pm 2 (CV \times Y)/100$.

1.2 Les coefficients de variation

Des estimations des coefficients de variation sont calculées pour les estimations de plusieurs caractéristiques collectées dans l'EDM. Les CV pour les estimations de la moyenne des dépenses par ménage (pour les catégories de dépenses détaillées), ainsi que pour les caractéristiques des logements et l'équipement ménager, sont disponibles à l'échelle nationale et provinciale dans la publication *Guide de l'utilisateur — Enquête sur les dépenses des ménages* (référence [3]).

On doit noter que les estimations de CV ne tiennent pas compte du fait que certaines données ont été imputées; par conséquent, les CV présentés peuvent sous-estimer les valeurs réelles. Pour la plupart des variables, le taux d'imputation est peu élevé (voir section 5) et les CV fournis représentent une bonne estimation des vrais CV. Cependant, il est important de tenir compte à la fois du CV et du taux d'imputation si on s'intéresse à la fiabilité de dépenses détaillées ayant un taux d'imputation élevé.

Le tableau 1.1 donne un aperçu des CV des estimations des moyennes par ménage à l'échelle provinciale ainsi qu'à l'échelle nationale pour l'estimation de quelques-unes des catégories de dépenses sommaires et pour le revenu.

Tableau 1.1
Coefficients de variation par province ainsi qu'à l'échelle nationale pour l'estimation de la moyenne des dépenses par ménage pour quelques catégories de dépenses sommaires et pour l'estimation du revenu moyen

Catégories de dépenses sommaires	Can.	T.-N.-L.	Î.-P.-É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sask.	Alb.	C.-B.
	%										
Dépenses totales	0,7	1,4	2,7	1,7	1,3	1,3	1,5	1,5	1,4	1,3	1,6
Dépenses de consommation	0,7	1,4	2,4	1,3	1,3	1,2	1,5	1,3	1,5	1,4	1,4
Alimentation	0,6	1,3	2,4	1,3	1,3	1,1	1,3	1,4	1,2	1,4	1,2
Logement	1,0	2,1	4,3	2,4	1,9	1,6	1,9	1,8	1,7	1,6	2,2
Entretien ménager	1,0	1,7	2,8	2,2	2,3	1,8	2,0	2,4	1,9	2,1	2,3
Ameublement	2,9	4,2	8,6	3,5	3,9	4,0	5,9	4,2	5,0	3,5	3,6
Habillement	1,3	2,1	3,7	2,9	3,1	2,1	2,7	2,9	2,6	2,7	2,9
Transport	1,6	3,8	5,0	2,8	3,4	2,8	3,1	3,2	3,4	4,1	3,2
Soins de santé	1,5	3,1	3,9	3,3	3,1	2,8	3,6	3,4	3,6	2,5	3,2
Soins personnels	1,3	1,9	3,5	2,3	2,8	2,7	2,5	2,8	2,5	3,1	2,6
Loisirs	1,9	4,1	8,0	3,7	3,2	4,2	3,8	4,5	3,9	3,8	3,3
Lecture et autres imprimés	1,8	3,7	6,5	4,2	4,5	3,1	3,9	3,1	3,5	4,1	4,0
Éducation	3,7	7,8	9,4	8,0	8,6	6,0	6,8	7,4	8,3	7,6	7,5
Tabac, boissons alcoolisées	2,0	4,2	3,9	4,0	4,8	3,5	4,2	4,7	4,6	3,7	3,9
Jeux de hasard (net)	5,1	7,3	18,4	8,8	7,6	15,4	7,4	8,1	7,5	9,9	6,4
Dépenses diverses	3,5	7,0	17,1	6,0	5,4	5,6	7,0	4,6	5,5	9,9	6,0
Impôts personnels	1,2	2,6	5,4	5,2	3,0	2,3	2,2	2,8	2,5	3,0	3,6
Assurances individuelles et cotisations de retraite	1,4	2,7	6,9	2,4	2,9	1,6	2,9	4,5	1,8	3,6	4,6
Dons et contributions	4,8	10,7	20,2	10,7	8,0	18,9	7,4	10,8	8,2	11,9	11,4
Revenu	0,7	1,2	2,5	1,2	1,5	1,2	1,4	1,3	1,5	2,0	1,6

Les coefficients de variation de l'estimation de la moyenne des dépenses totales par ménage varient entre 1,3 % et 2,7 % selon les provinces et il est de 0,7 % à l'échelle nationale.

Pour les différentes catégories de dépenses sommaires, les CV à l'échelle nationale sont inférieurs ou égaux à 1,9 % à l'exception des dépenses dans les catégories suivantes: ameublement, éducation, jeux de hasard, dépenses diverses ainsi que dons en argent et contributions. Ces catégories de dépenses représentent respectivement 2,9 %, 1,7 %, 0,4 %, 1,6% et 2,6 % du total des dépenses. De plus, si on fait exception de ces catégories, les CV sont généralement inférieurs ou égaux à 5 % à l'échelle provinciale. La taille de l'échantillon étant moindre à l'Île-du-Prince-Édouard, les CV ont tendance à être plus élevés que ceux des autres provinces.

Le tableau 1.2 donne un aperçu des CV pour les caractéristiques des logements et de l'équipement ménager à l'échelle provinciale ainsi qu'à l'échelle nationale pour l'estimation de quelques-unes des catégories.

Tableau 1.2
Coefficients de variation par province ainsi qu'à l'échelle nationale pour quelques catégories des caractéristiques des logements et l'équipement ménager

Catégories	Can.	T.-N.-L.	Î.-P.-É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sask.	Alb.	C.-B.
	%										
Propriétaire	1,0	1,6	3,0	2,4	2,0	2,2	2,0	2,7	1,8	2,2	2,0
Locataire	2,0	5,3	8,0	5,8	6,7	2,9	4,2	6,0	5,1	5,7	4,1
Machine à laver	0,7	1,1	1,9	1,7	1,1	1,1	1,7	2,1	1,1	1,7	1,8
Sècheuse	0,8	1,3	2,1	1,7	1,3	1,2	1,8	2,0	1,2	1,6	1,9
Lave-vaisselle	1,3	3,3	5,1	3,6	2,9	2,4	2,7	3,2	2,7	2,3	2,8
Congélateur	1,1	1,9	3,6	2,2	1,6	2,7	2,3	1,9	1,6	1,9	3,0
Four à micro-ondes	0,4	0,6	1,0	0,7	0,6	0,7	0,7	0,6	0,5	0,8	0,9
Téléphone cellulaire	1,0	2,2	4,7	2,4	2,5	2,4	1,7	2,5	2,4	2,1	2,0
Lecteur de disques compacts	0,7	1,6	2,2	1,3	1,4	1,4	1,3	1,6	1,5	1,3	1,3
Télédistribution	1,1	2,6	5,9	3,1	3,2	2,2	2,0	2,5	3,8	3,1	2,0
Antenne parabolique	2,7	6,8	7,8	6,3	5,5	4,3	6,0	7,0	5,4	5,8	9,1
Lecteur DVD	0,8	2,0	3,5	1,9	2,3	1,7	1,6	1,8	1,8	1,6	1,8
Ordinateur personnel	0,8	2,4	2,9	1,9	2,4	1,7	1,6	1,8	1,9	1,6	1,6
Connexion tél. régulière à un ordinateur (Modem)	3,3	7,2	6,8	9,3	5,9	5,6	6,4	11,8	9,9	10,7	9,4
Connexion tél. haute vitesse à un ordinateur	2,7	9,5	10,5	6,2	7,5	5,9	5,1	6,3	5,9	7,8	6,2
Connexion câble à un ordinateur	2,5	8,6	24,6	9,2	15,9	6,2	4,7	8,4	8,4	5,3	4,2
Utilisation d'Internet (maison)	1,0	3,0	3,7	2,5	3,1	2,1	1,9	2,4	2,3	1,8	1,9
Véhicules possédés (un)	1,5	3,4	4,7	3,4	3,5	2,9	2,9	3,4	3,7	3,9	2,9
Véhicules possédés (2 ou plus)	1,5	4,8	4,5	3,6	3,6	3,4	3,1	3,0	2,9	3,2	2,9

Les coefficients de variation pour les caractéristiques des logements et de l'équipement ménager sont généralement inférieurs à 4 % à l'échelle provinciale sauf pour les catégories : *locataire, antenne parabolique, connexion téléphonique*

régulière à un ordinateur (Modem), connexion téléphonique haute vitesse à un ordinateur et connexion câble à un ordinateur. L'Île-du-Prince-Édouard fait exception. La taille de l'échantillon étant moindre, les CV ont tendance à être plus élevés que ceux des autres provinces. Seule la province de Québec affiche un CV inférieur à 4% pour la catégorie locataire. C'est aussi dans cette province que l'on retrouve la proportion la plus élevée de locataires (43,4 %). Les CV des dépenses en connexion à un ordinateur varient selon le type de connexion. À l'Île-du-Prince-Édouard, le CV (6,8 %) est plus petit pour la connexion téléphonique régulière à un ordinateur, où dans ce cas, 30,0 % des ménages déclarent posséder ce type de connexion. En revanche, le CV (4,2 %) est plus petit en Colombie-Britannique pour la connexion câble à un ordinateur où 33,6 % des ménages de cette province déclarent posséder ce type de connexion.

Les CV pour les caractéristiques des logements et l'équipement ménager à l'échelle nationale sont inférieurs ou égaux à 2,0 % à l'exception des catégories suivantes : *antenne parabolique, connexion téléphonique régulière, connexion téléphonique haute vitesse à un ordinateur et connexion câble à un ordinateur.* On retrouve une proportion moindre de ménages possédant de tels équipements pour ces quatre catégories. Cette proportion représente respectivement 22,0 %, 15,7 %, 21,3 % et 22,0 % à l'échelle nationale.

1.3 Modèle pour dériver une approximation du CV

Des estimations pour différents domaines d'intérêt (par ex., par quintile de revenu) sont disponibles pour les catégories de dépenses sommaires dans la publication *Les habitudes de dépenses au Canada* (référence [4]). Des estimations pour différents domaines d'intérêt pour les catégories de dépenses détaillées sont également disponibles sur demande auprès de la Division de la statistique du revenu. (Pour plus de détails sur les tableaux disponibles auprès de la Division de la statistique du revenu, voir les références [3] ou [4].) Pour des raisons opérationnelles, il n'est pas possible de produire les CV pour tous les différents niveaux d'agrégation qui pourraient intéresser les utilisateurs.

1.3.1 Approximation du CV pour les estimations des domaines

Il est toutefois possible de calculer une approximation du CV en utilisant une relation entre le nombre de ménages de l'échantillon qui ont déclaré des dépenses pour une catégorie et le CV à un niveau agrégé. Cette relation, basée sur la tendance du CV à croître proportionnellement à une diminution de la racine carrée du nombre de ménages déclarant une dépense, est illustrée ci-après.

Formule d'approximation du CV pour un domaine (un sous-groupe de la population)

Si CV(Y) représente le CV pour l'estimation de la moyenne par ménage d'une certaine caractéristique pour toute la population, alors on peut calculer une approximation du CV de l'estimation de cette caractéristique pour un domaine (que l'on peut considérer comme un sous-groupe de la population tel qu'un type de ménage, un quintile de revenu, un niveau d'urbanisation) à partir de l'équation suivante :

$$CV(Y_d) = CV(Y) \times \sqrt{\frac{nP}{n_d P_d}}$$

où

- n : le nombre de ménages dans l'échantillon
- P : l'estimation de la proportion des ménages déclarant une valeur > 0 pour cette caractéristique dans la population
- n_d : le nombre de ménages de l'échantillon dans le domaine d
- P_d : l'estimation de la proportion des ménages déclarant une valeur > 0 pour cette caractéristique dans le domaine d

Généralement on utilise le CV, la taille n et la proportion P à l'échelle nationale pour calculer les approximations pour les différents domaines. Dans le cas où on cherche à calculer une approximation du CV pour un domaine entièrement contenu dans une seule province (par ex. région métropolitaine), il est préférable d'utiliser ces valeurs à l'échelle provinciale puisque les CV provinciaux sont publiés pour l'EDM de 2004 (référence [3]). Il est important de se rappeler que la valeur du CV obtenue par cette approche n'est qu'une approximation du CV.

1.3.2 Approximation du CV à partir du fichier de microdonnées

Les utilisateurs du fichier de microdonnées peuvent se servir d'une autre approche pour dériver une approximation du CV des estimations, qui sera généralement plus performante que celle présentée dans la section précédente pour les CV des catégories de dépenses détaillées. Cette approche est décrite en détails dans la documentation accompagnant le fichier de microdonnées de 2004. Elle est utilisable seulement à partir du fichier de microdonnées puisqu'il est nécessaire d'avoir les données et les poids de chaque ménage pour calculer cette approximation.

Le document sur la qualité des données de l'enquête de 1997 renferme des résultats de l'évaluation de la performance de ces deux méthodes d'approximation du CV.

1.4 Suppression des données non fiables dans les tableaux d'estimations

Comme le coefficient de variation est un indicateur de la fiabilité des données, on aimerait l'utiliser pour déterminer si les estimations devraient être publiées ou non. On considère que les estimations dont le CV est estimé à plus de 33 % ne sont pas suffisamment fiables pour être publiées. Toutefois les estimations des CV ne sont pas calculées pour une grande partie des estimations publiées, la règle de suppression pour les estimations de dépenses est donc basée sur le nombre de ménages qui déclarent une valeur supérieure à zéro.⁴

4. En pratique, on utilise l'estimation de la proportion des ménages déclarant une dépense que l'on multiplie par la taille d'échantillon.

On peut démontrer que les CV sont en général inférieurs à 33 % si le nombre de ménages qui déclarent une dépense est supérieur à 30. Comme il s'agit d'une règle approximative, certaines estimations peuvent être publiées même si le CV est supérieur à 33 % et certaines estimations ne seront pas publiées malgré un CV inférieur à 33 %. Le document sur la qualité des données de l'EDM 1997 donne les résultats de l'évaluation des risques d'erreurs de l'utilisation de la règle de suppression.

2 La non-réponse

Les erreurs dues à la non-réponse découlent du fait que certains répondants potentiels ne fournissent pas l'information nécessaire ou que cette information s'avère inutilisable. Lorsque le répondant a omis de répondre à certaines questions seulement, on parle alors de non-réponse partielle. Dans ce cas, les données manquantes sont imputées. Les erreurs liées à l'imputation sont présentées dans la section 5 portant sur les erreurs de traitement. Dans la présente section, la non-réponse englobe la non-réponse à la collecte, due principalement à l'impossibilité de contacter le ménage ou au refus des membres du ménage de participer à l'enquête, que ce soit partiellement ou complètement, ainsi que les ménages pour lesquels les données ont été collectées mais sont inutilisables.

Le principal impact de la non-réponse sur la qualité des données est qu'elle peut induire un biais dans les estimations si les caractéristiques des répondants et des non-répondants diffèrent et que cette différence a un impact sur les caractéristiques étudiées. Les taux de non-réponse peuvent être calculés facilement mais ils n'ont qu'une valeur indicative à l'égard de la qualité des données car ils ne permettent pas de mesurer l'importance du biais associé aux estimations. L'ampleur de la non-réponse peut être considérée comme une évaluation des risques de biais dans les estimations.

2.1 Le taux de réponse, les taux de non-réponse et les taux de vacance

Dans l'EDM, comme les unités sélectionnées sont des logements, les intervieweurs doivent d'abord identifier les logements inadmissibles, c'est-à-dire les logements occupés par des personnes ne faisant pas partie de la population cible, les logements qui n'existent plus (démoli, maison mobile déplacée ou logement converti en entreprise) et les logements vacants (inoccupés, saisonniers ou en construction).

Parmi les logements admissibles, on évalue ensuite la proportion des ménages qui n'ont pas répondu à l'enquête, qu'on appelle le taux de non-réponse à la collecte. Ces derniers comprennent les ménages qui ont refusé de participer à l'enquête et les ménages où il a été impossible d'établir un contact avec les répondants parce qu'ils étaient absents ou encore à cause de circonstances spéciales (problème de langue, maladie ou décès).

Toujours parmi les logements admissibles, on détermine également le taux de données inutilisables. Les données inutilisables correspondent au nombre de ménages pour lesquels le questionnaire était au moins partiellement complet mais qui ont été rejetés lors du traitement des données. Il existe deux causes principales de rejet. D'abord lorsqu'une partie importante des questions sur le revenu ou des

questions sur les dépenses a été laissée sans réponse, le questionnaire est classé incomplet et n'est pas utilisé. L'autre source de rejet correspond aux questionnaires pour lesquels la différence entre les entrées (revenu et autres sources d'argent du ménage) et les déboursés (dépenses et variation nette des actifs et passifs) est supérieure à 20 %. Ces questionnaires sont également exclus de l'estimation et considérés comme de la non-réponse.

Il est à noter que tous les taux de la présente section sont non pondérés. Pour l'Enquête sur les dépenses des ménages de 2004, le taux de réponse final est de 69,2 %. Le tableau 2.1-1 présente le taux de réponse final ainsi que la taille de l'échantillon (ménages admissibles) ventilée selon les refus, les non-contacts, les données inutilisables et les données utilisables. Ce taux est fourni à l'échelle nationale et provinciale.

Tableau 2.1-1
Taille de l'échantillon et taux de réponse par province ainsi qu'à l'échelle nationale

Province	Ménages admissibles	Non contactés	Ayant refusé de participer	Inutilisables	Utilisables	Taux de réponse final (à l'estimation) ¹
	nombre					%
Canada	20 446	1 751	3 759	782	14 154	69,2
Terre-Neuve-et-Labrador	1 829	170	206	92	1 361	74,4
Île-du-Prince-Édouard	778	67	98	24	589	75,7
Nouvelle-Écosse	2 080	204	401	151	1 324	63,7
Nouveau-Brunswick	1 861	182	316	114	1 249	67,1
Québec	2 645	172	526	32	1 915	72,4
Ontario	3 033	315	758	181	1 779	58,7
Manitoba	1 922	174	295	38	1 415	73,6
Saskatchewan	1 807	105	264	27	1 411	78,1
Alberta	2 038	166	385	30	1 457	71,5
Colombie-Britannique	2 453	196	510	93	1 654	67,4

1. Utilisables/ménages admissibles x 100

Le tableau 2.1-2 présente le taux de non-réponse final, le taux de non-réponse à la collecte, ventilé selon les refus et les non-contacts ainsi que le taux de ménages dont les données étaient inutilisables parce que le questionnaire était incomplet ou non équilibré. On y trouve également le taux de vacance. Ces taux sont fournis à l'échelle nationale et provinciale.

Notons que les taux de vacance présentés aux tableaux de la section 2 incluent les logements vacants (inoccupés, saisonniers ou en construction) ainsi que les logements qui n'existent plus (démoli, maison mobile déplacée ou logement converti en entreprise).

Tableau 2.1-2
Taux de non-réponse taux de vacance par province ainsi qu'à l'échelle nationale

Province	Taux de vacance	Taux de non-réponse à la collecte			Taux de données inutilisables			Taux de non-réponse final (à l'estimation)
		Total	Pas de contact	Refus	Total	Incomplet	Non équilibré	
%								
Canada	11,9	26,9	8,6	18,4	3,8	1,5	2,3	30,8
T.-N.-L.	14,4	20,6	9,3	11,3	5,0	0,9	4,1	25,6
Î.-P.-É.	19,2	21,2	8,6	12,6	3,1	1,0	2,1	24,3
N.-É.	12,9	29,1	9,8	19,3	7,3	1,0	6,3	36,3
N.-B.	14,1	26,8	9,8	17,0	6,1	1,2	4,9	32,9
Qc	9,4	26,4	6,5	19,9	1,2	1,2	0,0	27,6
Ont.	9,2	35,4	10,4	25,0	6,0	2,4	3,6	41,3
Man.	8,4	24,4	9,1	15,3	2,0	1,1	0,8	26,4
Sask.	14,0	20,4	5,8	14,6	1,5	1,2	0,3	21,9
Alb.	9,9	27,0	8,1	18,9	1,5	1,3	0,1	28,5
C.-B.	12,9	28,8	8,0	20,8	3,8	2,7	1,1	32,6

Le taux de non-réponse final au Canada est de 30,8 %. Il est dû aux refus (18,4 %), aux ménages qu'il a été impossible de rejoindre (8,6 %), et finalement aux ménages dont les données étaient inutilisables (3,8 %). Pour chacune des provinces, les refus représentent la cause majeure de non-réponse, suivie des non-contacts, puis des données inutilisables.

Les taux de non-réponse finaux varient selon les provinces. C'est en Saskatchewan qu'on observe le taux de non-réponse le plus bas, soit de 21,9 %. C'est également dans cette province que le taux de non-contacts est le plus faible (5,8 %). Par ailleurs, on remarque qu'au Québec il n'y avait pas de questionnaires non équilibrés. Les taux de non-réponse en Saskatchewan et à l'Île-du-Prince-Édouard sont inférieurs à 25 %, tandis que les taux sont supérieurs à 35 % en Nouvelle-Écosse et en Ontario. Le taux de non-réponse final plus élevé en Nouvelle-Écosse est en partie attribuable à un taux plus élevé de questionnaires non équilibrés (6,3 %). Le taux de non-réponse de l'Ontario est particulièrement élevé atteignant 41,3 %. C'est aussi dans cette province que l'on trouve le plus haut taux de non-contacts (10,4 %) et de refus (25,0 %).

Le taux de vacance est présenté au tableau 2.1-2, mais on doit considérer que les logements vacants ne contribuent pas au biais de l'échantillon dans la mesure où ils sont identifiés correctement. En analysant les taux de vacance, on décèle les problèmes d'identification des logements liés à la collecte. Le taux de vacance de l'EDM de 2004 est de 11,9 %.

2.2 La non-réponse selon le niveau d'urbanisation

La non-réponse varie selon le niveau d'urbanisation. Les divers taux à l'échelle nationale sont présentés par niveau d'urbanisation dans le tableau 2.2.⁵

5. Des tableaux portant sur les taux de non-réponse selon le niveau d'urbanisation par province sont disponibles sur demande auprès de la Division des méthodes d'enquêtes auprès des ménages.

Tableau 2.2
Taux de non-réponse et taux de vacance par niveau d'urbanisation

Catégorie d'urbanisation	Taux de vacance	Taux de non-réponse à la collecte			Taux de données Inutilisables			Taux de non-réponse final (à l'estimation)
		Total	Pas de contact	Refus	Total	Incomplet	Non équilibré	
Urbain		%						
1 000 000 ou plus	8,3	30,4	8,0	22,3	3,3	1,9	1,5	33,7
500 000 à 999 999	7,2	28,2	8,8	19,4	1,5	1,0	0,5	29,8
250 000 à 499 999	8,6	36,2	11,4	24,8	7,2	2,2	4,9	43,3
100 000 à 249 999	10,2	27,8	9,3	18,4	5,8	1,8	3,9	33,5
30 000 à 99 999	10,1	28,1	10,1	17,9	3,5	1,2	2,3	31,6
Moins de 30 000	11,7	22,3	7,3	15,0	3,0	1,3	1,7	25,3
Rural	22,2	20,4	7,1	13,3	3,9	1,2	2,6	24,3
Total	11,9	26,9	8,6	18,4	3,8	1,5	2,3	30,8

Le taux de non-réponse final augmente généralement avec le niveau d'urbanisation. C'est dans la catégorie d'urbanisation «250 000 à 499 999» que l'on retrouve le plus haut taux d'unités non contactés (11,4 %), de refus (24,8 %) et de données inutilisables (7,2 %). Le taux de non-réponse final de 43,3 % de cette catégorie d'urbanisation est dû à la fois à la non-réponse obtenue lors de la collecte (36,2 %), laquelle est due en partie à celle obtenue en Ontario (42,1 %, donnée non présentée), ainsi qu'au taux de données inutilisables (7,2 %), lequel est dû en partie à celui obtenu en Nouvelle-Écosse (8,5 %, donnée non présentée).

Le taux de non-réponse à la collecte a lui aussi tendance à croître avec le niveau d'urbanisation. Il y a un écart de près de 8 % entre les catégories d'urbanisation «Moins de 30 000» et «1 000 000 ou plus». Les refus comptent pour plus de 50 % de la non-réponse totale à tous les niveaux d'urbanisation.

En examinant les taux de vacance par niveau d'urbanisation, il ressort que le taux de vacance est près de deux fois plus élevé en région rurale (22,2 %) que dans les régions urbaines à faible population (11,7 %). Ces dernières régions ont aussi un taux de vacance supérieur à celui des régions urbaines à plus forte population. Ce phénomène s'observe également dans l'Enquête sur la population active (EPA) et s'explique sans doute par un plus grand nombre de logements saisonniers en milieu rural. Ceci explique entre autre pourquoi le taux de vacance est plus élevé dans les provinces de l'Atlantique, tel qu'illustré dans le tableau 2.1-2, et surtout à l'Île-du-Prince-Édouard puisqu'on y observe une proportion plus élevée de logements ruraux dans l'échantillon.

2.3 La non-réponse selon les strates de revenu

Il est impossible de comparer le taux de réponse selon le revenu car cette information n'est pas accessible pour les non-répondants. Toutefois, le plan d'échantillonnage de l'EPA, utilisé pour l'EDM, a été conçu de sorte à former, dans les régions métropolitaines de recensement, des strates d'aires géographiques présentant une plus forte concentration de ménages à revenu élevé. Bien que le nombre de strates à revenu élevé demeure relativement petit (51 sur un total de 1060 strates), la comparaison des taux de non-réponse dans ce groupe par rapport

à l'ensemble des autres strates fournit de l'information pertinente sur l'effet potentiel de la non-réponse.

L'échantillon de l'EDM 2004 a été sélectionné à partir du nouveau plan d'échantillonnage de l'EPA. Les changements qui ont été apportés à la stratification, plus particulièrement aux strates déterminées en fonction du revenu, ainsi que le fait qu'elles aient été redéfinies selon les données du Recensement de 2001 font en sorte que les taux de réponse selon les strates de revenu de l'EDM 2004 ne sont pas directement comparables à ceux des enquêtes précédentes.

Sous le nouveau plan, le nombre de strates à revenu élevé a été augmenté. Ces strates sont formées d'aires géographiques à plus forte concentration de ménages ayant un revenu supérieur à 125,000 \$. Elles comptent pour environ 5% de tous les ménages du Canada. Pour l'échantillon de l'EDM de 2004, 9,9 % des ménages admissibles provenaient de ces strates considérant que l'on suréchantillonne afin d'obtenir une meilleure représentation des ménages à revenu élevé dans l'échantillon.

Par ailleurs, selon le nouveau plan de sondage de l'EPA, il n'y a plus de base d'appartements et par conséquent de base d'immeubles d'appartements à faible revenu à partir de laquelle étaient formées les strates à faible revenu. Les immeubles d'appartements sont désormais intégrés au plan régulier de l'EPA. Ainsi, il n'y a donc plus de strates distinctes à faible revenu comme c'était le cas pour les années précédentes.

Le tableau 2.3 présente les taux de non-réponse et de vacance des strates à revenu élevé par rapport aux autres strates. Mentionnons que l'ensemble des autres strates comprend en plus des strates régulières, quatre nouveaux types de strates qui se sont ajoutés au nouveau plan d'échantillonnage de l'EPA. Ces types de strates sont les suivants : les strates à taux de vacance élevé, les strates à coût élevé, les strates ayant une concentration d'immigrants et les strates ayant une concentration d'autochtones. La portion de l'échantillon de l'EDM allouée pour ces quatre derniers types de strates étant plus faible, les résultats de ceux-ci ne sont pas présentés individuellement au tableau 2.3.

Tableau 2.3
Comparaison des taux de non-réponse et de vacance des strates à revenu élevé par rapport aux autres strates

Type de strate en fonction du revenu	Taux de vacance	Taux de non-réponse à la collecte			Taux de données inutilisables			Taux de non-réponse final (à l'estimation)
		Total	Pas de contact	Refus	Total	Incomplet	Non équilibré	
%								
Revenu élevé	5,8	34,8	9,5	25,3	4,1	1,8	2,3	38,9
Autres	12,5	26,1	8,5	17,6	3,8	1,5	2,3	29,9
Total	11,9	26,9	8,6	18,4	3,8	1,5	2,3	30,8

Dans les strates à revenu élevé, le taux de non-réponse final (38,9 %) est d'environ 30 % supérieur à celui des autres strates. Le taux de refus des strates à revenu élevé se situe à 25,3 %, ce qui est supérieur à ce qui est observé pour les autres

strates. Les ménages des strates à revenu élevé et des autres strates présentent toutefois des taux de données inutilisables semblables.

On remarque que le taux de vacance est plus faible pour les strates à revenu élevé que pour l'ensemble des autres strates. Ce phénomène était aussi observé pour les enquêtes précédentes.

2.4 L'ajustement pour la non-réponse

Pour compenser la non-réponse, les poids de l'EDM sont gonflés par l'inverse du taux de réponse pondéré à l'intérieur de certains groupes prédéfinis. À la suite du remaniement du plan de sondage de l'EPA, les groupes d'ajustement de non-réponse ont été remaniés. Tout comme pour les années précédentes, ces groupes sont définis en fonction des différents niveaux d'urbanisation dans chaque province et de régions géographiques infra-provinciales pour les provinces du Québec, de l'Ontario et de la Colombie-Britannique. De plus, des groupes spécifiques d'ajustement de non-réponse sont créés pour les strates à revenu élevé. Comme on l'a vu à la section précédente, le nombre de strates à revenu élevé a augmenté avec l'introduction du nouveau plan de sondage de l'EPA. Ainsi il est dorénavant possible de former dans toutes les provinces, à l'exception de l'Île-du-Prince-Édouard, des groupes d'ajustement de non-réponse pour les strates à revenu élevé. On ne peut former un tel groupe à l'Île-du-Prince-Édouard, puisque cette province ne possède pas de strates à revenu élevé.

Les taux pondérés diffèrent des taux présentés dans cette section puisqu'ils tiennent compte du poids de sondage de chaque ménage. Une description algébrique de l'ajustement pour la non-réponse est présentée à l'annexe A.

L'ajustement des poids pour la non-réponse permet de tenir compte des différences au niveau de la non-réponse par niveau d'urbanisation (tel qu'illustré dans la section 2.2) et régions géographiques ou par groupes de strates à revenu élevé. Il permettra de réduire le biais dans la mesure où les caractéristiques des répondants et des non-répondants sont similaires pour un même niveau d'urbanisation et région géographique ou encore pour un même groupe de strates à revenu élevé.

Il est à noter qu'un groupe d'ajustement de non-réponse peut être combiné à un autre groupe si le nombre de ménages est trop petit ou si le facteur d'ajustement est trop élevé.

3 Les erreurs de couverture

Lors de la conception de l'enquête, on a défini la population visée, qu'on appelle la population cible. Rappelons d'abord la définition de cette population cible pour l'EDM puisqu'une bonne compréhension de la population cible est nécessaire à une bonne interprétation des données de l'enquête. Il est important de préciser que l'EDM utilise la base de sondage de l'enquête sur la population active (EPA).

La population cible

La population cible correspond aux individus vivant dans les ménages privés. Les pensionnaires d'établissement institutionnel tel que les prisons, les hôpitaux pour malades chroniques, les résidences pour personnes âgées ainsi que les membres d'ordres religieux et d'autres groupes vivant en communauté, les membres des Forces Armées vivant dans les camps militaires et les individus vivant de façon permanente dans les hôtels ou les maisons de chambres sont donc exclus. On exclut aussi les représentants officiels de pays étrangers qui vivent au Canada et leurs familles ainsi que les résidents des réserves indiennes et des terres publiques (à l'exception des Territoires). L'enquête couvre donc près de 98 % de la population dans les 10 provinces. Les Territoires sont exclus de la population cible pour l'EDM de 2004, l'enquête ne couvrant cette région qu'une année sur deux.

Nous n'avons pas recueilli de renseignements auprès des personnes qui vivent temporairement loin de leur famille (par exemple, les étudiants d'université) parce que nous obtenons l'information auprès de leur famille, si elle est choisie dans l'échantillon.

Les erreurs de couverture proviennent d'une représentation inadéquate de la population cible à partir des unités de la base de sondage. Certaines unités de la population cible peuvent être omises de la base de sondage, on parle alors d'un sous-dénombrement. D'autres unités qui ne sont pas dans la population cible peuvent être incluses par erreur ou certaines unités peuvent être incluses plus d'une fois, ces unités sont responsables du surdénombrement.

3.1 Le sous-dénombrement et le surdénombrement : les taux de glissement

Dans l'EDM, la sélection de l'échantillon est effectuée en utilisant une liste de logements dans chaque grappe sélectionnée. L'omission de logements lors de la création de la liste, les nouveaux logements qui s'ajoutent entre la création de la liste et la visite des intervieweurs (principalement dans les secteurs en développement) ainsi que la classification erronée de logements vacants contribuent au sous-dénombrement. Le fait d'inclure des logements qui ne seraient pas dans les limites de la grappe est une source de surdénombrement. De façon similaire, des erreurs peuvent se glisser lors de la collecte des données, lors de l'identification des personnes qui sont membres du ménage sélectionné. Ces erreurs contribuent également au sous-dénombrement et au surdénombrement.

De plus, tel que décrit à la section 2.4, des méthodes de repondération sont mises en place pour tenir compte de la non-réponse. Lors de ces ajustements, il est cependant impossible de corriger les poids d'enquête pour faire en sorte que tous les sous-groupes de la population soient bien représentés.

Une bonne représentation de la population cible est indispensable à la production d'estimations de dépenses réalistes. Il est nécessaire que l'échantillon représente adéquatement les individus de la population cible et la distribution des ménages selon leur taille.

On observe généralement un sous-dénombrement net du nombre de personnes et du nombre de ménages dans l'EDM, que l'on rectifie par un ajustement des poids en utilisant des données auxiliaires, ou repères, qui sont basées sur des estimations démographiques post-censitaires. Le taux de glissement (voir annexe A) est une mesure du pourcentage d'écart entre les estimations provenant de ces données auxiliaires et les estimations de l'enquête calculées en utilisant des poids non ajustés avec ces données.⁶ Le glissement représente donc l'effet cumulé du sous-dénombrement et du débalancement dans certains sous-groupes de la population créé par la non-réponse à l'enquête qui n'a pu être corrigée à l'étape de la repondération.

Pour l'enquête de 2004, plusieurs changements ont été apportés à l'approche visant à ajuster les poids en utilisant des données auxiliaires. D'abord les estimations démographiques post-censitaires proviennent maintenant du Recensement de 2001 et non plus du Recensement de 1996, ce qui rend les taux de glissements de 2004 difficilement comparables à ceux des enquêtes précédentes. De plus, des modifications ont été apportées à la stratégie d'ajustement des poids [6]. Par exemple, dans les enquêtes précédentes les ajustements démographiques étaient effectués au niveau des croisements de neuf groupes d'âge selon le sexe, alors que la nouvelle stratégie se restreint à huit groupes d'âge. Des analyses ont démontré que le sexe ne semble pas être associé aux dépenses des ménages puisque les ménages sont en général constitués de personnes des deux sexes. Les ajustements selon le sexe ne contribuent donc pas à améliorer la qualité des estimations. Il est toutefois clair qu'il existe un effet du sexe sur les dépenses pour les ménages composés d'une seule personne. Par contre, l'absence de statistiques démographiques annuelles sur le nombre de ménages d'une seule personne, ventilé selon le sexe, ne nous permet pas de faire des ajustements pour ce cas particulier.

Taux de glissement par groupe d'âge et sexe à l'échelle nationale

Les taux de glissement par groupe d'âge et sexe à l'échelle nationale sont présentés dans le tableau 3.1-1. Un taux positif correspond à un surdénombrement du nombre de personnes dans l'enquête.

6. On utilise les sous-poids, c'est-à-dire les poids de sondage ajustés pour la non-réponse (voir Annexe A).

Tableau 3.1-1
Taux de glissement nationaux selon le groupe d'âge-sexe, Canada

Âge	Sexe		Total
	Hommes	Femmes	
	%		
0-6 ans	-2,9	2,8	-0,1
7-17 ans	-1,4	1,3	-0,1
18-24 ans	-14,8	-10,7	-12,8
25-34 ans	-17,1	-13,1	-15,1
35-44 ans	-13,9	-7,3	-10,6
45-54 ans	-10,3	-12,8	-11,6
55-64 ans	-15,1	-12,3	-13,7
65 ans et +	-9,0	-7,6	-8,2
Total	-10,8	-7,9	-9,3

Pour l'EDM de 2004, le taux de sous-dénombrement de la population était de 9,3 %. Les taux de glissement des enfants (0 à 6 ans et 7 à 17 ans) sont très différents de ceux des autres groupes d'âge. En effet, le taux de sous-dénombrement pour l'ensemble des enfants est de 0,1 %, alors qu'il est de 11,9 % pour les adultes (données non présentées). De plus, chez les filles de 0 à 17 ans on observe un léger surdénombrement. Pour les filles de 0 à 6 ans, ceci est dû aux surdénombrements obtenus en Ontario (4,3 %), en Alberta (13,0 %) et en Colombie-Britannique (14,9 %) pour ce groupe d'âge-sexe (voir le tableau 3.2-1). De la même façon, le surdénombrement chez les filles de 7 à 17 ans est dû à ceux obtenus au Québec (10,9 %) et en Alberta (16,7 %) pour ce groupe d'âge.

Les taux les plus élevés à l'échelle nationale sont observés chez les hommes de 18 à 24 ans, chez ceux de 25 à 34 ans, ainsi que chez les hommes de 55 à 64 ans. On remarque que le taux de sous-dénombrement chez les femmes est toujours inférieur à celui des hommes.

Tel que mentionné précédemment, l'EDM utilise la base de sondage de l'enquête sur la population active (EPA). Pour une même période, le taux de sous-dénombrement de l'EPA à l'échelle nationale est de 9,8 % (référence [5]), ce qui est légèrement inférieur au taux de 11,6 % de l'EDM observé lorsqu'on se restreint aux personnes de 15 ans et plus (données non présentées).

Puisque l'on n'ajuste plus les poids de sondage en fonction du sexe des personnes des ménages répondants, un sous-dénombrement ou surdénombrement net subsiste donc par groupe d'âge-sexe. Ce sous-dénombrement ou surdénombrement net est mesuré par le taux de glissement résiduel. Le taux de glissement résiduel est une mesure du pourcentage d'écart entre les estimations provenant des données auxiliaires et les estimations de l'enquête calculées en utilisant cette fois-ci les poids finaux. Les taux de glissement résiduels par groupe d'âge-sexe à l'échelle nationale sont présentés dans le tableau 3.1-2.

Tableau 3.1-2
Taux de glissement résiduels nationaux selon le groupe d'âge-sexe, Canada

Âge	Sexe		Total
	Hommes	Femmes	
	%		
0-6 ans	-1,8	1,9	0,0
7-17 ans	-1,5	1,5	0,0
18-24 ans	-0,8	0,9	0,0
25-34 ans	0,0	0,0	0,0
35-44 ans	-1,1	1,1	0,0
45-54 ans	1,0	-1,0	0,0
55-64 ans	-1,2	1,1	0,0
65 ans et +	-0,1	0,1	0,0
Total	-0,6	0,6	0,0

Les taux de glissement résiduels chez les hommes et les femmes sont nécessairement de signes opposés puisque les poids ont été ajustés pour correspondre aux estimations démographiques repères de chaque groupe d'âge. Bien que l'on ne tienne plus compte du sexe dans l'ajustement des poids, il demeure que de contrôler selon l'âge aide grandement à diminuer l'ampleur du sous-dénombrement par groupe d'âge-sexe observé au tableau 3.1-1. En effet, après avoir ajusté les poids pour chaque groupe d'âge, les taux de glissement résiduels sont inférieurs à 1,2 % pour les adultes et inférieurs à 1,9 % pour les enfants à l'échelle nationale.

Taux de glissement provinciaux par groupe d'âge et sexe

Les taux de glissement par groupe d'âge et sexe à l'échelle provinciale sont présentés dans le tableau 3.2-1.

Tableau 3.2-1
Taux de glissement pour les provinces selon le groupe d'âge-sexe

Taux de glissement par groupe d'âge-sexe										
Âge	Terre-Neuve-et-Labrador	Île-du-Prince-Édouard	Nouvelle-Écosse	Nouveau-Brunswick	Québec	Ontario	Manitoba	Saskatchewan	Alberta	Colombie-Britannique
Homme										
0-6	-8,2	12,0	2,7	-13,1	-1,2	-7,9	-12,2	-20,9	18,2	-0,5
7-17	-9,0	11,0	-7,9	3,5	8,0	-5,0	-8,0	-14,7	8,4	-8,9
18-24	-34,4	-29,6	-26,3	-22,4	0,0	-11,7	-25,0	-22,9	-23,3	-30,0
25-34	-20,2	-22,7	-13,9	-20,2	-16,8	-17,9	-23,3	-25,3	-4,4	-22,9
35-44	-19,4	-10,9	-19,2	-12,5	-3,8	-18,9	-13,1	-24,5	-2,8	-22,2
45-54	-21,2	-4,5	-17,0	-4,5	-8,8	-6,7	-13,6	-14,6	-14,3	-16,8
55-64	-11,5	-11,4	-7,4	-11,0	-19,6	-14,0	-6,7	-11,0	-23,4	-10,5
65 +	-2,2	-21,6	0,1	-13,1	-3,7	-10,7	-3,2	-9,8	-18,1	-11,1
Total	-16,1	-9,3	-11,9	-10,9	-6,1	-11,9	-13,1	-17,8	-6,8	-16,1
Femme										
0-6	-16,0	0,7	-1,3	4,0	-7,1	4,3	-6,8	-5,8	13,0	14,9
7-17	-3,5	-8,5	1,6	-4,2	10,9	-2,2	-5,9	-5,5	16,7	-12,2
18-24	-22,0	-7,1	-18,9	-26,5	1,5	-12,0	-9,9	-15,1	-13,9	-18,5
25-34	-8,6	-3,4	-8,1	-15,0	-8,5	-21,2	-17,4	-21,3	4,7	-9,6
35-44	-10,1	-0,2	-9,1	-12,2	-1,3	-9,2	-14,5	-11,9	4,7	-16,5
45-54	-22,2	3,6	-15,4	-5,2	-10,9	-13,6	-16,2	-12,3	-7,1	-17,2
55-64	-19,3	-12,3	11,1	-0,7	-19,2	-9,5	-6,3	-0,8	-21,0	-12,2
65 +	7,1	-20,1	-2,3	-19,6	-9,0	-4,4	0,8	-1,1	-16,5	-12,9
Total	-11,5	-6,1	-5,5	-10,4	-5,6	-9,2	-9,7	-9,2	-1,4	-12,2
Total	-13,8	-7,7	-8,6	-10,6	-5,8	-10,5	-11,3	-13,5	-4,1	-14,2

On observe un sous-dénombrement net pour chacune des provinces, les taux variant de 4,1 % à 14,2 %. C'est en Alberta que l'on retrouve le plus faible taux de sous-dénombrement (4,1 %). On doit mentionner qu'un faible taux de sous-dénombrement global n'est toutefois pas garant d'une meilleure couverture. Par exemple, le taux de glissement général observé en Alberta (4,1 %) cache le pire cas de surdénombrement à l'échelle provinciale des groupes d'âge-sexe (18,2 % chez les garçons) et les pires cas de sous-dénombrement pour le groupe d'âge des 55 à 64 ans.

À l'échelle provinciale, le taux de sous-dénombrement le plus élevé est observé chez les hommes de 18 à 24 ans à Terre-Neuve-et-Labrador (34,4 %). De plus, c'est chez les hommes de 18 à 24 ans que l'on retrouve les taux de sous-dénombrement les plus élevés, soit dans les provinces de l'Atlantique ainsi qu'en Colombie-Britannique. Bien que ce soit pour ce groupe d'âge-sexe que l'on observe les taux de sous-dénombrement les plus élevés, le taux le plus élevé à l'échelle nationale est plutôt observé chez les hommes de 25 à 34 ans (voir tableau 3.2-1). Ceci s'explique par le fait qu'au Québec, chez les hommes des 18 à 24 ans, il n'y a pas d'erreur de couverture (pour être plus précis on observe un taux de glissement de -0,014 %) avant l'ajustement des poids en fonction des groupes d'âge. Par ailleurs, c'est à l'Île-

du-Prince-Édouard que l'on observe les taux de glissement les plus élevés pour le groupe d'âge des 65 ans et plus. On remarque également que le schéma de variation des taux de glissement pour les groupes d'âge-sexe diffère sensiblement d'une province à l'autre.

Les taux de glissement résiduels par groupe d'âge-sexe à l'échelle provinciale sont présentés dans le tableau 3.2-2.

Tableau 3.2-2
Taux de glissement résiduel pour les provinces selon le groupe d'âge-sexe

Taux de glissement résiduel par groupe d'âge-sexe										
Âge	Terre-Neuve-et-Labrador	Île-du-Prince-Édouard	Nouvelle-Écosse	Nouveau-Brunswick	Québec	Ontario	Manitoba	Saskatchewan	Alberta	Colombie-Britannique
Homme										
0-6	0,9	5,8	1,9	-9,3	5,4	-5,3	-3,0	-10,3	4,1	-6,0
7-17	-4,8	10,2	-5,5	3,4	-2,0	-0,5	-1,3	-6,8	-5,3	1,0
18-24	-6,9	-11,6	-3,8	3,1	-0,8	2,5	-4,5	0,7	0,6	-10,1
25-34	-4,1	-6,5	-0,5	3,5	-3,8	4,3	-1,6	1,3	-2,3	-3,8
35-44	-4,9	-2,3	-4,2	2,7	2,7	-2,8	2,6	-2,4	-2,6	-1,4
45-54	1,2	-6,5	1,0	1,5	-0,7	4,3	1,2	-1,5	-5,5	0,7
55-64	5,8	-0,4	-8,3	-6,0	0,2	-2,2	0,4	-2,9	-0,7	0,5
65 +	-7,1	-0,5	1,3	4,1	5,3	-3,0	-1,4	-4,3	-1,9	1,1
Total	-2,5	-1,3	-2,5	1,0	0,4	0,0	-0,7	-3,1	-2,2	-1,7
Femme										
0-6	-0,9	-6,0	-2,0	9,8	-5,6	5,5	3,1	10,7	-4,3	6,3
7-17	5,1	-10,5	5,8	-3,6	2,1	0,5	1,4	7,3	5,6	-1,1
18-24	7,0	11,7	3,9	-3,2	0,8	-2,6	4,7	-0,8	-0,6	10,5
25-34	3,9	5,9	0,5	-3,3	3,9	-4,3	1,6	-1,3	2,4	3,7
35-44	4,7	2,2	4,0	-2,7	-2,8	2,8	-2,6	2,4	2,7	1,3
45-54	-1,2	6,1	-0,9	-1,5	0,7	-4,2	-1,2	1,5	5,7	-0,7
55-64	-5,7	0,4	8,0	6,0	-0,2	2,1	-0,4	2,9	0,7	-0,5
65 +	5,8	0,4	-1,0	-3,2	-4,0	2,4	1,1	3,4	1,6	-0,9
Total	2,4	1,3	2,3	-0,9	-0,4	0,0	0,7	3,1	2,3	1,7
Total	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

Tout comme pour les résultats observés à l'échelle nationale, l'ajustement des poids en fonction du groupe d'âge tend à atténuer l'ampleur du sous-dénombrement ou surdénombrement par groupe d'âge-sexe observé au tableau 3.2-1. Les taux de glissement résiduels les plus élevés sont de l'ordre de 10 % à 12 %. Ces taux se retrouvent parmi les enfants de 0 à 6 ans, chez ceux de 7 à 17 ans ainsi que pour les adultes de 18 à 24 ans. Tel que mentionné précédemment, des analyses ont démontré que le sexe ne semble pas associé aux dépenses des ménages, sauf peut-être pour les ménages composés d'une seule personne. Or, la presque totalité des personnes de ces trois derniers groupes d'âge ne constituent pas des ménages

d'une seule personne. Ainsi, bien que le taux de glissement résiduel soit plus élevé pour ces groupes d'âge-sexe, il n'y aurait vraisemblablement pas de biais dans les estimations découlant d'une moins bonne représentativité de ces groupes d'âge-sexe.

Notons par ailleurs que les taux de glissement globaux pour les femmes et les hommes de chaque province sont relativement faibles.

Taux de glissement selon la taille de ménage

Les taux de glissement selon les tailles de ménage que l'on utilise lors de l'ajustement des poids sont présentés dans le tableau 3.3. Un taux négatif correspond à un sous-dénombrement du nombre de ménages dans l'enquête.

Tableau 3.3
Taux de glissement pour les provinces selon la taille de ménage

Province	Taux de glissement			
	Ménages	Ménages d'une personne	Ménages de deux personnes	Ménages de trois personnes et plus
	%			
Canada	-8,3	-7,9	-9,0	-8,0
Terre-Neuve-et-Labrador	-11,7	-16,7	-5,8	-14,2
Île-du-Prince-Édouard	-6,2	-4,0	-4,6	-8,9
Nouvelle-Écosse	-6,3	-2,2	-4,6	-10,9
Nouveau-Brunswick	-10,6	-15,9	-11,0	-6,8
Québec	-4,9	-3,9	-4,4	-6,1
Ontario	-10,1	-8,7	-13,7	-8,2
Manitoba	-9,1	-5,0	-7,4	-13,9
Saskatchewan	-9,5	-9,3	-2,3	-16,5
Alberta	-4,1	-6,9	-4,5	-2,2
Colombie-Britannique	-12,1	-14,4	-11,6	-10,7

On observe une sous-estimation du nombre de ménages de 8,3 % à l'échelle nationale. Cette sous-estimation est légèrement inférieure à celle de 9,3% observée pour le nombre de personnes. De plus, on observe toujours un sous-dénombrement quelle que soit la taille de ménage, et ce, tant à l'échelle nationale que provinciale. À l'échelle nationale, il y a peu de différences dans le sous-dénombrement selon la taille de ménage, les taux variant de 7,9 % à 9,0 %.

À l'échelle provinciale, il y a une variation plus importante du taux de glissement. Cependant, l'écart observé est à peu près le même entre les tailles de ménage. En effet, pour les ménages d'une personne, le taux varie de -2,2 % pour la Nouvelle-Écosse jusqu'à -16,7 % pour Terre-Neuve-et-Labrador. Pour les ménages de deux personnes, les taux de glissement varient de -2,3 % pour la Saskatchewan à -13,7 % pour l'Ontario. Pour les ménages de trois personnes et plus, les taux de glissement varient de -2,2 % pour l'Alberta à -16,5 % pour la Saskatchewan.

À l'exception de l'Ontario, le sous-dénombrement le plus élevé est observé à peu près également entre les ménage d'une personne et les ménages de trois personnes et plus.

3.2 L'ajustement au niveau de la population et des ménages

Afin de corriger le problème de représentativité illustré dans les tableaux 3.1-1 et 3.2-1 et réduire le biais qui en découle, on ajuste les données de l'enquête lors de la pondération en utilisant les estimations démographiques pour les groupes d'âge définis dans ces tableaux, et ce pour chaque province. Pour plus de détails sur la méthodologie de l'ajustement, voir les références [1] et [6]. Cet ajustement réduit le biais mais ne l'élimine pas complètement si les caractéristiques des individus que l'on a omis de l'enquête diffèrent de celles des individus inclus pour un même groupe d'âge dans une province.

On notera également que l'efficacité de l'ajustement à partir des estimations démographiques dépend beaucoup de la qualité de ces estimations démographiques et de leur exactitude à bien représenter la population cible de l'enquête. Les estimations démographiques ne sont pas exemptes d'erreurs. Ce sont des estimations post-censitaires basées sur les comptes de population du Recensement de 2001, ajustées pour le sous-dénombrement net et qui tiennent compte des récentes statistiques sur la migration, les naissances, la mortalité, etc. Ces estimations démographiques sont ajustées pour tenir compte de certaines exclusions spécifiques aux enquêtes auprès des ménages comme les personnes vivant en institutions. Conceptuellement, elles diffèrent légèrement de la population cible de l'EDM en incluant les personnes habitant les ménages collectifs qui ne sont pas des institutions, par exemple les membres de groupes vivant en communauté et les individus vivant de façon permanente dans les hôtels ou les maisons de chambres. Cependant, on considère cette différence négligeable puisque ces individus représentent moins de 0,4 % de la population canadienne.

Pour corriger le problème de représentativité de l'échantillon par rapport au nombre de ménages selon leur taille, illustré dans le tableau 3.3, on ajuste les données de l'enquête à l'aide de données auxiliaires. En ajustant les poids de l'EDM pour refléter les estimations post-censitaires du nombre de ménages selon la taille, on vise à compenser le biais qui découle d'une représentation inadéquate des ménages. Toutefois on n'éliminera pas nécessairement ce biais si les caractéristiques des ménages non interviewés (omis ou non-répondants) diffèrent de celles des ménages répondants pour une même taille de ménage. Comme dans le cas des estimations démographiques de population, l'efficacité de l'ajustement dépend de la qualité des données auxiliaires sur le nombre de ménages.

En plus des estimations démographiques quant aux groupes d'âge par province, deux autres ensembles de données auxiliaires sont utilisés pour ajuster les données d'enquêtes lors de la pondération en vue d'en améliorer la représentativité. Le premier ensemble de données est utilisé pour contrôler le nombre d'enfants et d'adultes dans certaines grandes villes. Pour le second, des comptes par grandes catégories de revenu en salaires et traitements sont utilisés lors de l'ajustement des poids afin d'assurer une certaine cohérence entre les distributions de revenu de l'EDM et celles provenant de sources externes.

4 Les erreurs de réponse

Les erreurs de réponse correspondent au manque d'exactitude des réponses aux questions. Elles peuvent être attribuables à différents facteurs parmi lesquels se retrouvent un questionnaire qui nécessite des améliorations, une mauvaise interprétation des questions par les intervieweurs ou les répondants ainsi que des erreurs dans les déclarations des répondants.

Dans l'EDM, les erreurs dans les déclarations des répondants peuvent survenir pour différentes raisons. Premièrement, il y a les erreurs de rappel qui se produisent lorsque le répondant oublie les dépenses effectuées durant la période de l'enquête (qui correspond à l'année civile) ou fournit une valeur erronée à cause de l'intervalle de temps écoulé entre le moment de l'achat et la date de l'entrevue. Les erreurs de rappel sont probablement la source d'erreur de réponse la plus importante de l'enquête étant donné que la période de référence est longue (12 mois) et qu'une grande variété d'information est demandée.

Une des principales mesures utilisées pour minimiser l'erreur de rappel dans l'EDM est de calculer, pour chaque ménage, la différence entre les entrées d'argent (revenus et autres montants reçus par le ménage) et les déboursés (dépenses plus variation nette de l'actif et du passif). Lorsque la différence excède 15 % des entrées d'argent ou des déboursés, le montant le plus élevé étant retenu, les répondants sont contactés à nouveau pour obtenir des renseignements supplémentaires et tenter d'identifier les erreurs ou les omissions. On encourage également le répondant à consulter divers documents (factures, relevés bancaires,...) pour fournir des données plus exactes. Pour déterminer les dépenses pour de petits articles achetés à intervalles réguliers, les intervieweurs suggèrent généralement aux répondants d'estimer la fréquence des achats et le prix généralement payé pour dériver les dépenses sur une période de 12 mois.

Une seconde source d'erreur dans la déclaration des répondants est l'erreur télescopique qui consiste à inclure dans la période de référence des événements qui se sont produits avant ou après celle-ci. Dans l'EDM, on considère que le fait d'utiliser l'année civile est un bon repère pour le début de la période de référence. De plus, comme la période de référence est longue, l'impact de l'erreur télescopique est moins important.

Les réponses par procuration peuvent également contribuer à l'erreur de réponse. Le membre du ménage qui a effectué une dépense est généralement la personne qui peut la rapporter avec la meilleure précision. C'est sûrement le cas par exemple pour les achats personnels. Les dépenses rapportées par personne interposée sont plus susceptibles d'être entachées d'une erreur de réponse et ce type d'erreur risque d'affecter plus sérieusement certaines catégories de dépenses.

Parmi d'autres sources d'erreur de réponse, il ne faut pas oublier le degré de coopération du répondant. Il se peut que le répondant décide pour des raisons personnelles de ne pas mentionner certaines dépenses ou de déformer les faits.

Dans l'EDM, on considère également que le fardeau imposé au répondant, étant donné la longueur de l'entrevue et la grande variété d'articles à rapporter, ainsi que le rythme de l'entrevue peuvent entraîner la fatigue du répondant et avoir un impact sur la qualité des réponses obtenues. Le temps d'entrevue varie beaucoup d'un

ménage à l'autre selon la taille du ménage, le revenu et plusieurs autres caractéristiques. Pour certains ménages, l'entrevue peut prendre plus de cinq heures.

Quoique l'on considère que les erreurs de réponses sont une source importante d'erreur dans une entrevue rétrospective, elles demeurent l'aspect lié à la qualité des données le plus difficile à mesurer. Il faut généralement entreprendre des études spéciales très coûteuses pour tenter de les mesurer. On tente de les réduire en implantant des techniques de sondage développées à cet effet.

5 Les erreurs de traitement

Des erreurs peuvent se produire lors de toutes les manipulations des données. Les principales étapes du traitement des données sont le codage des réponses, la saisie des données, la vérification, l'imputation des non-réponses partielles et la pondération. Dans l'EDM, différentes procédures sont appliquées à chacune des étapes afin de réduire au minimum les erreurs de traitement et les estimations de l'enquête sont comparées avec d'autres sources avant d'être publiées. Les erreurs qui peuvent découler des ajustements effectués lors de la pondération ont été décrites dans les sections 2 et 3. La présente section couvre principalement les autres types d'erreurs de traitement.

Le codage des réponses est nécessaire pour très peu de questions. Ce travail est effectué par l'intervieweur, puis vérifié par l'intervieweur principal. Avant 2001, la saisie des données était vérifiée à l'aide d'un système de vérification automatisé qui consistait à grouper les questionnaires par lots et à choisir, dans chacun de ces lots, un échantillon de questionnaires qui seraient saisis une seconde fois. Toutes les erreurs relevées étaient corrigées. Si le nombre d'erreurs dépassait un certain seuil, le lot complet était de nouveau soumis à la saisie. Contrairement aux années précédentes aucune procédure de vérification des questionnaires par lot n'a été utilisée depuis 2001, à la suite de l'introduction d'un nouveau système de saisie des données (BLAISE). Toutefois, des règles de vérifications ont été implantées dans ce système afin de pouvoir assurer la cohérence des informations saisies. Les résultats préliminaires d'une étude sur la saisie semblent démontrer que les taux d'erreurs de saisie avec le nouveau système sont similaires à ceux avec l'ancien système.

Une première étape de vérification automatisée des questionnaires est effectuée après que chacun d'eux ait été vérifié manuellement par l'intervieweur et l'intervieweur principal. On s'assure que certaines règles essentielles de cohérence entre les réponses rapportées sur le questionnaire sont respectées. On identifie également les situations inhabituelles qui pourraient justifier des corrections. Cette étape de vérification automatisée est effectuée dans les bureaux régionaux de Statistique Canada, ce qui permet de communiquer avec les répondants lorsque des renseignements supplémentaires sont nécessaires pour résoudre des incohérences dans les réponses qu'ils ont fournies. Les problèmes identifiés au cours de cette vérification sont résolus par les membres des équipes de résolution des vérifications qui ont reçu une formation spéciale à cet égard. Par la suite, d'autres vérifications des données sont faites au bureau central et les réponses invalides sont corrigées.

Le traitement des données de l'EDM comprend également l'imputation de données pour la non-réponse partielle. Il y a non-réponse partielle lorsque le répondant refuse

de répondre ou ne connaît pas la réponse à certaines questions particulières. L'approche pour imputer les données diffère selon qu'il s'agit de données catégoriques ou continues. Les données catégoriques peuvent prendre uniquement quelques valeurs (comme les questions pour lesquelles la réponse est oui ou non et les questions sur le type de logement habité) alors que les données continues peuvent prendre n'importe quelle valeur numérique (comme les revenus et les dépenses).

Les données sur le revenu et les dépenses sont imputées au moyen de la technique du plus proche voisin. L'imputation se fait sur un groupe de variables à la fois, ce groupe étant formé de sorte à tenir compte des relations entre ces variables. Un groupe correspond généralement à une section du questionnaire. Pour chaque groupe, les valeurs manquantes d'un receveur (un ménage qui a des données manquantes pour au moins une de ces variables) sont imputées à partir des données du plus semblable parmi tous les donneurs (les ménages qui n'ont aucune donnée manquante pour ces variables). Pour chaque receveur, on détermine le donneur le plus semblable comme étant celui qui a la valeur minimale pour une certaine fonction de distance. Cette fonction est basée sur des variables d'appariement choisies parce qu'elles sont corrélées avec les variables à imputer. Par exemple, le revenu total du ménage fait partie des variables d'appariement pour presque toutes les sections portant sur les dépenses. Dans le choix du donneur, on tient compte également du fait que le receveur doit satisfaire certaines règles de cohérence après l'imputation de ces données manquantes. Pour la plupart des sections, l'imputation est effectuée au niveau du ménage mais dans certains cas, par exemple pour le revenu et les dépenses en vêtements, l'imputation est effectuée au niveau de la personne car c'est à ce niveau que l'on collecte l'information pour ces variables.

Notons que depuis 2001, l'imputation des données sur le revenu et les dépenses est effectuée à l'aide du Système canadien de contrôle et d'imputation du recensement (SCANCIR) de Statistique Canada. Le nouveau système utilise une méthodologie quelque peu différente de celle du système utilisé auparavant. Ce système permet une meilleure utilisation des données catégoriques lors de la sélection du donneur. De plus, ce système permet à la fois l'imputation des données continues et des données catégoriques. Des tests ont été effectués avant l'introduction du système et les résultats obtenus étaient semblables à ceux obtenus avec l'ancien système. À compter de 2003, les données catégoriques, qu'on trouve principalement dans les sections sur les caractéristiques et l'équipement du logement, sont imputées à l'aide du SCANCIR. Les données catégoriques étaient imputées auparavant à l'aide d'une technique "hot deck" où un ménage donneur est choisi de façon aléatoire parmi un groupe de ménages répondants possédant des caractéristiques semblables.

Le biais des estimations causé par l'imputation de la non-réponse partielle est difficile à évaluer. Il dépend de la différence entre les répondants et les non-répondants ainsi que de la capacité de la méthode d'imputation à produire des estimations non biaisées. Toutefois, les taux d'imputation donnent une indication de l'importance des non-réponses partielles. Ils sont présentés dans la section suivante.

5.1 Proportion de ménages ou personnes nécessitant imputation, à l'échelle nationale et provinciale

Une première indication de l'ampleur de la non-réponse partielle est la proportion de ménages nécessitant imputation et le nombre de variables imputées par ménage. Le questionnaire peut se diviser en deux grands groupes de variables, celles qui sont recueillies au niveau du ménage et celles collectées auprès des individus telles que le revenu et les dépenses en vêtements. Pour ces dernières catégories, il est important de noter que l'on accepte que le répondant fournisse seulement le revenu total ou les dépenses totales s'il est incapable de fournir la ventilation par source de revenu ou type de dépenses. Le niveau d'imputation pour les composantes du revenu et des dépenses vestimentaires est donc plus élevé mais n'affecte pas le revenu total, les dépenses vestimentaires totales ainsi que le total des dépenses.

Le pourcentage de ménages pour lesquels il a fallu imputer des dépenses de ménage (excluant les dépenses vestimentaires et les dépenses de la section sur l'impôts personnels, sécurité et dons en argent) est présenté dans la prochaine sous-section. Dans celle qui la suit, on présente le pourcentage de personnes pour lesquelles il a fallu imputer au moins une variable reliée aux dépenses vestimentaires, le pourcentage de personnes pour lesquelles au moins une variable de revenu a été imputée et le pourcentage de personnes pour lesquelles il a fallu imputer au moins une variable de la section sur l'impôts personnels, sécurité et dons en argent. Finalement dans la dernière sous-section on présente les résultats pour le pourcentage de ménages pour lesquels il a fallu imputer au moins une des variables catégoriques. À la suite de l'imputation des données par le système, certaines corrections peuvent s'avérer nécessaires, autant sur les variables imputées par le système que sur les autres, afin d'assurer la cohérence des données. Dans les faits, ces modifications représentent un très faible pourcentage. Les résultats sont fournis tant à l'échelle nationale que provinciale. Cela fournit une indication des provinces qui sont les plus affectées par l'imputation.

5.1.1 Imputation de dépenses du ménage par province

Le pourcentage de ménages utilisables pour lesquels il a fallu imputer au moins une variable de dépenses figure dans le tableau 5.1-1. Les ménages utilisables correspondent à tous les ménages demeurant dans des logements admissibles, à l'exception de ceux qui n'ont pu être contactés, qui ont refusé de participer à l'enquête, qui ont fourni des données incomplètes ou qui ne balançaient pas (voir définitions dans la section 2.1). Le tableau est réparti en fonction du nombre de variables qui ont été imputées (parmi les 240) pour un ménage.

Il importe de préciser que les paiements hypothécaires réguliers ainsi que les primes d'assurance hypothécaire sont inclus au titre des dépenses de logement et donc dans les dépenses totales. Depuis 2002, ces deux variables sont ajoutées au calcul des taux d'imputation présentés au tableau 5.1-1. Ceci a pour effet d'augmenter le taux global d'imputation.

Depuis 2004, un changement a été apporté au questionnaire en ce qui concerne les dépenses reliées aux services de communication au foyer (téléphone, téléphone cellulaire et accès à Internet), aux services de télédistribution et aux services de diffusion par satellite. En raison de l'utilisation grandissante des forfaits (services groupés), un ménage peut-être facturé pour des services combinés, de façon qu'il

est plus difficile pour celui-ci de fournir les dépenses pour les services individuels. On accepte alors que le ménage répondant fournisse seulement les dépenses totales de ces services, tout en indiquant quels services sont inclus dans le forfait. L'imputation des dépenses des services individuels s'effectue alors en deux étapes. Dans un premier temps on impute les ménages pour lesquels seulement quelques services sont manquants, suivis des ménages pour lesquels seulement le total des dépenses du forfait est disponible. Pour ces derniers ménages, les dépenses des services (ceux inclus dans le forfait) imputés sont ajustées au prorata afin que leur somme corresponde au total des dépenses du forfait fourni par le ménage répondant. Étant donné que ce changement a eu un impact important sur le taux global d'imputation des dépenses, les taux d'imputation du tableau 5.1-1 sont présentés séparément avec et sans les frais de services de communication au foyer, de location de services de télédistribution et de location de services de diffusion par satellite. De plus, puisque ce changement a eu un impact sur le niveau d'imputation des dépenses de ces cinq services, le taux d'imputation ainsi qu'une mesure d'impact de l'imputation pour chacun de ces services sont présentés au tableau 5.1-2.

Tableau 5.1-1
Ménages qui nécessitent l'imputation de dépenses par province

Province	Ménages qui nécessitent l'imputation de variables de dépenses ¹ (sauf pour les dépenses en vêtements et les dépenses de la section sur l'impôts personnels, sécurité et dons en argent)							
	Exclut les frais de services de communication au foyer et les frais de location pour télédistribution (câble) et pour la diffusion par satellite				Inclut les frais de services de communication au foyer et les frais de location pour télédistribution (câble) et pour la diffusion par satellite			
	Nombre de variables imputées (parmi 240)			Total	Nombre de variables imputées (parmi 245)			Total
	1	2	3 ou plus		1	2	3 ou plus	
	%							
Canada	10,0	1,9	1,3	13,2	7,8	15,1	8,2	31,1
T.-N.-L.	15,2	1,9	0,9	18,0	12,0	10,7	6,7	29,3
Î.-P.-É.	10,9	2,2	0,8	13,9	6,1	15,1	21,1	42,3
N.-É.	13,7	2,7	2,4	18,8	8,8	15,4	24,2	48,4
N.-B.	9,8	1,5	0,6	11,9	8,2	16,9	5,9	31,1
Qc	5,5	0,5	0,2	6,2	5,3	7,1	3,1	15,5
Ont.	8,9	3,2	2,8	15,0	8,2	14,3	8,3	30,7
Man.	5,1	1,1	0,4	6,6	4,5	13,6	2,5	20,6
Sask.	13,5	1,8	0,9	16,2	10,6	15,0	7,4	33,0
Alb.	15,0	2,3	1,4	18,6	11,3	18,1	6,4	35,8
C.-B.	5,6	2,1	2,0	9,7	4,1	25,8	6,4	36,3

1. Inclut les paiements hypothécaires réguliers et les primes d'assurance hypothécaire.

Le tableau 5.1-1 montre qu'il a fallu imputer des dépenses pour 31,1 % des ménages à l'échelle nationale. Ce taux plus élevé en 2004 est dû au changement apporté au questionnaire en ce qui concerne les dépenses reliées aux services de communication au foyer (téléphone, téléphone cellulaire et accès à Internet), aux

services de télédistribution et aux services de diffusion par satellite. En effet, 21,5 % (donnée non présentée) des ménages utilisables ont nécessité l'imputation d'au moins un de ces cinq services. La presque totalité des cas correspond à la situation où les ménages avaient déclaré avoir payé pour un forfait (services groupés) et que les dépenses associées aux services compris dans le forfait ont été imputées. Les taux d'imputation plus élevés lorsque l'on tient compte de ces cinq variables, tels que présentés à la colonne «2 variables imputées» et à la colonne «3 variables imputées ou plus», s'expliquent par le fait qu'un forfait comprend habituellement deux services ou plus. Le taux global d'imputation en excluant les dépenses reliées aux services de communication au foyer, aux services de télédistribution et aux services de diffusion par satellite est de 13,2 % à l'échelle nationale, ce qui est comparable aux taux obtenus les années précédentes. À elle seule, la variable représentant les primes d'assurance hypothécaire nécessite une imputation pour 5,4 % des ménages utilisables (soit 14,4 % des ménages lorsque l'on se restreint aux ménages ayant déclaré des hypothèques sur les logements dont ils étaient propriétaires et occupants).

Lorsque l'on exclut les dépenses reliées aux services de communication au foyer (téléphone, téléphone cellulaire et accès à Internet), aux services de télédistribution et aux services de diffusion par satellite, on constate que près de 75 % des ménages utilisables ont nécessité l'imputation d'une seule variable. De plus, il y a très peu de ménages à l'échelle nationale qui ont eu plus d'une variable imputée (3,2 %). Les provinces pour lesquelles on observe les plus basses proportions de ménages ayant requis l'imputation d'au moins une variable de dépense sont le Québec (6,2 %) et le Manitoba (6,6 %). On retrouve les taux les plus élevés en Nouvelle-Écosse (18,8 %), en Alberta (18,6 %) et à Terre-Neuve-et-Labrador (18,0 %). C'est en Ontario et en Colombie-Britannique où l'on trouve le pourcentage le plus élevé de ménages pour lesquels il a fallu imputer plus d'une variable de dépenses. Dans ces deux provinces, on a imputé au moins deux variables de dépenses pour plus de 40 % des ménages qui nécessitaient une imputation.

Si on exclut les paiements hypothécaires réguliers, les primes d'assurance hypothécaire, les dépenses reliées aux services de communication au foyer aux services de télédistribution et aux services de diffusion par satellite, le faible pourcentage de ménages pour lesquels des variables doivent être imputées, combiné à un nombre généralement peu élevé de variables à imputer lorsque l'imputation est nécessaire, suggère que l'impact des valeurs imputées sur les estimations ne devrait pas être trop élevé.

Puisque le niveau d'imputation des dépenses reliées aux services de communication au foyer, aux services de télédistribution et aux services de diffusion par satellite est plus élevé, il est important de mesurer l'effet de l'imputation sur les estimations de totaux de ces cinq variables. Cette mesure ainsi que le taux d'imputation permettront de faire un suivi dans le temps de l'ampleur de l'imputation effectuée pour ces variables. La popularité grandissante des forfaits (services groupés) dans la population fait en sorte que le niveau d'imputation devrait augmenter au cours du temps. La mesure d'impact de l'imputation est définie par le total pondéré des données imputées divisé par l'estimation totale (sommes des valeurs pondérées). Cette mesure représente la proportion de la valeur totale de l'estimation provenant de données imputées.

Tableau 5.1-2
Impact de l'imputation à l'échelle nationale des frais de services de communication au foyer, des frais de location pour la télédistribution et pour la diffusion par satellite

Caractéristiques	Impact de l'imputation	Ménages déclarants qui nécessitent l'imputation des frais de services de communication au foyer, des frais de location pour télédistribution (câble) ou pour la diffusion par satellite
		%
Services téléphoniques	7,2	12,2
Services de téléphones cellulaires, téléavertisseurs et appareils de messagerie textuelle portatifs	3,5	6,1
Location de services de télédistribution (câble)	20,4	20,5
Location de services pour la diffusion par satellite	6,8	3,7
Services d'accès à Internet	32,8	40,3

Selon le tableau 5.1-2, le taux d'imputation et l'impact de l'imputation sont plus importants pour les dépenses reliées aux services d'accès à Internet ainsi qu'à la location de services de télédistribution. Cela est principalement dû au fait qu'une proportion importante des forfaits incluait ces deux services parmi les ménages qui ont déclaré avoir payé pour un forfait. Le niveau élevé d'imputation effectuée sur les composantes du tableau 5.1-2 implique que les estimations de celles-ci pourraient être grandement affectées par l'imputation, alors que l'effet sur l'estimation du total de ces cinq services combinés sera négligeable, puisque les ménages doivent fournir le total des dépenses associées au forfait. Bien que le taux d'imputation ainsi que l'impact soient élevés pour les dépenses en services d'accès à Internet, la hausse observée en 2004 pour la moyenne des dépenses d'accès à Internet était cohérente avec les tendances observées provenant d'autres sources d'information indépendantes. Les services d'accès à Internet représentent 16 % de toutes les dépenses des ménages au chapitre des communications. Par ailleurs, le total des dépenses des cinq services combinés du tableau 5.1-2 ne représente que 2,6 % des dépenses totales des ménages.

5.1.2 Imputation des dépenses des personnes et du revenu par province

Étant donné qu'une partie des répondants fournissent uniquement le total pour les variables de dépenses en vêtements et les variables de revenu, celles-ci sont imputées, au niveau des individus, en deux étapes. Les individus pour lesquels il faut imputer seulement quelques composantes sont imputés en premier lieu, suivis par ceux pour lesquels seulement les totaux sont disponibles et qui nécessitent l'imputation de toutes les composantes. (Voir la référence [1] pour une description plus détaillée de ce processus).

Le pourcentage d'individus utilisables (personnes membres des ménages utilisables) pour lesquels il a fallu imputer au moins une variable de revenu est indiqué par

province dans le tableau 5.2. On y retrouve le pourcentage de personnes pour lesquelles exactement une variable a été imputée, ceux dont deux variables ou plus (mais pas toutes) ont été imputées et le pourcentage des personnes pour lesquelles seul le revenu total était disponible, nécessitant ainsi l'imputation de toutes les composantes. Le pourcentage total de personnes pour lesquelles une forme d'imputation de revenu a été effectuée est également donné. L'avant dernière colonne du tableau 5.2 contient ce même pourcentage total de personnes pour lesquelles une forme d'imputation a été effectuée, mais pour les dépenses en vêtements. La dernière colonne contient le pourcentage total de personnes pour lesquelles une forme d'imputation a été effectuée pour les variables de la section du questionnaire portant sur les impôts personnels, sécurité et dons en argent.

Il importe de préciser que seuls les membres du ménage âgés de 15 ans et plus au 31 décembre de l'année de référence doivent répondre aux questions portant sur le revenu personnel et celles de la section sur les impôts personnels, sécurité et dons en argent. Par conséquent, depuis l'année de référence 2003, le pourcentage de personnes qui nécessitent l'imputation de variables de revenu ou pour celles de la section sur l'impôts personnels, sécurité et dons en argent a été calculé en se restreignant aux personnes de 15 et plus et non sur l'ensemble des personnes comme pour les années précédentes. Ceci a pour effet d'augmenter légèrement le taux d'imputation pour ces variables. Tout comme pour les années précédentes, le pourcentage de personnes qui nécessitent l'imputation de dépenses vestimentaires est calculé sur l'ensemble des personnes, puisque tous les membres d'un ménage doivent répondre aux questions portant sur ce type de dépenses.

Tableau 5.2
Personnes qui nécessitent l'imputation du revenu, personnes qui nécessitent l'imputation de dépenses en vêtements et personnes qui nécessitent l'imputation de variables de la section sur les impôts personnels, sécurité et dons en argent par province

Province	Pourcentage de personnes qui nécessitent l'imputation de variables de revenu				Pourcentage de personnes qui nécessitent l'imputation d'au moins une des 11 variables de dépenses vestimentaires	Pourcentage de personnes qui nécessitent l'imputation d'au moins une des 15 variables de la section sur les impôts personnels, sécurité et dons en argent
	1 variable imputée	2 variables imputées ou plus (pas toutes)	Toutes les variables imputées (revenu total connu)	Total (toute forme d'imputation)		
	%					
Canada	0,7	0,2	4,0	5,0	20,2	3,8
T.-N.-L.	0,3	0,2	3,5	4,2	13,6	2,6
Î.-P.-É.	0,1	0,4	4,8	5,6	15,1	3,7
N.-É.	1,2	0,3	6,9	8,5	18,8	6,1
N.-B.	0,6	0,2	3,9	4,7	15,1	3,1
Qc	0,2	0,1	3,3	3,6	28,1	1,2
Ont.	1,8	0,8	2,8	5,4	16,6	7,0
Man.	0,2	0,1	2,1	2,4	22,5	2,9
Sask.	0,5	0,1	3,2	3,8	17,4	3,7
Alb.	0,8	0,1	3,7	4,7	23,7	3,7
C.-B.	0,8	0,2	6,5	7,6	25,3	3,9

À partir de ces résultats, on constate qu'il y a 5 % des personnes des ménages utilisables pour lesquelles il a fallu imputer au moins une variable de revenu. Pour 80 % d'entre elles, le revenu total était disponible mais toutes les composantes ont dû être imputées. Pour une très grande partie des autres personnes qui ont nécessité imputation, seulement une composante de revenu (une variable) a été imputée. À l'échelle provinciale, le pourcentage de personnes nécessitant l'imputation d'au moins une variable de revenu varie entre un minimum de 2,4 % pour le Manitoba et un maximum de 8,5 % pour la Nouvelle-Écosse. On observe un taux d'imputation plus élevé en Nouvelle-Écosse (8,5 %) et en Colombie-Britannique (7,6 %). Pour l'Ontario, l'imputation se divise principalement entre le cas où il n'y a qu'une seule composante à imputer et le cas où total du revenu est fourni par le répondant mais toutes les composantes doivent être imputées.

Dans l'avant dernière colonne du tableau, on constate qu'environ 20 % des personnes ont nécessité une imputation pour au moins une des variables de dépenses vestimentaires. Les taux provinciaux s'échelonnent de 13,6 % pour Terre-Neuve-et-Labrador à 28,1 % pour le Québec. Presque toutes ces personnes ont fourni le total de leurs dépenses vestimentaires, mais il a fallu imputer les composantes. Terre-Neuve-et-Labrador est la seule province où le taux d'imputation est inférieur à 15 %. On observe également qu'à l'exception de la Nouvelle-Écosse, ces taux sont un peu plus faibles dans les provinces de l'Atlantique que pour les autres provinces. Bien que le Québec affiche le taux le plus élevé d'imputation des dépenses vestimentaires, cette province présente des taux d'imputation parmi les plus faibles pour le revenu ou pour les variables de la section sur les impôts personnels, sécurité et dons en argent. Le niveau élevé d'imputation effectuée sur les composantes des dépenses vestimentaires implique que les estimations de ces composantes pourraient être grandement affectées par l'imputation, alors que l'effet sur les estimations du total sera négligeable.

Dans la dernière colonne du tableau, on constate que près de 4 % des personnes ont nécessité une imputation pour au moins une des variables de la section sur les impôts personnels, sécurité et dons en argent. À l'échelle provinciale, ce pourcentage est également peu élevé, variant entre un minimum de 1,2 % au Québec et un maximum de 7,0 % pour l'Ontario. Seules les provinces de la Nouvelle-Écosse et de l'Ontario affichent un taux supérieur à 4 %. Dans ces deux provinces, le taux d'imputation plus élevé est dû à la variable représentant l'impôt personnel sur le revenu de 2004 qui nécessite en proportion plus d'imputation que pour les autres provinces.

5.1.3 Imputation des variables catégoriques par province et territoire

Le pourcentage de ménages utilisables pour lesquels il a fallu imputer au moins une variable catégorique figure dans le tableau 5.3. Le tableau est réparti en fonction du nombre de variables qui ont été imputées (parmi les 58) pour un ménage. Les variables catégoriques qui sont imputées se trouvent aux sections suivantes du questionnaire: Caractéristiques du logements (à l'exception de la variable genre de logement); Équipement associé au logement; Mode d'occupation du logement (à l'exception des variables se rapportant au changement dans le mode d'occupation durant l'année de référence); Tabac et dépenses diverses, pour les variables se rapportant aux achats par le biais de vente directe (questions pour lesquelles la

réponse est oui ou non). Il est à noter que les autres variables catégoriques du questionnaire telles que celles se rapportant à la composition du ménage ou encore aux divers branchements du questionnaire sont vérifiées et validées par les spécialistes du sujet de la Division de la statistique du revenu. Par conséquent, ces dernières variables ne sont pas imputées au moyen de la technique du plus proche voisin.

Tableau 5.3
Ménages qui nécessitent l'imputation de variables catégoriques par province

Province	Ménages qui nécessitent l'imputation de variables catégoriques			
	Nombre de variables imputées (parmi 58)			Total
	1	2	3 ou plus	
%				
Canada	5,5	1,1	0,9	7,6
Terre-Neuve-et-Labrador	4,0	0,4	0,2	4,6
Île-du-Prince-Édouard	5,1	1,0	0,7	6,8
Nouvelle-Écosse	3,2	0,5	0,0	3,6
Nouveau-Brunswick	1,8	0,3	0,1	2,2
Québec	6,5	0,7	0,7	8,0
Ontario	5,4	1,3	1,3	8,0
Manitoba	5,7	1,3	1,1	8,1
Saskatchewan	7,2	1,1	1,6	9,9
Alberta	8,5	2,3	2,1	12,9
Colombie-Britannique	6,4	2,1	1,2	9,7

Le tableau 5.3 montre qu'à l'échelle nationale, il a fallu imputer au moins une variable catégorique se rapportant aux caractéristiques du logements et équipement associé au logement, au mode d'occupation du logement et aux achats par le biais de vente directe pour 7,6 % des ménages. Toutefois, pour environ 72 % d'entre eux seulement une variable a été imputée. On observe les plus basses proportions de ménages ayant requis l'imputation d'au moins une variable catégorique pour les provinces de l'Atlantique. À l'opposé, les provinces de l'Ouest canadien affichent les taux les plus élevés.

Références

- [1] Tremblay J. et Arsenault, S. 2001. *La méthodologie de l'Enquête sur les dépenses des ménages*. N° 62F0026MIF2001003 au catalogue. Ottawa. Division des méthodes d'enquêtes auprès des ménages, Statistique Canada.
- [2] Rao, J.N.K., Wu, C.F.J. et Yue K. 1992. « Quelques travaux récents sur les méthodes de rééchantillonnage applicables aux enquêtes complexes », *Techniques d'enquête*, 18(2), p.225-234 (Statistique Canada, N° 12-001 au catalogue).
- [3] Statistique Canada, Division de la statistique de revenu. 2004. *Guide de l'utilisateur — Enquête sur les dépenses des ménages*. N° 62F0026MIF au catalogue - N° 007. Ottawa.
- [4] Statistique Canada, Division de la statistique de revenu. 2005. *Les habitudes de dépenses au Canada*. N° 62-202 au catalogue. Ottawa.
- [5] Statistique Canada, Division des méthodes d'enquêtes auprès des ménages. 2004. *Enquête sur la population active, Rapport sur les opérations*. Enquête 200412. Ottawa.
- [6] Lessard, S. 2005. *Révision de la stratégie de calage de l'Enquête sur les dépenses des ménages*. Ottawa. Division des méthodes d'enquêtes auprès des ménages, Statistique Canada.

Annexe A

Notation algébrique

1. Ajustement pour la non-réponse

Le sous-poids (c-à-d. le poids de sondage ajusté pour la non-réponse) d'un ménage k , noté w_k^{NR} , est

$$w_k^{NR} = \pi_k^{-1} * \frac{1}{\text{taux}_g} \quad \text{avec} \quad \text{taux}_g = \frac{\sum_{k \in s_{g,r}} \pi_k^{-1}}{\sum_{k \in s_{g,r}} \pi_k^{-1} + \sum_{k \in s_{g,nr}} \pi_k^{-1}}$$

où

$s_{g,r}$ est l'ensemble des ménages répondants dans le groupe de non-réponse g ,

$s_{g,nr}$ est l'ensemble des ménages non-répondants (refus, non-contacts, données inutilisables) dans le groupe de non-réponse g , et

π_k^{-1} est le poids de sondage attribué au ménage k .

2. Calcul du taux de glissement

Le taux de glissement pour un groupe de contrôle c , noté taux_c , est,

$$\text{taux}_c = 100 * \frac{\left(\sum_{k \in s_{c,r}} w_k^{NR} \right) - t_c}{t_c}$$

où

$s_{c,r}$ est l'ensemble des ménages répondants dans le groupe de contrôle c ,

w_k^{NR} est le sous-poids du ménage k , et

t_c est le total de la donnée auxiliaire pour le groupe de contrôle c .