



Division de la statistique du revenu

62F0026MIF - 01001

Enquête sur les dépenses des ménages de 1998 Indicateurs de qualité des données

Préparé par :
Sophie Arsenault
Jean-Luc Bernier
Jean-Marc Fillion
Sunita Patel
Johanne Tremblay

Octobre 2001



Statistique
Canada

Statistics
Canada

Canada

Des données sous plusieurs formes

Statistique Canada diffuse les données sous formes diverses. Outre les publications, des totalisations habituelles et spéciales sont offertes. Les données sont disponibles sur Internet, disque compact, disquette, imprimé d'ordinateur, microfiche et microfilm, et bande magnétique. Des cartes et d'autres documents de référence géographiques sont disponibles pour certaines sortes de données. L'accès direct à des données agrégées est possible par le truchement de CANSIM, la base de données ordiolinguistique et le système d'extraction de Statistique Canada.

Comment obtenir d'autres renseignements

Toute demande de renseignements au sujet du présent produit ou au sujet de statistiques ou de services connexes doit être adressée à : Services aux clients, Division de la statistique du revenu, Statistique Canada, Ottawa, Ontario, K1A 0T6 ((613) 951-7355; (888) 297-7355: revenu@statcan.ca) ou à l'un des centres de consultation régionaux de Statistique Canada :

Halifax	(902) 426-5331	Regina	(306) 780-5405
Montréal	(514) 283-5725	Edmonton	(403) 495-3027
Ottawa	(613) 951-8116	Calgary	(403) 292-6717
Toronto	(416) 973-6586	Vancouver	(604) 666-3691
Winnipeg	(204) 983-4020		

Vous pouvez également visiter notre site sur le Web : <http://www.statcan.ca>

Un service d'appel interurbain sans frais est offert à **tous les utilisateurs qui habitent à l'extérieur des zones de communication locale** des centres de consultation régionaux.

Service national de renseignements	1 800 263-1136
Service national d'appareils detélécommunications pour les malentendants	1 800 363-7629
Numéro pour commander seulement (Canada et Etats-Unis)	1 800 267-6677

Renseignements sur les commandes et les abonnements

Les prix ne comprennent pas les taxes de vente

On peut se procurer ce produit n° 62F0026MIF-01001 au catalogue sur internet gratuitement. Pour obtenir un numéro de ce produit, les utilisateurs sont priés de se rendre à http://www.statcan.ca/cgi-bin/downpub/research_f.cgi.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois et dans la langue officielle de leur choix. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec le centre de consultation régional de Statistique Canada le plus près de chez vous.



Statistique Canada
Division de la statistique du revenu

Enquête sur les dépenses des ménages de 1998 Indicateurs de qualité des données

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2001

Tous droits réservés. Il est interdit de reproduire ou de transmettre le contenu de la présente publication, sous quelque forme ou par quelque moyen que ce soit, enregistrement sur support magnétique, reproduction électronique, mécanique, photographique, ou autre, ou de l'emmagasiner dans un système de recouvrement, sans l'autorisation écrite préalable des Services de concession des droits de licence, Division du marketing, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

octobre 2001

N° 62F0026MIF - 01001 au catalogue

Périodicité : Irrégulier

Ottawa

This publication is available in English upon request.

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.

Résumé:

Ce document présente les indicateurs de qualité produits pour l'enquête sur les dépenses des ménages de 1998. On y trouve les indicateurs de qualité usuels utiles aux utilisateurs pour l'interprétation des données tel que les coefficients de variation, les taux de non-réponse, les taux d'imputation et l'impact des données imputées sur les estimations. On y a également ajouté certains indicateurs moins fréquemment utilisés comme les taux de glissement et des mesures de la représentativité de l'échantillon pour certaines caractéristiques qui permettent d'évaluer la méthodologie de l'enquête

Les auteurs remercient l'équipe responsable de l'enquête sur les dépenses des ménages de la division des statistiques sur le revenu (DSR) ainsi que Sylvana Beaulieu et José Gaudet pour leur collaboration dans la préparation de ce document.

TABLE DES MATIÈRES

FAITS SAILLANTS.....	9
INTRODUCTION	11
1. LES ERREURS D'ÉCHANTILLONAGE.....	12
1.1 Les mesures de l'erreur d'échantillonnage.....	12
1.2 Les coefficients de variation.....	13
Tableau 1.1	
Coefficients de variation (%) par province et territoire ainsi qu'au niveau du Canada pour l'estimation de la moyenne des dépenses par ménage pour quelques catégories de dépenses sommaires et pour l'estimation du revenu moyen	14
Tableau 1.2	
Coefficients de variation (%) par province et territoire ainsi qu'au niveau du Canada pour quelques catégories des caractéristiques des logements et l'équipement ménager.	15
1.3 Modèle pour dériver une approximation du CV	15
1.4 Suppression des données non fiables dans les tableaux d'estimations	17
2. LA NON-RÉPONSE.....	17
2.1 Les taux de non-réponse et les taux de vacance	18
Tableau 2.1	
Taux de non-réponse (%) et taux de vacance (%) par province ou territoire ainsi qu'au niveau Canada	19
2.2 La non-réponse selon le niveau d'urbanisation	20
Tableau 2.2	
Taux de non-réponse (%) et taux de vacance (%) par niveau d'urbanisation ...	20
2.3 La non-réponse selon les strates de revenu.....	21
Tableau 2.3	
Comparaison des taux de non-réponse (%) et de vacance (%) des strates à revenu élevé et à faible revenu par rapport aux autres strates.....	21
2.4 L'ajustement pour la non-réponse.....	21
3. LES ERREURS DE COUVERTURE	22
3.1 Le sous-dénombrement et le surdénombrement : les taux de glissement	23
Tableau 3.1	
Taux de glissement nationaux selon le groupe d'âge croisé avec le sexe.....	24
Tableau 3.2	
Taux de glissement pour les provinces et les territoires selon le groupe d'âge-sexe croisé.....	25
Tableau 3.3	
Taux de glissement pour les provinces et les territoires selon la taille de ménage.....	26
3.2 L'ajustement pour l'erreur de couverture au niveau de la population et du nombre de ménages.....	26
4. LES ERREURS DE RÉPONSE.....	27

TABLE DES MATIÈRES (fin)

5. LES ERREURS DE TRAITEMENT	28
5.1 Mesure d'impact.....	30
5.2 Proportion de ménages ou personnes imputés au niveau national et provincial.....	31
Tableau 5.1 Ménages qui nécessitent l'imputation de dépenses par province	32
Tableau 5.2 Personnes qui nécessitent l'imputation du revenu et personnes qui nécessitent l'imputation de dépenses en vêtements par province	33
BIBLIOGRAPHIE.....	35
ANNEXE A	
LISTE DE TABLEAUX DÉTAILLÉS DISPONIBLES SUR DEMANDE.....	36
ANNEXE B	
NOTATION ALGÈBRE.....	37

FAITS SAILLANTS

Les erreurs d'échantillonnage

- Les coefficients de variation (CV) de l'estimation de la moyenne des dépenses par ménage varient entre 1,5% et 2,6% selon les provinces. Les CV sont plus élevés dans les territoires soit 5,2% au Yukon, 5,4% dans les territoires du Nord-Ouest et 8,9% au Nunavut.

La non-réponse

- Le taux de non-réponse est de 23,6%, attribuable à près de 5% des ménages qu'on n'a pas réussi à contacter, à 16% de refus et à un peu moins de 3% de ménages pour lesquels les données ont été considérées inutilisables soit parce qu'elles étaient incomplètes (1,6%) ou encore qu'elles ont été rejetées parce que les différences entre les entrées et les déboursés rapportés sur le questionnaire excédaient 20% (1,3%).
- L'analyse des taux de réponse dans les strates d'aires géographiques à revenu élevé et à faible revenu provenant du plan d'échantillonnage révèle que le taux de non-réponse dans les strates à revenu élevé est environ 65% supérieur à celui observé dans les strates régulières surtout à cause des refus et des données incomplètes qui sont beaucoup plus fréquents.
- Les ménages des strates à faible revenu ont un taux de non-réponse final environ 6 points de pourcentage inférieur à ceux des strates régulières et cette différence provient presque entièrement d'un comportement différent à la collecte puisqu'il y a 5 points de pourcentage de moins au niveau du taux de refus pour les strates à revenu faible par rapport aux strates régulières.

Les erreurs de couverture

- Le taux de glissement pour les individus de 15 ans et plus indique un sous-dénombrement de 9,6% de la population, un taux similaire à celui que l'on observe dans l'Enquête sur la population active (EPA) pour une même période. Le taux de glissement pour les enfants (0 à 14 ans) est très différent de celui des autres groupes d'âge. On observe un sur-dénombrement des enfants de 4,9%.
- Les poids de l'enquête sont ajustés pour tenir compte du glissement mais le biais sera réduit dans la mesure où les caractéristiques des individus omis sont similaires à celles des individus inclus dans les groupes d'âge provinciaux utilisés pour l'ajustement.

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'EDM, 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 moyen pour compléter l'entrevue pour les ménages répondants qui ont fourni des données utilisables est de 1 heure 50 minutes. Le temps total d'entrevue varie selon les caractéristiques des ménages et peut prendre plus de 5 heures pour certains ménages.

Les erreurs de traitement reliées à l'imputation

i) Des variables de dépenses

- Il a fallu imputer au moins une variable de dépenses pour 5,9% des ménages mais pour la majorité d'entre eux, seulement une ou deux variables parmi les 237 variables de dépenses (on exclue ici la section sur les dépenses vestimentaires) ont été imputées.
- Les valeurs imputées contribuent à seulement 0,4% des dépenses totales et à moins de 1% des estimations du total des dépenses dans chaque catégorie de dépenses sommaire, sauf pour les dépenses diverses (1,3%).
- Un peu plus de 15% 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 les composantes ont été imputées.

ii) Des variables de revenu

- Il a fallu imputer au moins une variable de revenu pour moins de 3% des individus. Pour un peu plus de 60% d'entre eux, le revenu total était fourni par le répondant mais toutes les composantes (sources de revenu) ont été imputées.
- Les valeurs imputées contribuent à seulement 0,2% de l'estimation du revenu total.

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 auprès d'un échantillon d'environ 24 000 ménages répartis dans les 10 provinces et les trois territoires¹. 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. Cette enquête remplace l'enquête périodique sur les dépenses des familles avec comme principales différences que le niveau de détail a été réduit dans le questionnaire, que l'échantillon a été augmenté pour satisfaire les besoins d'estimations provinciales du Projet d'Amélioration des Statistiques Économiques Provinciales (PASEP) et que l'enquête est maintenant annuelle. Elle devient également le véhicule de collecte pour les caractéristiques des logements et l'équipement ménager en remplacement de l'Enquête sur l'Équipement Ménager. L'enquête de 1997, qui réfère aux dépenses effectuées au cours de l'année 1997, correspond à la première année de l'enquête suite à ce remaniement.

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 1998. On y trouve les indicateurs de qualité usuels généralement utiles aux utilisateurs pour l'interprétation des données tel que les coefficients de variation, les taux de non-réponse, les taux d'imputation et l'impact des données imputées sur les estimations. On y a également ajouté certains indicateurs moins fréquemment utilisés comme les taux de glissement et des mesures de la représentativité de l'échantillon pour certaines caractéristiques qui permettent d'évaluer la méthodologie de l'enquête. Finalement on y trouve les résultats de quelques évaluations notamment des modèles fournis aux utilisateurs pour calculer une approximation des CV et de la règle de suppression qui sert à déterminer si une estimation est suffisamment fiable pour être publiée.

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. Finalement, les erreurs de réponse et les erreurs de traitement sont traitées dans les sections 4 et 5 respectivement.

¹ L'EDM de 1998 est la première année pour laquelle nous avons des estimations pour le Nunavut. Dans les années antérieures, le Nunavut était inclus dans les Territoires du Nord-Ouest.

Ce document est axé sur la qualité des données. Pour une description détaillée de la méthodologie de l'enquête, on peut consulter la référence [6]. On notera également que plusieurs tableaux plus détaillés que ceux présentés dans ce rapport peuvent être obtenus auprès de la Division des méthodes d'enquêtes auprès des ménages. La liste de ces tableaux est fournie dans l'annexe A.

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 Jackknife. Pour plus de détails sur cette méthode, voir les références [1] et [7].

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 $(Y \pm 2 SE)^2$. 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.

² 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. Pour une même caractéristique de dépense, plusieurs estimations sont produites comme par exemple l'estimation du total des dépenses des ménages, l'estimation de la moyenne des dépenses par ménage, l'estimation du nombre de ménages déclarant une valeur supérieure à 0 ou encore l'estimation de la moyenne des dépenses pour les ménages déclarant une valeur supérieure à 0. Généralement dans l'EDM, on publie le CV de l'estimation de la moyenne des dépenses par ménage. À partir de l'EDM de 1998, ce CV est disponible dans la publication *Les habitudes de dépenses au Canada* (référence [8]), au niveau national pour chacune des dépenses détaillées collectées ainsi qu'au niveau provincial ou territorial pour plusieurs catégories de dépenses sommaires³. Les CV des moyennes des dépenses détaillées au niveau provincial ou territorial sont disponibles sur demande auprès de la Division des statistiques sur le revenu. On peut obtenir également les CV pour les caractéristiques des logements et l'équipement ménager.

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, l'effet de l'imputation est minime (voir section 6) et les CV fournis représentent une bonne estimation des vrais CV mais si on s'intéresse à la fiabilité de dépenses détaillées avec un taux d'imputation élevé, il est important de tenir compte à la fois du CV et de l'impact des données imputées sur les estimations.

Le tableau 1.1 donne un aperçu des CV des estimations des moyennes par ménage au niveau provincial ou territorial ainsi qu'au niveau national pour l'estimation de quelques-unes des catégories de dépenses sommaires et pour le revenu.

³ Dans les enquêtes précédentes, on publiait les CV au niveau national et pour différents groupes de revenu.

Tableau 1.1
Coefficients de variation (%) par province et territoire ainsi qu'au niveau du Canada 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.	Î.P.É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sas k.	Alb.	C.-B.	Yn	T.N.-O.	Nt
Dépenses totales	0,8	2,5	2,6	2,1	1,9	1,6	1,5	1,8	1,8	1,6	2,0	5,2	5,4	8,9
Dépenses de consommation	0,6	1,9	2,0	1,7	1,5	1,3	1,1	1,4	1,5	1,4	1,7	4,3	3,6	8,0
Alimentation	0,6	1,3	1,5	1,3	1,7	1,1	1,1	1,2	1,3	1,2	1,3	4,1	4,4	7,9
Logement	0,8	2,9	2,9	2,1	2,5	1,4	1,5	1,7	2,1	1,8	1,8	3,0	7,4	17,2
Entretien ménager	1,0	2,4	2,8	2,6	2,4	1,9	1,9	1,8	2,2	1,9	2,4	5,7	5,9	10,4
Ameublement	2,2	3,6	4,2	3,5	3,5	3,5	4,2	3,8	3,7	3,8	6,2	7,6	10,1	10,6
Habillement	1,1	2,6	3,2	2,9	2,6	2,1	2,1	2,5	2,5	3,5	2,5	7,7	6,3	4,5
Transport	1,4	4,1	4,4	3,6	3,3	2,9	2,6	3,7	3,2	3,2	4,0	9,1	5,0	30,6
Soins de santé	1,3	3,7	3,9	3,5	2,6	2,2	3,2	3,3	3,0	2,2	3,0	7,0	5,7	19,3
Soins personnels	1,0	2,4	3,0	2,8	2,1	2,0	1,9	2,7	2,3	2,1	2,4	5,6	6,4	12,3
Loisirs	1,6	3,7	6,1	3,6	3,8	4,0	2,7	3,6	3,5	3,6	3,9	8,9	9,1	9,1
Lecture et autres imprimés	1,4	4,4	4,2	3,6	3,1	2,9	2,7	3,9	3,0	2,9	2,9	12,3	5,5	19,7
Éducation	3,3	8,5	13,7	11,6	10,3	6,1	6,6	8,2	7,7	5,7	7,1	24,3	9,9	42,2
Tabac, boissons alcoolisées	1,5	4,2	5,7	5,1	3,6	2,5	3,2	4,0	4,4	4,0	4,0	7,1	9,5	25,1
Jeux de hasards (net)	3,2	6,0	8,6	8,3	8,2	4,9	7,0	8,7	11,3	8,6	6,5	17,3	23,5	15,0
Dépenses diverses	3,2	7,1	7,7	7,4	6,2	8,2	6,0	10,3	6,5	4,1	6,7	10,0	9,3	8,7
Impôts personnels	1,8	5,0	5,2	4,2	4,0	3,1	3,6	4,2	4,0	3,0	4,3	9,2	8,6	17,0
Assurances personnelles	1,5	5,3	4,7	3,6	3,6	3,1	3,0	5,2	7,9	2,3	3,1	8,9	7,0	11,3
Dons et contributions	3,3	7,0	16,9	8,0	7,4	8,1	6,5	7,9	7,4	6,5	7,5	30,2	44,2	33,8
Revenu	1,0	2,6	2,6	2,2	2,1	1,8	1,9	2,0	2,1	1,6	2,0	6,2	5,0	8,1

Les coefficients de variation de l'estimation de la moyenne des dépenses par ménage varient entre 1,5% et 2,6% selon les provinces. Les CV sont plus élevés dans les territoires soit 5,2% au Yukon, 5,4% dans les territoires du Nord-Ouest et 8,9% au Nunavut.

Pour les différentes catégories de dépenses sommaires, les CV au niveau national sont également inférieurs ou égaux à 2,2% à l'exception des dépenses dans les catégories suivantes: *éducation, jeux de hasards, dépenses diverses* ainsi que *dons en argent et contributions*. Ces catégories de dépenses représentent respectivement 1,3%, 0,5%, 1,6% et 2,2% du total des dépenses.

Le tableau 1.2 donne un aperçu des CV pour les caractéristiques des logements et l'équipement ménager au niveau provincial ou territorial ainsi qu'au niveau national pour l'estimation de quelques unes des catégories .

Tableau 1.2
Coefficients de variation (%) par province et territoire ainsi qu'au niveau du Canada pour quelques catégories des caractéristiques des logements et l'équipement ménager.

CATÉGORIES	Can.	T.-N.	Î.P.É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sask	Alb.	C.-B.	Yn	T.N.-O.	Nt
Propriétaire	1,0	1,9	3,3	2,1	2,3	1,8	2,1	2,3	1,9	2,0	1,9	8,3	13,8	5,6
Locataire	1,8	6,2	7,7	5,6	6,3	2,4	4,2	5,1	4,3	4,4	3,4	15,4	12,4	1,7
Machine à laver (automatique)	0,6	1,4	2,1	1,7	2,0	0,9	1,5	1,7	1,5	1,2	1,5	6,8	6,3	10,2
Sécheuse	0,6	1,2	2,3	1,7	2,0	1,0	1,5	1,7	1,4	1,2	1,5	6,8	7,7	9,2
Lave-vaisselle encastré	1,5	5,9	6,0	4,8	5,1	2,5	3,3	3,8	3,8	2,8	2,9	11,5	11,7	27,0
Congélateur	0,9	1,6	3,0	2,2	2,3	2,2	1,9	1,8	1,4	1,8	2,1	8,1	7,0	13,4
Four à micro-ondes	0,4	1,1	1,5	1,0	0,9	0,8	0,8	1,0	1,0	0,7	1,0	2,3	3,8	9,2
Téléphone cellulaire	2,0	7,8	9,6	6,1	6,4	5,4	3,8	5,1	4,5	2,9	3,6	16,4	21,0	-
Lecteur de disques compacts	0,7	2,0	2,7	2,2	2,1	1,5	1,5	2,0	2,0	1,4	1,6	4,2	2,9	7,7
Télédistribution	1,1	2,0	4,2	2,0	2,3	1,8	2,5	1,7	2,7	1,8	1,5	11,7	4,2	10,4
Ordinateur personnel	1,3	4,0	6,0	3,8	4,7	2,9	2,4	30,0	3,4	2,4	2,4	5,7	5,4	11,5
Modem	1,8	4,8	8,0	5,5	5,9	4,1	3,4	4,0	4,6	3,2	3,3	6,9	10,2	22,5
Utilisation d'Internet (maison)	2,1	5,9	9,5	6,6	6,5	4,9	4,0	4,8	5,7	3,8	4,3	8,4	10,9	28,9
Véhicules possédés (un)	1,3	3,8	4,8	3,5	3,9	2,3	2,9	2,8	3,0	3,0	2,9	8,9	11,1	25,7
Véhicules possédés (2 ou plus)	1,5	4,8	4,7	4,0	3,8	3,4	3,1	3,2	2,9	2,5	2,9	10,4	20,2	60,0

Les coefficients de variation CV pour les caractéristiques des logements et l'équipement ménager sont généralement inférieurs à 6% pour toutes les provinces sauf pour les catégories : *locataire*, *téléphone cellulaire*, *utilisation d'Internet et modem* dans certaines provinces de l'Atlantiques. Les CV sont plus élevés dans les territoires où l'on retrouve une proportion moindre des équipements.

Les CV pour les caractéristiques des logements et l'équipement ménager au niveau national sont également inférieurs à 2,1%.

1.3 Modèle pour dériver une approximation du CV

Pour des raisons opérationnelles, il n'est pas possible de produire les CV pour toutes les caractéristiques collectées par l'enquête à tous les différents niveaux d'agrégations qui peuvent intéresser les utilisateurs comme par exemple, par quintile de revenu, par type de ménage, par niveau d'urbanisation, par mode d'occupation du logement ou pour certaines régions métropolitaines

sélectionnées. Ces estimations sont toutefois disponibles à ces différents niveaux pour les catégories de dépenses sommaires dans la publication *Les habitudes de dépenses au Canada* (référence [8]) et pour les dépenses détaillées auprès de la Division des statistiques sur le revenu.

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 à la page suivante.

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 ménage 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 ménage 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 au niveau national pour calculer les approximations pour les différents domaines. Dans le cas où on cherche à calculer une approximation du CV pour une région métropolitaine, on peut utiliser ces valeurs au niveau provincial puisque le domaine est entièrement contenu dans une seule province et que les CV provinciaux sont publiés pour l'EDM de 1998. 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 micro-données

Les utilisateurs du fichier de micro-données 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 micro-données de 1998. Elle est utilisable seulement à partir du fichier de micro-données 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 des deux méthodes d'approximation du CV que nous venons de décrire.

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 ménage qui déclarent une valeur supérieure à 0⁴.

On peut démontrer que les CV atteignent en général environ 33% lorsque le nombre de ménages qui déclarent une dépense s'approche de 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%.

Comme les CV sommaires calculés pour l'évaluation du modèle dans la section 1.3 comportent peu de cas où le nombre de ménages qui déclarent une dépense est inférieur à 30, seulement les données au niveau détaillé ont été utilisées pour estimer les risques d'erreurs lors de la suppression. 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 non utilisables.

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

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 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 non admissibles 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 ont été laissées 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 reçus par le 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.

Le tableau 2.1 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 données inutilisables, ventilé selon les incomplets et les non balancés. On y trouve également le taux de vacance. Ces taux sont fournis au niveau national ainsi qu'au niveau provincial ou territorial.

Tableau 2.1
Taux de non-réponse (%) et taux de vacance (%) par province ou territoire
ainsi qu'au niveau Canada

Province ou territoire	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 balancé	
Canada	10,5	20,7	4,9	15,8	2,9	1,6	1,3	23,6
Terre-Neuve	15,1	14,1	4,2	9,9	2,8	1,5	1,3	16,8
I.-P.-É.	17,4	16,5	2,0	14,5	1,5	1,2	0,2	18,0
N.-É.	15,0	18,9	4,6	14,3	5,9	2,7	3,2	24,8
N.-B.	11,2	14,6	3,0	11,6	1,6	0,9	0,7	16,2
Québec	8,6	22,8	4,7	18,1	0,6	0,4	0,2	23,5
Ontario	7,1	26,3	6,3	20,0	4,2	2,4	1,8	30,5
Manitoba	12,5	18,2	3,3	14,9	1,8	1,5	0,3	20,0
Saskatchewan	10,9	14,9	4,2	10,7	1,8	1,6	0,2	16,7
Alberta	6,6	22,1	5,2	17,0	1,7	1,3	0,4	23,8
C.-B.	7,2	28,2	7,5	20,7	5,5	2,0	3,6	33,7
Yukon	17,2	20,6	6,8	13,8	7,3	3,7	3,7	27,9
T.N.-O.	16,2	10,7	1,8	8,9	0,8	0,5	0,3	11,5
Nunavut	13,1	3,5	2,0	1,5	2,0	2,0	0,0	5,5

Le taux de non-réponse final au niveau national est de 23,6%. Il est principalement attribuable aux refus (16%), aux ménages qu'on n'a pas réussi à contacter (5%), et finalement aux ménages pour lesquels les données étaient inutilisables (3%). Les ménages avec des données inutilisables se répartissent à peu près également entre ceux avec des données incomplètes et ceux avec des questionnaires non balancés.

Le taux de non-réponse final varie beaucoup selon les provinces. Le Nunavut et les Territoires du Nord-Ouest décrochent les plus faibles taux avec respectivement 5,5% et 11,5%, Terre-Neuve, l'Île du Prince Édouard, le Nouveau-Brunswick, le Manitoba et la Saskatchewan varient entre 16% et 20% alors que des taux supérieurs à 30% sont observés pour l'Ontario et la Colombie-Britannique. L'Ontario et la Colombie-Britannique affichent un taux de non-réponse à la collecte élevé par rapport aux autres provinces ou territoires. Le taux de données inutilisables varie beaucoup d'une province à l'autre. Ce taux est particulièrement faible au Québec, ainsi qu'aux Territoires du Nord-Ouest à moins de 1%. Ce taux est supérieur à 5% en Nouvelle-Écosse, en Colombie-Britannique et au Yukon.

Le taux de vacance est présenté dans le tableau 2.1 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 peut déceler des problèmes d'identification des logements liés à la collecte. Le taux de vacance de l'EDM de 1998 est de 10,5%, ce qui est légèrement inférieur au taux de l'enquête de l'année précédente qui était d'environ 11%. On peut noter que le taux de vacance de l'EDM de 1998 est légèrement plus bas que le taux de l'Enquête sur la population active (EPA) pour la même période.

2.2 La non-réponse selon le niveau d'urbanisation

La non-réponse varie selon le niveau d'urbanisation. Les divers taux au niveau national sont présentés par niveau d'urbanisation dans le tableau 2.2.

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

Niveau 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 balancé	
URBAIN								
1 000 000 ou plus	4,1	28,6	7,2	21,3	3,2	1,5	1,7	31,7
500 000 à 999 999	3,8	22,9	4,2	18,7	1,4	1,1	0,3	24,3
250 000 à 499 999	5,4	26,3	6,8	19,5	5,7	2,5	3,1	32,0
100 000 à 249 999	7,0	19,8	5,2	14,6	2,7	1,7	1,0	22,5
30 000 à 99 999	5,9	18,5	4,3	14,2	3,0	1,5	1,5	21,5
Moins de 30 000	10,7	16,2	3,8	12,4	2,7	1,8	0,9	18,9
RURAL	22,4	15,1	3,4	11,6	3,1	1,6	1,4	18,2
TOTAL	10,5	20,7	4,9	15,8	2,9	1,6	1,3	23,6

Généralement, le taux de non-réponse croît avec le niveau d'urbanisation. Selon les données du tableau 2.2, il semblerait que le groupe «500 000 à 999 999» habitants brise cette règle avec un taux de non-réponse final de 24%.

Pour le taux de non-réponse à la collecte, il y a une différence notable entre les niveaux d'urbanisation de 250 000 habitants et plus et les autres niveaux avec une différence d'au moins 3%. Les refus compte pour plus de 60% de la non-réponse totale à tous les niveaux d'urbanisation.

En examinant le taux de vacance par niveau d'urbanisation, il est clair que le taux de vacance est beaucoup plus élevé en région rurale (22%) et dans les régions urbaines de moins de 30 000 habitants (11%) que dans les régions urbaines ayant une concentration élevée de population. Ce phénomène s'observe également dans l'EPA et s'explique par un plus grand nombre de logements saisonniers en milieu rural. Cela explique également pourquoi le taux de vacance est plus élevé dans les provinces de l'Atlantique, tel qu'illustré dans le tableau 2.1, principalement à l'Île-du-Prince-Édouard où on observe une plus forte proportion de logements ruraux. Comme l'échantillon de l'EDM est plus concentré dans les régions métropolitaines ayant une concentration élevée de population que l'EPA, on peut s'attendre à observer un taux de vacance national légèrement inférieur.

2.3 La non-réponse selon les strates de revenu

Quoiqu'il ne soit pas possible de comparer le taux de réponse selon le revenu puisque cette information n'est pas disponible pour les non-répondants, le plan d'échantillonnage de l'EPA, utilisé pour l'EDM, a été conçu de sorte à former, dans 7 grandes villes, des strates d'aires géographiques dont le revenu moyen des ménages excède 100 000\$ ainsi que des strates composées d'appartements habités par des ménages ayant un revenu moyen inférieur à 20 000\$. Même si le nombre de telles strates est petit et compte seulement pour un petit nombre de logements dans l'échantillon de l'EDM (environ 280 et 200 pour les revenus élevés et faibles respectivement, soit moins de 2% de l'échantillon), la comparaison des taux de non-réponse dans ces deux groupes par rapport à l'ensemble des autres strates est révélatrice. Le tableau 2.3 présente ces résultats.

Tableau 2.3
Comparaison des taux (%) de non-réponse et de vacance des strates à revenu élevé et à faible revenu 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 balancé	
Revenu élevé	3,6	35,3	8,1	27,2	3,0	2,1	0,9	38,3
Régulières	10,6	20,6	4,8	15,7	2,9	1,6	1,3	23,5
Revenu faible	5,4	16,1	5,4	10,7	1,8	0,6	1,2	17,9
TOTAL	10,5	20,7	4,9	15,8	2,9	1,6	1,3	23,6

Le taux de non-réponse final dans les strates à revenu élevé est presque 65% supérieur à celui observé dans les strates régulières, à un peu plus de 38%. Le taux de refus est d'un peu plus d'un ménage sur quatre (27%), soit un taux presque 75% supérieur aux strates régulières. Seul le taux de non-réponse due aux données non balancées n'est pas plus haut pour la strate à revenu élevé par rapport aux deux autres strates.

Les ménages des strates à faible revenu ont un taux de non-réponse final environ 6 points de pourcentage inférieur à ceux des strates régulières, résultant presque entièrement d'un comportement différent à la collecte puisqu'il y a 5 points de pourcentage de moins au niveau du taux de refus pour les strates à revenu faible par rapport aux strates régulières.

Comme pour l'EDM de 1997, on constate que le taux de vacance est plus élevé pour les strates régulières que les deux autres types de strate, quoique l'écart y est beaucoup moins grand pour cette année par rapport à l'année précédente.

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 définis en fonction des différents niveaux d'urbanisation dans chaque province ou territoire. 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 B.

L'ajustement des poids pour la non-réponse permet de tenir compte des différences au niveau de la non-réponse par niveau d'urbanisation mentionnées dans la section 2.2. 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.

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 noter que dans l'EDM, cette population est différente pour les provinces et les territoires.

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. L'enquête couvre donc près de 98% de la population dans les 10 provinces. Au Yukon, les personnes vivant dans les petites communautés ou dans les régions non organisées sont également exclues et l'enquête couvre environ 81% de la population. La couverture des Territoires du Nord-Ouest correspond à 92% de la population et la couverture au Nunavut est de 89%.⁵

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.

⁵ En terme de ménages, la couverture du Yukon, des Territoires du Nord-Ouest et de Nunavut correspondent respectivement à 80%, 93% et 90% des ménages.

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 interviewers (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 l'identification des personnes qui sont membres du ménage sélectionné lors de la collecte des données. Ces erreurs contribuent également au sous-dénombrement et au surdénombrement.

Une bonne représentation de la population cible est indispensable à la production d'estimation de dépenses réalistes. Le nombre de personnes par ménage est également une caractéristique importante dans l'estimation des dépenses moyennes des ménages. Il est donc nécessaire que l'échantillon, non seulement représente adéquatement les individus de la population cible, mais également la distribution des ménages selon leur taille. L'EDM 1998 a fait l'objet d'une repondération à l'aide d'une nouvelle méthode de pondération introduite pour l'enquête de 1999. Cette méthode utilise de nouveaux groupes de contrôle afin de mieux corriger la représentativité de la population visée.

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. Le taux de glissement (voir annexe B) représente le pourcentage d'écart entre les estimations de l'enquête calculées en utilisant des poids non ajustés avec les données externes⁶ et les estimations provenant de données auxiliaires. Un taux positif correspond à un surdénombrement du nombre de personnes ou de ménages dans l'enquête. Les taux de glissement par groupe d'âge-sexe au niveau national puis au niveau provincial et territorial sont présentés dans les tableaux 3.1 et 3.2 alors que les taux de glissement selon les tailles de ménages que l'on utilise lors de l'ajustement des poids se trouvent dans le tableau 3.3.

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

Tableau 3.1
Taux de glissement nationaux selon le groupe d'âge croisé avec le sexe

Taux de glissement (%) par groupe d'âge-sexe croisé				
	Âge	Sexe		Total
		Homme	Femme	
Canada	0-6 ans	7,4	7,3	7,4
	7-17 ans	6,7	0,4	3,6
	18-24 ans	-12,0	-12,17	-12,1
	25-34 ans	-11,5	-6,5	-9,0
	35-54 ans	-8,0	-6,3	-7,1
	55-59 ans	-13,9	-7,8	-10,8
	60-64 ans	-10,7	-8,2	-9,4
	65-69 ans	-9,7	-6,8	-8,2
	70 ans et +	-7,7	-10,3	-9,3
	Total		-5,7	-5,3

Pour l'EDM de 1998, le taux de sous-dénombrement de la population était 5,5 %. Ce taux variait de 1% à 12% dans les provinces et atteignait un niveau beaucoup plus élevé dans les territoires (17%, 20% et 20% pour le Yukon, les Territoires du Nord-Ouest et le Nunavut respectivement). Au niveau des groupes d'âge, on constate que les taux de glissement pour les enfants (0 à 6 ans et 7 à 17 ans) sont très différents de ceux des autres groupes d'âge. En effet, on observe un sur-représentation des enfants alors que chez les adultes il y a toujours sous-représentation. Les taux les plus élevés au niveau national sont observés chez les 18 à 34 ans et les 55 à 64 ans pour les hommes, et chez les 18 à 24 ans et les 70 ans et plus pour les femmes. Si on ne tient pas compte du sexe, les taux les plus élevés se situent dans les groupes 18-24 et 55-59. On observe également que les taux de glissement sont généralement différents selon le sexe

Au niveau provincial (tableau 3.2), on observe un sous-dénombrement total légèrement inférieur au Nouveau-Brunswick, au Québec, en Ontario et au Manitoba comparativement au sous-dénombrement observé au niveau national, alors que l'inverse est observé pour les autres provinces et pour les territoires. On doit toutefois 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 au Manitoba (-1,2 %) cache les deux pires cas de surdénombrement provincial par groupe âge-sexe (20,7 % chez les femmes de 17-24 et 18,2 % chez les hommes de 0-6 ans). 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.

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

Taux de glissement (%) par groupe d'âge-sexe croisé														
Sexe	Âge	Terre-Neuve	Île du Prince-Édouard	Nouvelle-Écosse	Nouveau-Brunswick	Québec	Ontario	Manitoba	Saskatchewan	Alberta	Colombie-Britannique	Yukon	Territoires du Nord-Ouest	Nunavut
Homme	0-6	12,3	-27,9	4,1	4,9	9,8	13,3	4,4	4,1	-6,1	-0,1	*	-13,7	-15,5
	7-17	-13,8	-5,5	-6,6	14,2	7,8	11,6	20,7	0,5	3,9	-4,9			
	18-24	-27,0	-24,9	-25,8	-14,6	1,8	-14,0	-9,6	-23,4	-12,5	-23,0			
	25-34	-14,0	-24,2	-18,9	-21,5	-6,6	-10,9	-8,1	-11,6	-18,0	-13,6			
	35-54	-17,5	-5,0	-4,7	-0,4	-9,5	-4,4	-6,3	-12,7	-9,5	-14,5	-28,5	-24,4	-26,4
	55-59	2,3	-18,4	-19,7	-21,1	-16,5	-9,1	-7,5	-13,7	-31,8	-11,7			
	60-64	-6,6	19,8	4,2	3,6	-10,1	-12,3	-3,8	5,9	-9,2	-21,8			
	65-69	9,2	-12,4	-5,7	-11,3	-7,7	-15,5	14,2	4,5	3,2	-17,6			
	70 +	-22,6	-1,7	-2,6	-11,9	-1,4	-11,5	-5,9	-8,4	-9,9	-5,9			
Total	-13,1	-11,6	-8,6	-4,4	-3,6	-3,6	-0,7	-8,1	-9,1	-12,2	*	-21,0	-21,7	
Femme	0-6	-4,3	-9,6	-8,6	14,5	3,1	13,2	18,2	5,9	-3,8	7,2	*	-12,7	-13,0
	7-17	-7,9	-10,8	-3,7	1,0	4,4	3,6	-3,5	1,3	-10,5	-4,0			
	18-24	-28,2	-28,3	-22,2	-12,8	-18,7	-4,5	-11,8	-5,4	-10,9	-18,9			
	25-34	-6,6	-6,5	-25,6	-2,1	0,8	-7,1	-9,2	-7,5	-6,7	-12,9			
	35-54	-14,9	-11,5	-1,8	0,3	-6,6	-4,4	0,3	-10,9	-9,9	-10,0	-16,1	-20,3	-21,1
	55-59	-1,6	16,8	-12,1	4,8	-7,7	-1,9	0,2	-8,0	-17,6	-23,7			
	60-64	5,8	7,8	-11,1	-8,9	-13,8	-4,0	-8,1	-2,3	-10,0	-11,1			
	65-69	6,2	-15,0	-7,5	2,7	-1,7	-6,1	1,1	-8,6	4,2	-30,8			
	70 +	-11,9	-4,2	-6,8	-5,6	-11,1	-17,5	-1,8	5,1	-8,7	1,0			
Total	-10,8	-9,5	-9,5	-0,8	-5,0	-3,3	-1,7	-4,0	-8,8	-9,4	*	-17,8	-17,5	
Total		-11,9	-10,5	-9,1	-2,6	-4,3	-3,4	-1,2	-6,0	-9,0	-10,8	-17,1	-19,5	-19,7

* Le taux de glissement pour les enfants de 0-17 ans au Yukon est de -2,3%. Ce taux n'est pas calculé selon le sexe au Yukon pour ce groupe d'âge.

On observe une sous-estimation du nombre de ménages de 5,1 % au niveau national (tableau 3.3). Ce taux de glissement est comparable au taux de glissement au niveau des personnes qui est de -5,5 %. On constate que la sous-estimation des ménages d'une personne est plus élevée alors que celle des ménages de deux personnes est légèrement plus faible que celle des ménages de taille 3 ou plus.

Le Manitoba (0,1 %) et le Nouveau-Brunswick (-1,5 %) présentent les meilleures couvertures en termes de ménages. En ce qui a trait aux ménages d'une personne, le Yukon (-30,7 %), l'Ontario (-9,4 %) et le Québec (-7,1%) sont les seuls à présenter un taux de sous-dénombrement plus élevé que le taux national.

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

Province ou territoire	Taux de glissement (%)			
	Nombre de ménages	Nombre de ménages d'une personne	Nombre de ménages de deux personnes	Nombre de ménages de trois personnes et plus
Canada	-5,1	-6,8	-4,0	-4,8
Terre-Neuve	-8,4	-6,3	-1,7	-12,8
Île du Prince Edouard	-3,9	-1,2	7,6	-12,9
Nouvelle-Écosse	-5,6	-5,9	1,8	-11,1
Nouveau-Brunswick	-1,5	-2,0	1,0	-3,0
Québec	-4,7	-7,1	-4,0	-3,6
Ontario	-4,2	-9,4	-5,8	-0,6
Manitoba	0,1	0,9	2,5	-2,3
Saskatchewan	-3,8	-3,4	1,0	-7,9
Alberta	-6,3	-5,1	-3,1	-9,1
Colombie-Britannique	-8,9	-5,2	-5,9	-13,6
Yukon	-16,6	-30,7	-9,3	-13,1
Territoires du Nord-Ouest	-7,6	-1,6	10,2	-16,4
Nunavut	-32,4		-32,4 ⁷	

3.2 L'ajustement pour l'erreur de couverture au niveau de la population et du nombre de ménages

Pour corriger le problème de couverture illustré dans les tableaux 3.1 et 3.2 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-sexe définis dans ces tableaux, et ce pour chaque province ou territoire. Pour plus de détails sur la méthodologie de l'ajustement, voir la référence [6]. Cet ajustement réduit considérablement le biais engendré par les erreurs de couverture mais il n'élimine pas complètement le biais 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 ou un territoire.

On notera également que l'efficacité de l'ajustement de couverture 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 1996 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

⁷ Pour le Nunavut, seul le nombre total de ménages est utilisé.

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. On notera également que l'efficacité de l'ajustement dépend de la qualité des données auxiliaires.

En plus des estimations démographiques au niveau des groupes âge-sexe par province et territoire, trois 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. Le second est conçu pour contrôler le nombre de ménages monoparentaux et de couples avec enfants par province. Finalement, 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 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 interviewers 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 10% 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 interviewers suggèrent généralement aux répondants d'estimer la fréquence des achats et le prix généralement payé pour dériver les dépenses sur une période de douze 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 moyen nécessaire pour compléter l'entrevue pour les ménages répondants qui ont fourni des données utilisables est de 1 heure 50 minutes. 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 5 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 cherche à les combattre en implantant des techniques de sondage qui visent à les réduire.

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. Cette 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. La saisie des données est vérifiée à l'aide d'un système de vérification automatisé qui consiste à grouper les questionnaires par lots et à choisir, dans chacun de ces lots, un échantillon de questionnaires qui seront saisis une seconde fois. Toutes les erreurs relevées sont corrigées. Si le nombre d'erreurs dépasse un certain seuil, le lot complet est soumis à la saisie à nouveau.

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 de cohérence essentielles 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 qui se répondent par 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 catégoriques qu'on trouve principalement dans les sections sur les caractéristiques et l'équipement du logement, sont imputées 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.

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 variable 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 des receveurs (les ménages qui ont 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 généralement parti des variables d'appariement pour presque toutes les sections portant sur les dépenses. Dans le choix du donneur, on tient compte également de la contrainte que le receveur doit satisfaire après l'imputation de ces données manquantes, selon certaines règles de cohérence. 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.

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, le taux d'imputation donne une indication de l'importance des non-réponses partielles. La mesure de l'impact des valeurs imputées sur les estimations globales est aussi une bonne indication du biais potentiel des estimations. Ces indicateurs de la qualité sont présentés dans les sections suivantes.⁸

5.1 Mesure d'impact

Dans l'EDM, la non-réponse partielle correspond principalement à des répondants qui déclarent avoir effectué un certain type de dépenses mais qui n'en connaissent pas la valeur totale pour la période de référence demandée. Les valeurs imputées doivent donc être strictement positives. Pour des types de dépenses ou des sources de revenu peu fréquentes, où une large proportion de ménages déclarent une valeur de 0, la proportion de la valeur totale de l'estimation provenant de données imputées devient une mesure plus pertinente de l'effet de l'imputation. Cette mesure, appelée ici mesure d'impact de l'imputation, est définie par le total pondéré des données imputées divisé par l'estimation totale (sommées des valeurs pondérées). Une description algébrique de la mesure d'impact est donnée à l'annexe B.

5.1.1 Mesures d'impact des sources de revenu et des dépenses par section

L'impact de l'imputation offre l'avantage qu'il peut être calculé pour les dépenses agrégées. L'impact sur le revenu total est d'environ 0,2% tandis que pour les dépenses totales, il se situe à environ 0,4%. Les impacts les plus grands pour les dépenses sont sur les dépenses en vêtements, les dépenses diverses, les impôts personnels et les paiements d'assurance et cotisations de retraite qui ont des impacts respectifs d'environ 0,7%, 1,3%, 0,9% et 0,6%. Pour le reste des catégories de dépenses, les impacts sont inférieurs à 0,5%.

⁸ Pour des raisons opérationnelles, ces indicateurs de qualité ne sont pas disponibles pour les données catégoriques comme les caractéristiques et l'équipement du logement pour l'EDM de 1998.

5.2 Proportion de ménages ou personnes imputés au niveau national et provincial

Une première indication de l'ampleur de la non-réponse partielle est la proportion de ménages nécessitant de l'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 cette dernière catégorie, 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) 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 et le pourcentage de personnes pour lesquelles au moins une variable de revenu a été imputée. Les résultats sont fournis à la fois au niveau national et provincial ou territorial. Cela fournit une indication des provinces ou territoires qui sont les plus affectées par l'imputation en comparant chacune d'elle aux autres provinces ainsi qu'aux résultats au niveau national.

5.2.1 Imputation de dépenses du ménage par province ou territoire

Le pourcentage de ménages utilisables pour lesquels il a fallu imputer au moins une variable de dépenses (excluant les dépenses vestimentaires) figure dans le tableau 5.1. Les ménages utilisables correspondent à tous les ménages échantillonnés à l'exception de ceux qui n'ont pu être contactés, qui ont refusé, qui ont fourni des données incomplètes ou qui ne balancent pas selon les définitions de la section 2. Le tableau est réparti en fonction du nombre de variables qui ont été imputées (parmi les 237) pour un ménage.

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

Province ou territoire	Ménages (%) qui nécessitent l'imputation de VARIABLES DE DÉPENSES (sauf pour les dépenses en vêtements)			
	Nombre de variables imputées (parmi 237)			TOTAL
	1	2	3 ou plus	
Canada	4,7	0,8	0,4	5,9
Terre-Neuve	2,2	0,2	0,1	2,5
Î.-P.-É.	2,9	0,5	0,2	3,6
N. -É.	5,1	0,9	0,9	6,9
N. -B.	2,8	0,2	0,0	3,0
Québec	1,7	0,3	0,0	2,0
Ontario	8,0	2,4	1,5	11,9
Manitoba	4,9	0,5	0,6	6,0
Saskatchewan	3,5	0,6	0,1	4,2
Alberta	5,1	0,6	0,2	5,9
C.-B.	6,4	0,9	0,4	7,7
Yukon	11,8	0,3	0,0	12,1
T. N.-O.	9,5	0,0	0,0	9,5
Nunavut	6,8	2,6	1,1	10,5

Le tableau 5.1 montre qu'il a fallu imputer des dépenses (mis à part la section sur les vêtements) pour 6% des ménages au niveau national mais pour un peu plus de 75% d'entre eux, seulement une variable a été imputée. Il y a très peu de ménages au niveau national qui ont eu plus de 1 variable imputée (1,2%). Les résultats au niveau provincial sont généralement légèrement supérieurs ou inférieurs au taux national de 5,9% mais trois provinces ou territoires ont plus de 10% de ménages ayant au moins une imputation, ce sont l'Ontario (11,9%), le Yukon (12,1%) et le Nunavut (10,5%). Le Québec et les provinces de l'atlantique à l'exception de la Nouvelle-Écosse ont les plus bas têt d'imputation étant tous inférieurs à 4%.

5.2.2 Imputation de dépenses vestimentaires et de 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, par une procédure à deux degrés. 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 [6] 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 trouve 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 lesquels seul le revenu total était disponible et qui nécessitaient une imputation de toutes les composantes. Le pourcentage total de personnes pour lesquelles une forme d'imputation de revenu a été effectuée est également

donné. La 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.

Tableau 5.2
Personnes qui nécessitent l'imputation du revenu et personnes qui nécessitent l'imputation de dépenses en vêtements par province

Province ou territoire	Pourcentage de personnes qui nécessitent l'imputation de VARIABLES DE REVENU				Pourcentage de personnes qui nécessitent l'imputation de VARIABLES DE DÉPENSES VESTIMENTAIRES (sur 11)
	1 variable imputé	2 variables imputées ou plus (pas toutes)	Toutes les variables imputées (revenu total connu)	TOTAL (toute forme d'imputation)	
Canada	0,8	0,1	1,6	2,6	15,2
Terre-Neuve	0,4	0,0	1,4	1,8	8,5
I.-P.-E.	0,0	0,1	1,5	1,6	21,2
N.-É.	0,7	0,1	2,8	3,6	12,4
N.-B.	0,1	0,0	3,3	3,5	18,1
Québec	0,5	0,1	0,8	1,3	25,5
Ontario	1,0	0,1	1,4	2,5	11,0
Manitoba	4,1	0,2	2,0	6,3	13,7
Saskatchewan	0,4	0,0	0,7	1,1	14,6
Alberta	0,0	0,1	1,8	2,0	12,5
C.-B.	0,5	0,1	2,2	2,7	15,8
Yukon	0,2	0,0	0,2	0,4	12,0
T. N.-O.	0,0	0,4	0,2	0,5	13,9
Nunavut	4,9	0,0	0,1	5,1	15,7

À partir de ces résultats, on constate qu'il y a moins de 3% des personnes des ménages utilisables pour lesquelles il a fallu imputer au moins une variable de revenu. Pour un peu plus de 60% d'entre eux, le revenu total était fourni par le répondant mais toutes les composantes ont dû être imputées. Pour une très grande partie des autres personnes qui ont nécessité de l'imputation, seulement une composante de revenu (une variable) a été imputée. Au niveau provincial et territorial, le pourcentage de personnes nécessitant l'imputation d'au moins une variable de revenu est pareillement peu élevé, variant entre un maximum d'un peu plus de 6% et 5% pour le Manitoba et le Nunavut respectivement et un minimum d'environ un demi pourcent pour le Yukon et les Territoires du Nord-Ouest.

À partir de la dernière colonne du tableau, on constate qu'un peu plus de 15% des personnes nécessitent une forme quelconque d'imputation pour les variables des dépenses vestimentaires. Les taux provinciaux et territoriaux sont tous près du niveau national sauf pour Terre-Neuve, qui est moins élevé avec un taux d'environ 9%, et pour l'Île-du-Prince-Édouard et le Québec, qui ont des taux atteignant respectivement environ 21.2% et 25.5%. Presque toutes ces personnes ont fourni le total des dépenses vestimentaires mais il a fallu imputer les composantes. La proportion des personnes pour lesquels il a fallu imputer le total n'est pas présenté dans ce tableau mais elle représente moins de 1% des personnes imputées dans chaque province ou territoire. Le niveau élevé

d'imputation effectué 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.

BIBLIOGRAPHIE

- [1] Méthodologie de l'enquête sur la population active du Canada, No. 71-526-XPB au catalogue
- [2] Enquête sur la population active, Rapport sur la qualité, Enquêtes 0798 à 1298, Division des méthodes d'enquêtes auprès des ménages.
- [3] Lemaître, G. et Dufour, J. (1987). Une méthode intégrée de pondération des personnes et des familles, *Technique d'enquête*, Vol.13, n 2, pp.211-220, Statistique Canada
- [4] The Nation: 1996 Census of Population 93F0030XDB96009, Statistics Canada, Electronic Shelf.
- [5] The Nation: 1996 Census of Population 93F0029XDB96002, Statistics Canada, Electronic Shelf.
- [6] Tremblay, J. et Arsenault, S. (2001). La méthodologie de l'enquête sur les dépenses des ménages, Division des méthodes d'enquêtes auprès des ménages, Statistique Canada.
- [7] Wolter, K.M. (1985). Introduction to Variance Estimation, Springer-Verlag New-York Inc.
- [8] Les habitudes de dépenses au Canada (1998), No. 62-202-XPB au catalogue.

ANNEXE A

Liste de tableaux détaillés disponibles sur demande

1. Taux de non-réponse (non-contact, refus, données non utilisables) selon le niveau d'urbanisation par province ou territoire
2. Taux de non-réponse (non-contact, refus, données non utilisables) selon les communautés pour les territoires
3. Impact des données imputées sur les estimations pour toutes les variables de dépenses
4. Impact des données imputées sur les estimations pour toutes les variables de revenu par province
5. Taux d'erreur de la règle de suppression dans les régions métropolitaines

ANNEXE B

Notation algébrique

1. Notation pour les poids

B_k^{-1} : *Poids de sondage* : l'inverse de la probabilité d'inclusion pour un ménage k.

w_k^{NR} : *Sous-poids* : poids ajusté pour la non-réponse

w_k^f : *Poids final* : poids ajusté pour la non-réponse et les contrôles (estimateur de régression)

2. Ajustement pour la non-réponse

Les poids ajustés pour la non-réponse d'un ménage k, noté w_k^{NR} , sont

$$w_k^{NR} = p_k^{-1} * \frac{1}{\text{taux}_g} \quad \text{avec} \quad \text{taux}_g = \frac{\sum_{k=1}^{n_{r,g}} p_k^{-1}}{\sum_{k=1}^{n_{s,g}} p_k^{-1}}$$

Où

$n_{r,g}$ est le nombre de répondants dans le groupe de non-réponse g,

$n_{s,g}$ est le nombre de ménages éligibles dans l'échantillon dans le groupe de non-réponse g, et

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

3. Calcul du taux de glissement

$$\text{taux}_c = 100 * \frac{w_c^{NR} - t_c}{t_c}$$

Où

w_c^{NR} est le sous-poids ajusté pour la non-réponse pour le groupe de contrôle c,

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

4. Calcul de la mesure d'impact de l'imputation sur les estimations

La mesure d'impact de l'imputation sur les estimations est définie par

$$I = \frac{\sum_{k=1}^{n_{s,I}} w_k^f \hat{y}_k}{\sum_{k=1}^{n_{s,R}} w_k^f y_k + \sum_{k=1}^{n_{s,I}} w_k^f \hat{y}_k}$$

Où

$n_{s,R}$ est le nombre de ménages utilisables qui ont fourni une valeur pour la caractéristique y ,

$n_{s,I}$ est le nombre de ménages utilisables pour lesquels une valeur a été imputée pour la caractéristique y

y_k est la valeur fournie par le $k^{\text{ième}}$ ménage,

y_k^f est la valeur imputée pour le $k^{\text{ième}}$ ménage, et

w_k^f est le poids final du $k^{\text{ième}}$ ménage.