



**Document de recherche**

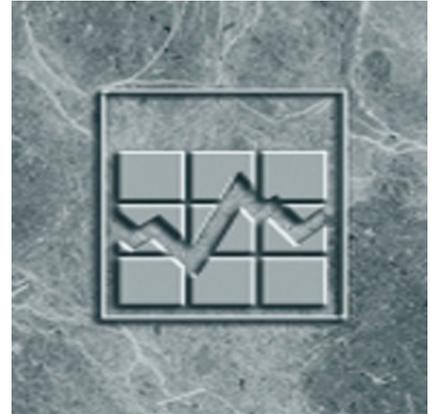
# Enquête sur les dépenses des ménages de 2000 Indicateurs de qualité des données

by Jean-Claude Boies, Sylvie DeBlois, José Gaudet,  
Bruno Lapierre, Christian Nadeau

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

Telephone: 613 951-7355

Toutes les opinions émises par l'auteur de ce document ne reflètent pas nécessairement celles de Statistique Canada.



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](mailto: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.

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

## Renseignements sur les commandes et les abonnements

### Les prix ne comprennent pas les taxes de vente

On peut se procurer ce produit n° 62F0026MIF2002001 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](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 2000 Indicateurs de qualité des données

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2002

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.

juin 2002

N° 62F0026MIF2002001 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.*

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



## Résumé

Ce document présente les indicateurs de qualité produits pour l'Enquête sur les dépenses des ménages de 2000. On y trouve les indicateurs de qualité usuels utiles aux utilisateurs pour l'interprétation des données, tels que les coefficients de variation, les taux de non-réponse, les taux de glissement et les taux d'imputation.

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.

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



# TABLE DES MATIÈRES

<b>FAITS SAILLANTS .....</b>	<b>9</b>
<b>INTRODUCTION .....</b>	<b>11</b>
<b>1. LES ERREURS D'ÉCHANTILLONNAGE .....</b>	<b>12</b>
1.1 Les mesures de l'erreur d'échantillonnage .....	12
1.2 Les coefficients de variation .....	12
Tableau 1.1	
Coefficients de variation (%) par province ainsi qu'au niveau national pour l'estimation de la moyenne des dépenses par ménage pour quelques catégories de dépenses sommaires et pour l'estimation du revenu moyen. ....	13
Tableau 1.2	
Coefficients de variation (%) par province ainsi qu'au niveau national pour quelques catégories des caractéristiques des logements et l'équipement ménager .....	14
1.3 Modèle pour dériver une approximation du CV .....	14
1.4 Suppression des données non fiables dans les tableaux d'estimations .....	16
<b>2. LA NON-RÉPONSE .....</b>	<b>16</b>
2.1 Les taux de non-réponse et les taux de vacance .....	17
Tableau 2.1	
Taux de non-réponse (%) et taux de vacance (%) par province ainsi qu'au niveau national .....	18
2.2 La non-réponse selon le niveau d'urbanisation .....	19
Tableau 2.2	
Taux de non-réponse (%) et taux de vacance (%) par niveau d'urbanisation ..	19
2.3 La non-réponse selon les strates de revenu .....	19
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 .....	20
2.4 L'ajustement pour la non-réponse .....	20
<b>3. ERREURS DE COUVERTURE .....</b>	<b>21</b>
3.1 Le sous-dénombrement et le surdénombrement : les taux de glissement .....	21
Tableau 3.1	
Taux de glissement nationaux selon le groupe d'âge-sexe .....	22
Tableau 3.2	
Taux de glissement pour les provinces selon le groupe d'âge-sexe .....	23
Tableau 3.3	
Taux de glissement pour les provinces selon la taille de ménage .....	24
3.2 L'ajustement pour l'erreur de couverture au niveau de la population et du nombre de ménages .....	24
<b>4. LES ERREURS DE RÉPONSE .....</b>	<b>25</b>

## TABLE DES MATIÈRES (fin)

<b>5. LES ERREURS DE TRAITEMENT .....</b>	<b>26</b>
5.1 Proportion de ménages ou personnes nécessitant imputation, au niveau national et provincial.....	28
Tableau 5.1	
Ménages qui nécessitent l'imputation de dépenses par province .....	29
Tableau 5.2	
Personnes qui nécessitent l'imputation du revenu et personnes qui nécessitent l'imputation de dépenses en vêtements par province .....	30
<b>BIBLIOGRAPHIE.....</b>	<b>31</b>
<b>ANNEXE A</b>	
<b>Notation algébrique .....</b>	<b>32</b>

# FAITS SAILLANTS

## Les erreurs d'échantillonnage

- Les coefficients de variation (CV) de l'estimation de la moyenne des dépenses totales par ménage varient entre 1,2 % et 1,6 % selon les provinces et il est de 0,8 % au niveau national.
- Les coefficients de variation (CV) de l'estimation de la moyenne pour les différentes catégories de dépenses sommaires sont inférieurs ou égaux à 1,8 % au niveau national, et généralement inférieurs à 5 % au niveau provincial. Les résultats pour les différentes catégories de caractéristiques des logements et l'équipement ménager sont similaires.

## La non-réponse

- Le taux de non-réponse est de 29,8 %. La non-réponse est due aux refus (18,5 %), aux ménages qu'on n'a pas réussi à contacter (7,5 %) et aux ménages dont les données ont été considérées inutilisables (3,8 %).
- Le taux de non-réponse final a tendance à croître avec le niveau d'urbanisation. On observe un taux de non-réponse de 23,6% en milieu rural, et un taux de 37,6% dans les centres urbains d'un million d'habitants et plus. Cette tendance s'observe également dans les taux de non-réponse à la collecte.
- L'analyse des taux de réponse finaux 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 (44,2 %) est plus élevé que ceux observés dans les strates régulières (29,4 %) et dans les strates à faible revenu (28,9 %).

## Les erreurs de couverture

- On observe un sous-dénombrement des ménages de 8,1 % au niveau national. On observe aussi un sous-dénombrement des ménages pour chacune des provinces, les taux variant de 5,2 % à 13,2 %.
- On observe un sous-dénombrement des personnes de 9,9 % au niveau national. On observe aussi un sous-dénombrement des personnes pour chacune des provinces, les taux variant de 6,6 % à 15,4 %.
- Les taux de glissement nationaux pour les enfants (0 à 6 ans et 7 à 17 ans) sont très différents de ceux des autres groupes d'âge. En effet, chez les enfants, on observe un surdénombrement ou un léger sous-dénombrement, alors que chez les adultes il y a toujours sous-dénombrement. Le taux de sous-dénombrement pour l'ensemble des enfants est de 1,8 %, alors qu'il est de 12,2 % pour les adultes.

## **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 total d'entrevue varie selon les caractéristiques des ménages et peut prendre plus de cinq 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 (excluant les dépenses vestimentaires) pour 8,9 % des ménages. Pour la majorité d'entre eux, seulement une ou deux variables parmi les 237 variables de dépenses ont été imputées.
- Environ 17 % 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.

### **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 70 % d'entre eux, le revenu total était fourni par le répondant mais toutes les composantes (sources de revenu) ont été imputées.

# 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 2000 est composé de 24 180 logements<sup>1</sup> répartis dans les 10 provinces<sup>2</sup>. 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<sup>er</sup> janvier au 31 décembre de l'année précédente. Suite à 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 2000. 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 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. Finalement, 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].

---

<sup>1</sup> De ces 24 180 logements, on doit identifier et enlever les logements inadmissibles (voir section 2.1) pour en arriver à 20 989 ménages auprès desquels on collecte les dépenses et les revenus.

<sup>2</sup> Les territoires n'ont pas été couverts par l'EDM 2000 mais le seront en 2001. Ils seront par la suite couverts à tous les deux ans.

# 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 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  $\pm 2$  erreur-type ( $Y \pm 2 SE$ )<sup>3</sup>. 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.

## 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 aux niveaux national et provincial dans la publication *Guide de l'utilisateur – Enquête sur les dépenses des ménages* (référence [3]).

---

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

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 au niveau provincial ainsi qu'au niveau national 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'au niveau national 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.	Sask.	Alb.	C.-B.
Dépenses totales	0,8	1,4	1,6	1,5	1,2	1,6	1,6	1,3	1,6	1,2	1,6
Dépenses de consommation	0,6	1,3	1,6	1,5	1,2	1,2	1,1	1,2	1,5	1,2	1,4
Alimentation	0,6	1,1	1,6	1,1	1,2	1,1	1,2	1,2	1,9	1,2	1,1
Logement	0,8	1,6	2,2	1,8	1,7	1,7	1,5	1,7	1,9	1,5	2,2
Entretien ménager	0,9	1,7	2,1	2,0	2,1	2,1	1,8	1,7	1,9	2,0	1,9
Ameublement	1,8	3,7	5,3	4,8	4,5	3,6	3,4	3,6	3,8	3,6	3,5
Habillement	1,1	2,8	3,6	2,9	2,4	2,1	2,2	2,6	3,6	2,4	2,1
Transport	1,5	3,2	4,7	3,6	3,1	3,1	2,9	3,6	3,7	3,0	3,2
Soins de santé	1,3	3,1	3,4	3,1	2,9	2,4	3,2	2,5	4,6	2,4	2,6
Soins personnels	1,0	2,0	3,2	2,2	2,1	2,0	2,0	2,6	2,6	2,1	2,1
Loisirs	1,4	3,1	3,9	5,6	4,0	2,9	2,7	2,7	3,7	3,7	3,5
Lecture et autres imprimés	1,5	3,8	3,9	3,7	3,6	3,3	2,9	3,5	3,7	3,1	3,0
Éducation	3,3	6,4	10,8	11,0	10,4	6,0	6,2	7,2	7,5	7,5	6,9
Tabac, boissons alcoolisées	1,7	3,7	5,0	4,8	4,2	3,4	3,6	4,1	4,7	3,8	3,8
Jeux de hasard (net)	4,9	8,2	14,0	7,8	7,9	11,1	9,2	6,6	11,5	11,4	5,5
Dépenses diverses	2,6	7,0	11,0	6,9	11,4	5,2	5,2	6,5	5,4	5,8	6,9
Impôts personnels	1,8	2,9	4,1	3,1	2,3	2,9	3,6	3,2	2,9	2,3	3,7
Assurances personnelles	6,1	2,5	2,9	2,3	2,3	1,6	13,7	3,0	2,6	6,1	2,8
Dons et contributions	5,3	5,1	12,7	12,2	10,8	17,4	9,3	5,8	7,5	9,6	10,9
Revenu	0,9	1,0	1,6	1,2	1,0	1,1	1,9	1,1	0,9	0,8	1,0

Les coefficients de variation de l'estimation de la moyenne des dépenses totales par ménage varient entre 1,2 % et 1,6 % selon les provinces et il est de 0,8 % au niveau national.

Pour les différentes catégories de dépenses sommaires, les CV au niveau national sont inférieurs ou égaux à 1,8 % à l'exception des dépenses dans les catégories suivantes: *éducation, jeux de hasard, dépenses diverses, paiements d'assurance individuelle et*

cotisations de retraite ainsi que dons en argent et contributions. Ces catégories de dépenses représentent respectivement 1,5 %, 0,5 %, 1,5 %, 5,6% et 2,3 % du total des dépenses (données non présentées). De plus, si on fait exception de ces catégories, les CV sont généralement inférieurs ou égaux à 5% au niveau provincial.

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

**Tableau 1.2**  
**Coefficients de variation (%) par province ainsi qu'au niveau national 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.
Propriétaire	1,0	1,8	2,9	2,1	1,6	2,2	2,1	1,8	2,0	1,8	1,8
Locataire	1,8	5,3	6,8	4,9	4,6	2,8	4,0	4,4	4,9	4,6	3,3
Machine à laver (automatique)	0,7	1,2	2,5	1,6	1,2	1,0	1,7	1,5	1,3	1,1	1,4
Sécheuse	0,7	1,2	2,5	1,9	1,2	1,1	1,8	1,6	1,2	1,2	1,4
Lave-vaisselle encastré	1,4	4,3	6,2	4,4	3,7	2,5	3,3	3,3	4,0	2,5	2,4
Congélateur	1,0	1,3	2,6	2,2	1,8	2,5	2,2	1,6	1,6	1,7	2,1
Four à micro-ondes	0,4	1,0	1,3	0,9	0,8	0,9	0,8	0,9	0,9	0,6	0,8
Téléphone cellulaire	1,4	4,5	5,7	4,4	4,2	4,0	2,6	3,3	3,5	2,5	2,7
Lecteur de disques compacts	0,6	1,7	2,1	1,8	1,7	1,4	1,3	1,8	1,8	1,2	1,3
Télédistribution	1,0	1,6	4,3	2,2	2,5	2,1	2,1	1,8	2,4	2,0	1,3
Ordinateur personnel	1,0	3,0	4,3	2,7	3,3	2,4	1,9	2,4	2,6	2,0	1,7
Modem	1,2	3,7	5,0	3,2	3,7	2,9	2,2	2,9	3,2	2,5	2,1
Utilisation d'Internet (maison)	1,3	4,0	5,5	3,4	4,2	3,3	2,4	3,6	3,9	2,7	2,3
Véhicules possédés (un)	1,4	3,3	4,5	3,4	3,6	3,0	2,7	3,2	3,1	3,2	2,7
Véhicules possédés (2 ou plus)	1,5	4,6	3,3	3,5	3,1	3,8	3,2	3,4	2,9	2,5	2,7

Les coefficients de variation pour les caractéristiques des logements et l'équipement ménager sont généralement inférieurs à 5% pour les provinces de l'Atlantique et généralement inférieurs à 4% pour les autres provinces. La catégorie *locataire* fait cependant exception à cette règle, avec plusieurs CV provinciaux dépassant ces seuils.

Les CV pour les caractéristiques des logements et l'équipement ménager au niveau national sont inférieurs ou égaux à 1,8 %.

### 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  $P$  au niveau national 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 au niveau provincial puisque les CV provinciaux sont publiés pour l'EDM de 2000 (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 2000. 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 0<sup>4</sup>.

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.

---

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

## 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 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.

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 équilibrés. On y trouve également le taux de vacance. Ces taux sont fournis au niveau national ainsi qu'au niveau provincial.

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

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	
<b>Canada</b>	11,0	26,0	7,5	18,5	3,8	2,2	1,6	29,8
Terre-Neuve	19,1	21,0	8,8	12,2	2,0	0,8	1,2	22,9
Î.-P.-É.	20,0	14,9	4,3	10,6	0,5	0,1	0,4	15,4
N.-É.	14,1	27,6	9,3	18,4	6,6	3,4	3,2	34,3
N.-B.	13,6	23,9	5,7	18,2	3,7	2,1	1,7	27,7
Québec	7,0	25,9	6,2	19,7	1,1	0,7	0,4	26,9
Ontario	7,0	34,3	10,1	24,2	5,4	2,1	3,3	39,7
Manitoba	12,3	18,8	4,4	14,4	3,2	2,0	1,1	22,0
Saskatchewan	10,5	23,7	7,6	16,1	3,5	1,9	1,6	27,2
Alberta	7,0	29,3	6,8	22,5	2,5	1,3	1,2	31,8
C.-B.	8,9	26,9	8,4	18,5	6,3	5,5	0,8	33,2

Le taux de non-réponse final au Canada est de 29,8 %. Il est dû aux refus (18,5 %), aux ménages qu'il a été impossible de rejoindre (7,5 %), et finalement aux ménages dont les données étaient inutilisables (3,8%). Pour chacune des provinces, les refus représentent la cause majeure de non-réponse, suivis des non-contacts, puis des données inutilisables.

Les taux de non-réponse finaux varient selon les provinces. Le taux de l'Île-du-Prince-Édouard est inférieur à 20 %, tandis que les taux sont supérieurs à 30 % en Nouvelle-Écosse, en Ontario, en Alberta et en Colombie-Britannique. Le taux de non-réponse de l'Ontario est particulièrement élevé atteignant presque 40%. C'est aussi dans cette province que l'on trouve le plus haut taux de non-contacts (10,1 %) et de refus (24,2 %).

Le taux de vacance est présenté au 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 décèle les problèmes d'identification des logements liés à la collecte. Le taux de vacance de l'EDM de 2000 est de 11,0 %. Le taux de vacance de l'EDM est légèrement plus bas que celui 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<sup>5</sup>.

**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	
<b>URBAIN</b>								
1 000 000 ou plus	5,1	32,8	9,9	22,9	4,8	3,0	1,9	37,6
500 000 à 999 999	4,7	28,2	6,9	21,3	2,6	1,6	1,0	30,8
250 000 à 499 999	6,3	29,4	9,7	19,7	6,5	3,1	3,4	35,9
100 000 à 249 999	7,4	29,8	8,7	21,1	2,7	1,5	1,1	32,5
30 000 à 99 999	4,8	25,1	7,0	18,1	2,8	2,1	0,8	27,9
Moins de 30 000	7,9	20,2	6,2	14,0	3,8	2,3	1,5	24,0
<b>RURAL</b>	23,4	19,7	5,5	14,2	3,8	2,1	1,7	23,6
<b>TOTAL</b>	11,0	26,0	7,5	18,5	3,8	2,2	1,6	29,8

Le taux de non-réponse final augmente généralement avec le niveau d'urbanisation. Selon les résultats du tableau 2.2, il n'y a que le groupe des 500 000 à 999 999 qui ne suit pas cette tendance.

Le taux de non-réponse à la collecte a lui aussi tendance à croître avec le niveau d'urbanisation. Il y a un écart important de 12,6 % 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 trois fois plus élevé en région rurale (23,4 %) que dans les régions urbaines à faible population (7,9 %). 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'EPA et s'explique sans doute par un plus grand nombre de logements saisonniers en milieu rural. Comme l'échantillon de l'EDM est plus concentré dans les régions urbaines à forte population que celui de 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

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 sept grandes villes, des strates d'aires géographiques où le revenu moyen des ménages excède 100 000 \$ ainsi

<sup>5</sup> 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.

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 qu'il compte pour seulement un petit nombre de logements dans l'échantillon de l'EDM (environ 560 et 160 pour les strates de revenus élevés et faibles respectivement, soit moins de 3 % 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 équilibré	
Revenu élevé	3,9	38,5	8,3	30,3	5,6	2,4	3,2	44,2
Régulières	11,2	25,7	7,5	18,2	3,8	2,2	1,6	29,4
Revenu faible	4,8	24,8	12,1	12,8	4,0	3,4	0,7	28,9
<b>TOTAL</b>	<b>11,0</b>	<b>26,0</b>	<b>7,5</b>	<b>18,5</b>	<b>3,8</b>	<b>2,2</b>	<b>1,6</b>	<b>29,8</b>

Dans les strates à revenu élevé, le taux de non-réponse final est de beaucoup supérieur à celui dans les strates régulières, à 44,2 %. Le taux de refus des strates à revenu élevé est de plus de 30%, taux nettement supérieur à ce qui est observé pour les autres types de strates.

Les ménages des strates à faible revenu ont un taux de non-réponse final et un taux de non-réponse à la collecte similaires à ceux des strates régulières. Mais les causes de la non-réponse sont différentes pour les deux types de strates. Pour les strates régulières, la non-réponse à la collecte est majoritairement due aux refus tandis que pour les strates à faible revenu, la non-réponse à la collecte se divise également entre les refus et les non-contacts.

Comme pour les enquêtes de 1997 à 1999, le taux de vacance est plus élevé pour les strates régulières que pour chacun des deux autres types de strate. Dans les strates à revenu élevé et à revenu faible, les taux de vacance sont semblables.

## 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. 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 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. 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.

#### *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 dix provinces. Les Territoires sont exclus de la population cible pour l'EDM 2000, l'enquête ne couvrant cette région qu'une année sur deux.

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.

Une bonne représentation de la population cible est indispensable à la production d'estimations 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.

En 1999, une nouvelle méthode de pondération a été introduite utilisant de nouveaux groupes de contrôle. Cette nouvelle méthode nous permet de mieux corriger la représentativité de la population visée en ciblant des groupes d'âge et de sexe mieux délimités que ceux utilisés précédemment et pour lesquels la couverture varie d'un groupe à l'autre.

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, 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<sup>6</sup>. Les taux de glissement par groupe d'âge-sexe au niveau national puis au niveau provincial 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. Un taux positif correspond à un surdénombrement du nombre de personnes ou de ménages dans l'enquête.

**Tableau 3.1**  
**Taux de glissement nationaux selon le groupe d'âge-sexe**

Taux de glissement national par groupe d'âge-sexe				
	Âge	Sexe		Total
		Hommes	Femmes	
Canada	0-6 ans	2,7	-2,3	0,3
	7-17 ans	-5,0	-0,9	-3,0
	18-24 ans	-11,6	-15,5	-13,5
	25-34 ans	-16,7	-12,7	-14,7
	35-54 ans	-15,5	-11,3	-13,4
	55-59 ans	-7,8	-10,0	-8,9
	60-64 ans	-10,2	-7,6	-8,8
	65-69 ans	-13,7	-11,9	-12,8
	70 ans et +	-7,8	-4,7	-6,0
	<b>Total</b>	<b>-10,9</b>	<b>-8,8</b>	<b>-9,9</b>

Pour l'EDM de 2000, le taux de sous-dénombrement de la population était 9,9 %. En analysant le tableau 3.1 au niveau des groupes d'âge, on constate qu'au niveau national, 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, chez les enfants, on observe un surdénombrement ou un léger sous-dénombrement, alors que chez les adultes il y a toujours sous-dénombrement. Le taux de sous-dénombrement pour l'ensemble des enfants est de 1,8 %, alors qu'il est de 12,2 % pour les adultes (données non présentées). Les taux les plus élevés au niveau national sont observés chez les hommes de 25 à 34 ans, chez ceux de 35 à 54 ans, ainsi que chez les femmes de 18 à 24 ans.

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

Pour l'EDM, on utilise la base de sondage de l'enquête sur la population active (EPA). Pour une même période, le taux de glissement de l'EPA au niveau national est de 9,7 % (référence [5]), ce qui est légèrement inférieur au taux de 12,2 % de l'EDM observé lorsqu'on se restreint aux personnes de 15 ans et plus (données non présentées).

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

Taux de glissement par groupe d'âge-sexe											
Sexe	Age	Terre-Neuve	Île-du-Prince-Édouard	Nouvelle-Écosse	Nouveau-Brunswick	Québec	Ontario	Manitoba	Saskatchewan	Alberta	Colombie-Britannique
Hommes	0-6	-8,0	-4,9	1,1	6,4	6,9	8,0	-8,5	14,3	-1,4	-16,5
	7-17	-16,2	-11,5	9,6	3,1	-7,0	-3,2	-12,7	8,9	-0,8	-15,4
	18-24	-22,2	-9,5	-28,8	-25,9	-3,7	-8,3	-2,7	-32,3	-15,0	-21,3
	25-34	-27,5	-32,0	-18,2	-16,0	-9,8	-21,3	-14,7	-9,2	-11,3	-19,5
	35-54	-12,7	-12,7	-15,4	-13,3	-14,3	-16,3	-10,9	-17,7	-16,1	-17,0
	55-59	3,8	5,3	-10,6	-25,5	2,2	-9,1	2,2	-21,5	-11,1	-19,9
	60-64	-26,7	-14,5	-8,9	-21,2	-6,2	-10,2	-15,8	-16,6	-18,3	-5,2
	65-69	-17,3	6,0	2,4	-16,9	-14,6	-10,3	-16,9	-3,3	-22,2	-21,4
	70 +	-3,9	4,5	-12,0	-0,1	-0,8	-14,7	6,0	-5,2	-6,2	-7,2
	<b>Total</b>	<b>-15,2</b>	<b>-11,2</b>	<b>-10,4</b>	<b>-11,1</b>	<b>-7,7</b>	<b>-11,2</b>	<b>-9,1</b>	<b>-9,2</b>	<b>-10,9</b>	<b>-16,6</b>
Femmes	0-6	-0,4	-15,0	12,3	17,8	-2,9	-10,1	0,8	-0,3	16,2	-0,8
	7-17	6,2	-11,2	-6,6	-0,3	6,4	-1,3	-14,7	-2,9	-4,1	-4,8
	18-24	-7,4	-19,4	-20,5	-25,8	-6,1	-17,7	-9,0	-11,6	-10,3	-31,4
	25-34	-16,8	-21,1	-21,6	-5,0	-4,7	-17,3	-14,1	-4,5	-0,3	-21,6
	35-54	-7,0	-11,4	-4,3	-9,4	-7,6	-12,0	-9,5	-9,0	-14,2	-17,4
	55-59	-21,1	2,9	-13,0	-23,2	-18,1	-1,9	-12,0	-1,1	-12,6	-12,0
	60-64	-16,9	9,9	-19,4	1,9	-2,7	-5,4	-14,0	-18,8	-9,2	-16,5
	65-69	1,9	16,0	-2,1	0,7	-15,1	-12,9	-1,1	-9,9	-15,0	-11,2
	70 +	-7,3	-15,6	-6,3	-5,9	-7,5	-7,6	2,5	1,6	10,2	-2,5
	<b>Total</b>	<b>-7,0</b>	<b>-11,4</b>	<b>-8,5</b>	<b>-6,5</b>	<b>-5,5</b>	<b>-10,4</b>	<b>-8,6</b>	<b>-5,8</b>	<b>-5,4</b>	<b>-14,3</b>
<b>Total</b>	<b>-11,1</b>	<b>-11,3</b>	<b>-9,4</b>	<b>-8,8</b>	<b>-6,6</b>	<b>-10,8</b>	<b>-8,9</b>	<b>-7,5</b>	<b>-8,1</b>	<b>-15,4</b>	

On observe un sous-dénombrement net pour chacune des provinces, les taux variant de 6,6 % à 15,4 %. 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 Saskatchewan (-7,5 %) cache le pire cas de sous-dénombrement au niveau provincial des groupes d'âge-sexe (32,3 % chez les hommes de 18 à 24 ans) et le quatrième pire cas de surdénombrement (14,3 % chez les garçons de 0 à 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 3.3**  
**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
<b>Canada</b>	<b>-8,1</b>	<b>-7,4</b>	<b>-4,7</b>	<b>-11,2</b>
Terre-Neuve	-8,3	4,5	-4,6	-14,8
Île-du-Prince-Édouard	-9,0	-19,6	10,2	-16,7
Nouvelle-Écosse	-8,5	-8,4	-3,8	-12,4
Nouveau-Brunswick	-6,6	-4,8	-0,5	-12,2
Québec	-5,4	-8,8	3,6	-10,3
Ontario	-9,2	-6,5	-9,4	-10,6
Manitoba	-5,2	4,5	-5,2	-11,9
Saskatchewan	-5,7	2,9	-11,0	-7,3
Alberta	-6,2	-6,2	-3,4	-8,4
Colombie-Britannique	-13,2	-14,0	-8,4	-16,7

On observe une sous-estimation du nombre de ménages de 8,1 % au niveau national. Ce taux de glissement est comparable au taux de glissement au niveau des personnes qui est de -9,9 %.

Au niveau provincial, il y a une variation importante du taux de glissement pour les ménages d'une personne. Le taux varie de -19,6 % pour l'Île-du-Prince-Édouard jusqu'à 4,5 % pour Terre-Neuve et pour le Manitoba. Le même phénomène est observable pour les ménages de deux personnes, où les taux de glissement varient de -11,0 % pour la Saskatchewan à 10,2 % pour l'Île-du-Prince-Édouard. Le taux de glissement pour les ménages de trois personnes est moins variable, les valeurs allant de -7,3 % pour la Saskatchewan à -16,7% pour l'Île-du-Prince-Édouard. Pour toutes les provinces à l'exception de l'Île-du-Prince-Édouard et de la Saskatchewan, le sous-dénombrement le plus élevé est observé pour les ménages de trois personnes et plus.

### **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-sexes définis dans ces tableaux, et ce pour chaque province. Pour plus de détails sur la méthodologie de l'ajustement, voir la référence [1]. 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.

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 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, 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 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 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 intervieweurs 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 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 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. 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 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 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 quant à elles 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.

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.<sup>7</sup>

## **5.1 Proportion de ménages ou personnes nécessitant imputation, au niveau national et provincial**

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) 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 aux niveaux national et provincial. 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 (excluant les dépenses vestimentaires) figure dans le tableau 5.2. 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 balancent pas (voir définitions dans la section 2.1). Le tableau est réparti en fonction du nombre de variables qui ont été imputées (parmi les 237) pour un ménage.

---

<sup>7</sup> Pour des raisons opérationnelles, ces indicateurs de qualité ne sont pas disponibles pour les données catégoriques comme les caractéristiques du logement et l'équipement ménager.

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

Province	Ménages (%) qui nécessitent l'imputation de <b>VARIABLES DE DÉPENSES</b> (sauf pour les dépenses en vêtements)			
	Nombre de variables imputées (parmi 237)			TOTAL
	1	2	3 ou plus	
<b>Canada</b>	<b>6,9</b>	<b>1,2</b>	<b>0,7</b>	<b>8,9</b>
Terre-Neuve	12,4	0,8	0,5	13,6
Î.-P.-É.	5,4	0,4	0,0	5,8
N.-É.	8,9	1,9	0,6	11,4
N.-B.	7,9	1,4	0,9	10,3
Québec	2,4	0,5	0,4	3,3
Ontario	7,1	1,9	1,2	10,2
Manitoba	4,7	0,7	0,9	6,3
Saskatchewan	6,1	0,8	0,3	7,2
Alberta	4,5	0,2	0,4	5,1
C.-B.	10,3	2,6	1,6	14,5

Le tableau 5.1 montre qu'il a fallu imputer des dépenses (mise à part la section sur les vêtements) pour 8,9 % des ménages au niveau national mais pour près de 80 % d'entre eux, seulement une variable a été imputée. Il y a très peu de ménages au niveau national qui ont eu plus d'une variable imputée (1,9 %). 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 (3,3 %), l'Alberta (5,1 %) et l'Île-du-Prince-Édouard (5,8 %). On retrouve les taux les plus élevés à Terre-Neuve (13,6 %) et en Colombie-Britannique (14,5 %). Le faible pourcentage de personnes pour lesquelles 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.

#### 5.1.2 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, 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 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 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. 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	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
	1 variable imputée	2 variables imputées ou plus (pas toutes)	Toutes les variables imputées (revenu total connu)	<b>TOTAL</b> (toute forme d'imputation)	
<b>Canada</b>	<b>0,5</b>	<b>0,2</b>	<b>2,0</b>	<b>2,7</b>	<b>16,6</b>
Terre-Neuve	0,3	0,1	1,2	1,6	13,4
Î.-P.-É.	0,0	0,2	2,9	3,2	20,1
N.-É.	0,3	0,2	2,0	2,5	11,1
N.-B.	0,0	0,2	4,4	4,6	19,3
Québec	0,5	0,0	1,0	1,6	20,1
Ontario	1,1	0,3	1,8	3,2	11,9
Manitoba	1,5	0,2	1,9	3,6	14,7
Saskatchewan	0,2	0,2	0,7	1,1	20,9
Alberta	0,1	0,2	2,4	2,7	11,5
C.-B.	0,7	0,2	1,9	2,8	24,9

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 70 % 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. Au niveau provincial, le pourcentage de personnes nécessitant l'imputation d'au moins une variable de revenu est également peu élevé, variant entre un minimum de 1,1 % pour la Saskatchewan et un maximum de 4,6 % pour le Nouveau-Brunswick.

Dans la dernière colonne du tableau, on constate qu'environ 17 % des personnes ont nécessité une imputation pour au moins une des variables des dépenses vestimentaires. Les taux provinciaux s'échelonnent d'environ 11 % (Nouvelle-Écosse, Ontario et Alberta) à près de 25 % pour la Colombie-Britannique. Presque toutes ces personnes ont fourni le total de leurs dépenses vestimentaires, mais il a fallu imputer les composantes. 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.

## BIBLIOGRAPHIE

- [1] 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.
- [2] Wolter, K.M. (1985). Introduction to Variance Estimation, Springer-Verlag New-York Inc.
- [3] Guide de l'utilisateur – Enquête sur les dépenses des ménages, 2000, No. 62F0026MIF2001004 au catalogue.
- [4] Les habitudes de dépenses au Canada (2000), No. 62-202-XPF au catalogue.
- [5] Enquête sur la population active, Rapport mensuel sur les opérations, Enquête 200012, Division des méthodes d'enquêtes auprès des ménages.

# 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

$\pi_k^{-1}$  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  $\text{taux}_c$ , 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  $c$ .