



N° 62F0026MIF au catalogue — N° 004

ISSN : 1708-8887

ISBN : 0-662-74522-1

Document de recherche

Série de documents de recherche sur les dépenses des ménages

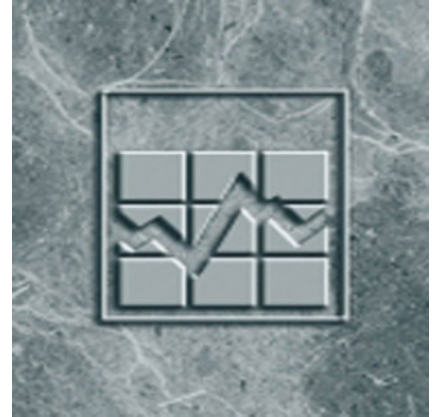
La méthodologie de l'Enquête sur les dépenses alimentaires

2001

par Sylvie Auger, Khushnood Khan et Johanne Tremblay

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

Téléphone: 613 951-7355



Statistique
Canada

Statistics
Canada

Canada

Comment obtenir d'autres renseignements

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

Pour obtenir des renseignements sur l'ensemble des données de Statistique Canada qui sont disponibles, veuillez composer l'un des numéros sans frais suivants. Vous pouvez également communiquer avec nous par courriel ou visiter notre site Web.

Service national de renseignements	1 800 263-1136
Service national d'appareils de télécommunications pour les malentendants	1 800 363-7629
Renseignements concernant le Programme des services de dépôt	1 800 700-1033
Télécopieur pour le Programme des services de dépôt	1 800 889-9734
Renseignements par courriel	infostats@statcan.ca
Site Web	www.statcan.ca

Renseignements pour accéder au produit

Le produit n° 62F0026MIF au catalogue est disponible gratuitement. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.ca et de choisir la rubrique Nos produits et services.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois, et ce, 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 Statistique Canada au numéro sans frais 1 800 263-1136. Les normes de service sont aussi publiées dans le site www.statcan.ca sous À propos de Statistique Canada > Offrir des services aux Canadiens.



Statistique Canada
Division de la statistique du revenu

Série de documents de recherche sur les dépenses des ménages

La méthodologie de l'Enquête sur les dépenses alimentaires

2001

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2005

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

Juillet 2005

N° 62F0026MIF2005004 au catalogue

ISSN : 1708-8887
ISBN : 0-662-74522-1

Périodicité : hors série

Ottawa

This publication is available in English upon request (Catalogue no. 62F0026MIE).

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

L'Enquête sur les dépenses alimentaires (EDA) est une enquête périodique recueillant des informations auprès des ménages sur les habitudes de dépenses alimentaires. La collecte des renseignements se fait principalement à partir de journaux hebdomadaires des achats que les répondants doivent remplir quotidiennement durant deux semaines consécutives.

Ce document présente une description détaillée de la méthodologie de cette enquête. On y décrit d'abord brièvement le plan d'échantillonnage qui est principalement basé sur le plan de l'Enquête sur la population active. Par la suite, on y présente les méthodes de collecte, de traitement des données, de pondération, d'estimation de la variance ainsi que de suppression des données non fiables dans les tableaux d'estimations.

Table des matières

1.	Introduction	6
2.	Population cible	6
3.	Plan d'échantillonnage.....	7
3.1	Taille et répartition de l'échantillon de l'EDA	8
3.2	Plan d'échantillonnage de l'EPA (sélection des grappes)	9
3.3	Sélection de l'échantillon de l'EDA	11
4.	Collecte des données	12
4.1	Méthode de collecte des données.....	12
4.2	Entrevue et procédures de suivi	13
4.3	Supervision et contrôles	13
4.4	Non-réponse à l'EDA.....	14
5.	Traitement des données.....	15
5.1	Codage, saisie des données et vérification	15
5.2	Imputation	16
6.	Pondération et estimation.....	17
6.1	Poids d'échantillonnage.....	18
6.2	Ajustement pour la non-réponse totale.....	18
6.3	Ajustement à l'information auxiliaire	19
6.4	Autres ajustements	21
6.5	Estimations	21
7.	Estimation de l'erreur d'échantillonnage.....	22
7.1	Modèle pour dériver une approximation du CV pour les estimations des domaines	22
7.2	Modèle pour dériver une approximation du CV à partir des fichiers de microdonnées.....	22
8.	Suppression de données et confidentialité.....	23
8.1	Suppression des données non fiables dans les tableaux d'estimations.....	23
8.2	Confidentialité des fichiers de microdonnées	23
9.	Changements dans la méthodologie de l'enquête	23
	Bibliographie	25
	Annexe A.....	26
	Formule pour le calcul de la variance des estimations à partir du jackknife	
	Annexe B.....	27
	Formule d'approximation du CV pour un domaine (un sous-groupe de la population)	

1. Introduction

L'enquête sur les dépenses alimentaires (EDA) est une enquête périodique qui a lieu à tous les 4 à 6 ans. Cette enquête recueille des informations auprès des ménages canadiens sur les habitudes de dépenses alimentaires.

Cette enquête sert à compléter l'Enquête sur les dépenses des ménages (EDM)¹ puisque l'EDA procure des renseignements détaillés sur les dépenses alimentaires qu'il n'est pas possible d'obtenir avec la méthodologie utilisée pour l'EDM. L'Enquête sur les dépenses alimentaires et l'Enquête sur les dépenses des ménages servent à mettre à jour les pondérations utilisées dans l'Indice des prix à la consommation.

Les données de l'EDA sont recueillies à partir d'entrevues personnelles au moyen d'un questionnaire papier (FE2), ainsi que par l'entremise de deux journaux hebdomadaires des achats (FE3) que le répondant doit remplir quotidiennement durant deux semaines consécutives. Ces données sont recueillies auprès d'un échantillon de ménages couvrant les dix provinces ainsi que les villes de Whitehorse et Yellowknife. En 2001, la ville d'Iqaluit, a été ajoutée. L'enquête se déroule tous les mois au cours d'une même année civile. Les données recueillies sont par la suite vérifiées et pondérées. Les produits de l'enquête comprennent la publication, les tableaux et les fichiers de microdonnées nécessaires aux divers utilisateurs.

L'objectif de ce document est de fournir une description détaillée de la méthodologie de l'enquête quant au plan d'échantillonnage, à la collecte et au traitement des données, à la production des estimations et des autres produits ainsi qu'aux règles régissant leur diffusion.

2. Population cible

La population cible de l'EDA comprend les individus vivant dans les ménages privés canadiens à l'exclusion des représentants officiels de pays étrangers qui vivent au Canada et leurs familles ainsi que les résidents des réserves indiennes et des terres publiques. La restriction aux ménages privés implique que les pensionnaires d'établissements institutionnels tels 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 exclus.

1. L'EDM est une enquête annuelle où l'on demande à un échantillon de ménages canadiens de faire état de toutes les dépenses effectuées au cours d'une année civile. Les résultats de l'EDM sont publiés dans *Les habitudes de dépenses au Canada*, n° 62-202 au catalogue.

Nous n'avons pas recueilli de renseignements auprès des personnes qui vivent temporairement loin de leur famille (par exemple, les étudiants d'université) parce que nous obtenons l'information auprès de leur famille, si elle est choisie dans l'échantillon. Donc, si un ménage est constitué uniquement de personnes vivant temporairement loin de leur famille, ce ménage est considéré hors du champ de l'enquête.

De plus, en 2001, puisque la taille de l'échantillon est beaucoup plus petite que lors des enquêtes précédentes, les membres des ménages habitant des logements situés dans les régions éloignées sont exclus de l'enquête. Ces régions, en général peu peuplées, sont situées principalement dans la portion nordique de certaines provinces. Ces régions sont difficilement accessibles et par conséquent, la collecte de données auprès de ces ménages est très dispendieuse.

L'enquête couvre près de 97% de la population dans les dix provinces. Au Yukon, Territoires du Nord-Ouest et Nunavut, l'enquête ne couvre que les villes de Whitehorse, Yellowknife et Iqaluit ce qui représente 62%, 44% et 17% de leur population respective selon le Recensement de 1996.

3. Plan d'échantillonnage

Les données sur les dépenses alimentaires sont recueillies auprès de ménages provenant d'un échantillon de logements sélectionnés selon un plan d'échantillonnage stratifié à plusieurs degrés. Ce plan varie selon le niveau d'urbanisation, mais il s'agit généralement d'un échantillon à deux degrés, dont le premier correspond à un échantillon aréolaire, c'est-à-dire un échantillon de régions géographiques appelées grappes.

La sélection des grappes est effectuée pour l'ensemble de l'année. Cet ensemble de grappes sélectionnées est divisé également en 12 sous-ensembles, afin de permettre la collecte des données au cours des 12 mois de l'année, soit un sous-ensemble par mois. La collecte est effectuée durant toute l'année, afin de tenir compte du caractère saisonnier des dépenses alimentaires. La méthode utilisée pour obtenir les 12 sous-ensembles doit faire en sorte que les trois échantillons mensuels pour chaque trimestre couvrent l'ensemble du Canada.

Au deuxième degré, avant chaque mois de collecte, les logements sont sélectionnés à partir d'une liste de tous les logements privés des grappes du sous-ensemble retenu pour le mois concerné. Tous les logements sélectionnés où habitent des personnes de la population cible constituent l'échantillon de l'enquête.

Afin de réduire les coûts de collecte, l'EDA utilise dans une large mesure le même plan d'échantillonnage que l'Enquête sur la population active (EPA). Les logements de l'échantillon de l'EDA sont sélectionnés à partir des grappes de l'échantillon de l'EPA, mais les deux enquêtes utilisent des échantillons de logements différents. Les principaux aspects du plan d'échantillonnage en

grappes de l'EPA sont décrits à la section 3.2. Une description plus détaillée figure dans la publication faisant état de la méthodologie de l'EPA [1]. L'approche utilisée pour la sélection des logements de l'EDA, à partir des grappes de l'échantillon de l'EPA, est abordée à la section 3.3. La section suivante explique comment l'échantillon est réparti entre les cinq régions canadiennes et les trois villes des territoires.

3.1 Taille et répartition de l'échantillon de l'EDA

En raison de restrictions budgétaires, la taille de l'échantillon de l'enquête de 2001 a été réduite de façon considérable comparativement à l'enquête précédente de 1996. Lors de l'enquête de 1996, l'échantillon comptait 17 000 logements, soit environ 15 000 ménages, une fois exclus les logements vacants et les ménages hors du champ de l'enquête (par ex., les ménages constitués seulement d'individus ayant une résidence permanente ailleurs). Avec un tel échantillon, on s'attendait à ce que les estimations aux niveaux national et provincial, ainsi que celles pour les trois villes des territoires², soient fiables.

À l'étape de la planification de l'EDA de 2001, il a été décidé de réduire la taille de l'échantillon, pour la faire passer de 17 000 à 14 000 logements. En dépit de cette réduction, on s'attendait toujours à des estimations relativement fiables au niveau national, provincial et pour les trois villes des territoires. Quelques semaines après le début de la collecte des données, en janvier 2001, on a dû réduire la taille de l'échantillon pour la faire passer de 14 000 à 9 000 logements en raison de contraintes budgétaires. Étant donné que la collecte était déjà en cours, il n'a pas été possible de réduire l'échantillon de janvier. L'échantillon a été réduit de façon appropriée à partir du mois de février. On s'attendait dès lors à pouvoir produire des estimations fiables au niveau régional³ plutôt qu'au niveau provincial.

L'échantillon total, à l'exclusion de l'échantillon des trois villes des territoires, a été réparti entre les cinq grandes régions canadiennes, afin d'obtenir un compromis entre la fiabilité au niveau national et celle au niveau régional pour le *Total des dépenses alimentaires hebdomadaires*. De façon plus particulière, la répartition de l'échantillon est établie en fonction de la variabilité des dépenses au titre des aliments dans chaque région et, dans une moindre mesure, du nombre de ménages au niveau régional.

Dans chaque région, l'échantillon a été réparti entre les provinces selon la proportion du nombre de ménages. Chaque échantillon provincial a été réparti en proportion directe avec la taille de la population pour les régions métropolitaines de recensement. La portion de l'échantillon allouée à l'extérieur des régions métropolitaines de recensement a suivi la répartition de l'EPA [1].

Le tableau 3.1 présente la taille de l'échantillon de chacune des provinces et des trois villes des territoires pour les enquêtes de 1996 et 2001. La taille de l'échantillon est présentée selon le nombre de logements et le nombre de

2. Dans les territoires, seules les villes Yellowknife, Whitehorse et Iqaluit sont couvertes par l'enquête.

3. Les cinq régions sont l'Atlantique, le Québec, l'Ontario, les Prairies et la Colombie-Britannique.

ménages (une fois exclus les logements vacants et les ménages hors du champ d'enquête).

Tableau 3.1
Taille de l'échantillon, selon la province et les trois villes des territoires

Région géographique	Taille de l'échantillon			
	1996		2001	
	Nombre de logements	Nombre de ménages	Nombre de logements	Nombre de ménages
Canada	17 000	14 993	9 488	8 414
Terre-Neuve-et-Labrador	691	592	233	199
Île-du-Prince-Édouard	621	516	79	67
Nouvelle-Écosse	1 347	1 143	396	329
Nouveau-Brunswick	1 124	941	304	261
Région de l'Atlantique	3 783	3 192	1 012	856
Région du Québec	2 395	2 139	1 816	1 615
Région de l'Ontario	3 886	3 479	2 632	2 404
Manitoba	1 023	895	451	387
Saskatchewan	1 369	1 211	393	335
Alberta	1 461	1 319	948	868
Région des Prairies	3 853	3 425	1 792	1 590
Région de la Colombie-Britannique	2 212	2 038	1 558	1 422
Total des régions	16 129	14 273	8 810	7 887
Whitehorse	435	363	250	202
Yellowknife	436	357	252	192
Iqaluit	0	0	176	133
Total des 3 villes des territoires	871	720	678	527

3.2 Plan d'échantillonnage de l'EPA (sélection des grappes)

Le plan d'échantillonnage de l'EPA est fondé sur les données du recensement et est remanié après chaque recensement décennal, afin de tenir compte des changements démographiques. Le plan d'échantillonnage actuel est fondé sur les données du Recensement de 1991.

Les principes qui sous-tendent le plan d'échantillonnage de l'EPA sont les mêmes pour chaque province. Tout d'abord, chaque province est divisée en un certain nombre de régions géographiques, constituées de régions économiques (RE), de régions économiques de l'assurance-emploi (RÉAE) et de leurs intersections. De façon plus particulière, chaque région métropolitaine de recensement constitue une région géographique, étant donné qu'il s'agit d'une RÉAE.

Chaque région géographique est divisée par la suite en types de secteurs, principalement des secteurs urbains, des secteurs ruraux et des régions éloignées. Le plan d'échantillonnage varie selon le type de secteur. Les régions éloignées représentent principalement la partie nord de certaines provinces, et

sont, pour la plupart, peu peuplées. Ces régions sont exclues de l'EDA. Le plan d'échantillonnage des secteurs urbains et ruraux est décrit ci-dessous.

Secteurs urbains

Dans certaines villes importantes comportant un grand nombre d'immeubles à appartements, on a eu recours à une liste des appartements et à une base aréolaire. Dans d'autres secteurs urbains, on a eu recours uniquement à une base aréolaire.

Une base aréolaire est formée d'une liste de régions géographiques. Ces régions sont combinées pour former des strates. On compte jusqu'à trois niveaux de stratification. Au niveau supérieur, l'objectif consiste généralement à constituer des strates compactes et contiguës au niveau géographique, tandis qu'au niveau inférieur, on vise la constitution de strates finales les plus homogènes possibles en ce qui a trait à certaines caractéristiques socio-économiques. Dans quelques grandes villes⁴, des strates distinctes sont constituées à partir des secteurs de dénombrement où le revenu moyen des ménages est élevé (environ 100 000 \$ ou plus).

Afin de réduire les coûts de la collecte, les logements qui constituent la strate finale ne sont pas sélectionnés directement. La strate est d'abord divisée en grappes. Dans les secteurs urbains, les grappes peuvent être des combinaisons de côtés d'îlot, de secteurs de dénombrement (SD) ou de parties de SD. Puis, les grappes sont sélectionnées (habituellement six, parfois un multiple de six) dans chaque strate, avec une probabilité proportionnelle à la taille de la grappe. Par exemple, si la taille d'une grappe est deux fois plus grande que celle d'une autre grappe, elle sera deux fois plus susceptible d'être sélectionnée que la deuxième.

La base de sondage d'appartements est une liste d'appartements dressée à partir de données de la Société canadienne d'hypothèques et de logement. Cette base de sondage fournit une meilleure représentation des personnes qui habitent en appartement et réduit les effets de l'augmentation des grappes découlant de la construction de nouveaux immeubles à appartements. Dans certaines villes⁵, les strates des appartements sont divisées en deux catégories : à faible revenu (revenu moyen des ménages inférieur à 20 000 \$) et régulière. Pour chaque strate de la base de sondage, les immeubles à appartements sont sélectionnés dans l'échantillon au premier degré, avec une probabilité proportionnelle au nombre d'appartements compris dans l'immeuble.

Dans les secteurs urbains à faible densité, à savoir principalement les villes très dispersées, on utilise un plan d'échantillonnage différent. L'échantillonnage se fait en trois étapes : tout d'abord, on sélectionne les villes à l'intérieur de la strate ; puis, les grappes (côtés d'îlot) sont choisies dans les villes ; enfin, les logements sont sélectionnés dans les grappes.

4. Montréal, Ottawa, Toronto, Hamilton, London, Winnipeg, Calgary et Vancouver.

5. Montréal, Ottawa-Hull, Toronto, Winnipeg, Calgary, Edmonton et Vancouver.

Secteurs ruraux

Dans les secteurs ruraux, on a uniquement recours à une base aréolaire. Des strates géographiques sont constituées en combinant deux ou trois divisions de recensement, qui sont par la suite subdivisées, lorsque leur nombre le permet, afin de constituer des strates qui sont homogènes du point de vue de certaines caractéristiques socio-économiques. Au premier degré d'échantillonnage, les secteurs de dénombrement sont sélectionnés à l'intérieur de chaque strate finale, selon une probabilité proportionnelle au nombre des logements compris dans le SD.

Dans les secteurs ruraux à faible densité, on a recours à une variante du plan d'échantillonnage. Deux ou trois unités primaires d'échantillonnage constituées d'un groupe de six SD sont sélectionnées au premier degré, puis un échantillon des logements est choisi à l'intérieur de chacun des SD des unités primaires d'échantillonnage sélectionnées.

3.3 Sélection de l'échantillon de l'EDA

Les intervieweurs visitent les grappes sélectionnées dans le plan d'échantillonnage de l'EPA et dressent une liste de tous les logements privés qu'elles comprennent. À partir de cette liste, on choisit un échantillon pour l'EPA et un autre pour l'EDA. Les logements sont sélectionnés grâce à un échantillonnage systématique.

Étant donné que l'EDA utilise un échantillon beaucoup plus restreint que l'EPA, les logements ne sont pas sélectionnés dans chaque grappe de l'EPA. L'EPA est une enquête par panel, dans laquelle les ménages demeurent dans l'échantillon pendant six mois. L'échantillon de l'EPA a été conçu afin de pouvoir être divisé en six sous-échantillons représentatifs, en vue de permettre le renouvellement d'un sixième de l'échantillon chaque mois. C'est pourquoi six grappes (ou un multiple de six) sont sélectionnées dans chaque strate finale, une par groupe de renouvellement. Cette méthode facilite la sélection d'un échantillon plus petit pour une autre enquête, étant donné que l'on peut utiliser un sous-ensemble des groupes de renouvellement. Il s'agit généralement de l'approche utilisée pour les enquêtes supplémentaires de l'EPA. Dans le cas de l'EDA, le nombre de groupes de renouvellement est déterminé au niveau de la strate, selon les besoins propres à l'enquête en ce qui a trait à la répartition provinciale et infraprovinciale de l'échantillon. Dans certains cas, seulement une partie d'un groupe de renouvellement est nécessaire. Lorsque tel est le cas, les logements sont supprimés de l'échantillon, grâce à une suppression systématique à partir d'une liste classée géographiquement.

L'échantillon des logements est obtenu après l'opération de listage des grappes. Étant donné que les taux d'échantillonnage sont déterminés au préalable, il peut exister une différence entre les tailles prévues et réelles des échantillons, dans les cas où le nombre de logements de la liste diffère du nombre utilisé pour l'élaboration du plan d'échantillonnage de l'enquête. Pour réduire les coûts de la collecte (étant donné que les tailles des grappes ont tendance à augmenter) et pour prévenir les disparités significatives dans les charges de travail des

intervieweurs, on a recours à deux méthodes pour contrôler la taille des échantillons.

La première méthode, qui est généralement utilisée, consiste à corriger le problème en supprimant systématiquement une partie des logements sélectionnés à l'origine. Le processus servant à maintenir la taille de l'échantillon au niveau souhaité est appelé stabilisation de l'échantillon. La deuxième méthode est utilisée lorsque le nombre de logements augmente de façon marquée dans certains secteurs urbains. Dans ce cas, on a plutôt recours au sous-échantillonnage des grappes. Il existe trois options, selon l'importance de l'augmentation et le niveau de similitude entre les nouveaux logements et ceux figurant dans la même strate : constituer des sous-grappes, créer une nouvelle strate, ou sous-échantillonner les logements de la grappe.

4. Collecte des données

L'EDA vise à recueillir des renseignements sur les dépenses alimentaires des ménages canadiens sur une base volontaire.

4.1 Méthode de collecte des données

La collecte des données est effectuée à partir d'entrevues personnelles au moyen d'un questionnaire papier (FE2), ainsi qu'avec deux journaux hebdomadaires des achats (FE3) que le répondant doit remplir quotidiennement durant deux semaines consécutives. L'enquête se déroule tous les mois au cours d'une même année civile et les données sont recueillies à partir des ménages provenant d'un échantillon différent de logements à chaque mois.

Le questionnaire sert essentiellement à recueillir des données sur certaines caractéristiques socio-économiques (par exemple, le nombre de personnes, le revenu du ménage), ainsi que des renseignements sur les habitudes d'achat des ménages et sur leurs dépenses alimentaires au cours du mois précédent.

Les journaux servent à recueillir les dépenses alimentaires quotidiennes aux magasins et aux restaurants. Dans le cas des dépenses effectuées aux magasins, les détails de ces dépenses comprennent la description de l'article acheté, le type de conditionnement (frais, congelé, en conserve, séché, autre), le type de magasin (supermarché, magasin spécialisé, dépanneur, autre), le nombre d'unités achetées, le poids net ou le volume par unité (en mesures métriques ou impériales) et le coût total de l'achat. Dans le cas des dépenses effectuées aux restaurants, les détails de ces dépenses comprennent le type de repas (déjeuner, dîner, souper, collation), le nombre de repas achetés, le type de restaurant (service aux tables, restauration rapide, cafétéria, autre) et le coût total de l'achat excluant les coûts, taxes et pourboires concernant les boissons alcoolisées.

En 2001, les répondants ont été fortement encouragés à joindre leurs reçus d'achat à la page du journal correspondant au jour de l'achat. Il y a eu aussi un autre élément nouveau en 2001, soit le carnet. Tous les membres du ménage

âgés de 10 ans et plus disposaient d'un petit carnet à apporter constamment avec eux pour y inscrire les produits au moment de l'achat.

Les membres d'un ménage

Pour obtenir les dépenses relatives à un ménage, il est essentiel de bien identifier les membres qui en font partie. La personne ou le groupe de personnes qui occupe un logement constitue un ménage. Pour l'EDA, les membres de ce ménage sont les personnes qui vivent dans le logement au moment de l'entrevue qui n'ont pas une résidence permanente ailleurs et ne sont pas membres d'un autre ménage.

4.2 Entrevue et procédures de suivi

Les entrevues sont menées par des intervieweurs de Statistique Canada dont certains sont également en charge de recueillir l'information de l'EPA. Ces intervieweurs reçoivent une formation particulière à l'EDA.

Une à deux semaines avant le mois de l'entrevue, le bureau régional envoie par la poste une lettre d'introduction aux occupants des logements sélectionnés soulignant l'importance de l'enquête et la confidentialité des renseignements recueillis ainsi qu'un dépliant donnant plus d'information concernant l'enquête. L'intervieweur se rend ensuite au logement. Lorsqu'il n'y a personne à la maison, d'autres tentatives sont effectuées auprès du ménage, par exemple en le visitant à différentes heures de la journée ou en tentant de le rejoindre par téléphone. Le numéro de téléphone, s'il est disponible, fait partie de l'information au sujet du logement qui est fournie à l'intervieweur. Si le numéro de téléphone est manquant, l'intervieweur essaiera de le retracer.

Lors du rendez-vous initial, l'intervieweur remplit le questionnaire (FE2) en transcrivant lui-même les réponses obtenues auprès du répondant. Par la suite, il remet les journaux (FE3) au répondant en donnant des indications sur la façon de bien remplir le journal et fixe un rendez-vous afin de revenir huit jours plus tard pour recueillir le journal complété. Lors de cette deuxième visite, l'intervieweur récupère le premier journal, rappelle au répondant de remplir le deuxième journal et fixe un dernier rendez-vous afin de récupérer celui-ci.

Si une personne refuse de participer à l'EDA, l'intervieweur complètera un rapport de non-interview et en informera l'intervieweur principal le plus tôt possible. Selon les commentaires fournis, l'intervieweur principal décidera s'il entame ou non les tentatives de conversion de refus. À noter qu'envoyer une lettre destinée aux refus mentionnant l'importance d'obtenir des données peut être utile mais a ses limites en raison de la contrainte de temps ; les données doivent être recueillies sur une période de deux semaines au cours du mois de l'entrevue.

4.3 Supervision et contrôles

Tous les intervieweurs de l'EDA travaillent sous la supervision d'un intervieweur principal qui est chargé de s'assurer à ce que les intervieweurs connaissent bien

les concepts et les méthodes de l'enquête ainsi que de contrôler périodiquement le travail et de revoir les documents remplis. Les intervieweurs principaux travaillent à leur tour sous la supervision des gestionnaires de programme qui sont postés à chacun des bureaux régionaux de Statistique Canada.

À la fin de chaque période de déclaration hebdomadaire, l'intervieweur retourne chez le répondant pour prendre et réviser le journal de la semaine précédente afin de s'assurer qu'il est complet et exact. Si le journal n'a pas été rempli, l'intervieweur tente de compléter le journal en demandant au répondant de se rappeler ses dépenses. Il en profite pour recueillir des renseignements sur les facteurs qui auraient pu influencer sur les habitudes et les dépenses du répondant. Parmi ces renseignements, on retrouve l'absence de membres du ménage pendant la semaine, le nombre de repas servis à des invités ou à d'autres personnes ne faisant pas partie du ménage, le nombre de repas reçus gratuitement, les dons d'aliments ou les aliments provenant de l'exploitation agricole ou du jardin du répondant, ou les produits de la chasse ou de la pêche.

4.4 Non-réponse à l'EDA

Malgré toutes les tentatives effectuées par les intervieweurs pour obtenir l'information, il reste toujours un certain nombre de ménages non répondants. Par exemple, il a été impossible de les contacter ; une entrevue n'a pas été menée dû à des circonstances incontrôlables ; ou parce que les membres du ménage ont refusé de participer à l'enquête. Les taux de non-réponse à la collecte observés au cours des deux dernières enquêtes sont présentés au tableau 4.1.

Tableau 4.1
Taux de non-réponse à l'EDA

Année de référence	Taux de non-réponse			
	À la collecte			Total final
	Total	Pas de contact	Refus	
1996	21,8	4,6	17,2	23,7
2001	28,5	11,4	17,0	28,5

En 2001, le taux de non-réponse final est égal à celui de la collecte. En 1996, le taux de non-réponse final est 23,7 % en raison de 280 questionnaires répondus mais qui étaient inutilisables. Cette différence s'explique par un changement de traitement des données entre 1996 et 2001.

Il y a également les répondants qui ne rapportent qu'une semaine de données. En 2001, au niveau des dix provinces, il y en a eu un peu moins de 5 %. En 1996, ils étaient un peu moins de 2 %.

5. Traitement des données

Les principales étapes du traitement des données de l'EDA sont le codage des réponses, la saisie des données, la vérification, l'imputation pour la non-réponse partielle et la pondération. Cette dernière est abordée à la section 6.

5.1 Codage, saisie des données et vérification

Le codage constitue une étape importante de l'EDA. On procède à un codage de base du questionnaire, et à plusieurs opérations de codage des journaux : les descriptions de tous les articles achetés dont il est fait mention dans les journaux doivent être codées. À noter qu'en 2001, le codage a été effectué par le personnel spécialisé de cette enquête, ce qui n'était pas le cas en 1996.

Le codage est effectué au bureau central de Statistique Canada (SC). Dans le cas des journaux, les codeurs utilisent un logiciel élaboré à la Division de la statistique du revenu de SC, afin de les aider à choisir les codes d'aliments appropriés et inscrivent les codes en rouge directement dans les journaux. Lorsque la description de l'article acheté n'est pas fournie ou n'est pas suffisamment détaillée, on attribue le code d'aliment générique le plus approprié, qui par la suite sera imputé au moyen du processus d'imputation automatisée dont il est question dans la section suivante.

Dans le cas des aliments achetés au magasin, chaque entrée d'un journal est composée des champs suivants : la description de l'article acheté, le type de conditionnement, le type de magasin, le nombre d'unités achetées, le poids net ou le volume par unité et le coût total de l'achat. Tous les champs qui sont manquants, à l'exception de la description de l'article, sont imputés par le codeur, à partir des autres renseignements figurant dans le journal. Afin de contrôler la qualité, certains questionnaires et journaux des répondants sont sélectionnés de façon aléatoire et vérifiés par les codeurs principaux.

Une fois le codage effectué, on procède à la saisie des questionnaires et des journaux. La principale vérification automatisée qui est effectuée au moment de la saisie consiste à vérifier la validité des codes des aliments. Si un code d'aliment n'est pas valide, le responsable de la saisie des données vérifie s'il s'agit d'une erreur de saisie. Si ce n'est pas le cas, le responsable de la saisie entre un code spécifique qui fait en sorte que, lorsque la série de vérifications automatisées suivant la saisie des données aura été effectuée, le code d'aliment comportera une étiquette indiquant qu'il nécessite d'être corrigé.

Du point de vue de la saisie des données, la charge de travail d'un commis à la saisie est constituée de 20 questionnaires et journaux répartis en quatre lots de cinq. Afin de vérifier la qualité des données saisies, on sélectionne un lot de façon aléatoire. Si la qualité des données du lot est acceptable, l'ensemble des questionnaires est approuvé ; autrement, les trois autres lots sont vérifiés.

Enfin, les données saisies sont soumises à une série de vérifications automatisées. Lorsque des erreurs sont décelées, certaines font l'objet d'une imputation automatique tandis que les autres, après examen final des

questionnaires ou des journaux, sont imputées ou étiquetées manuellement pour indiquer qu'elles doivent être soumises au processus d'imputation automatisée.

5.2 Imputation

Comme il est mentionné dans la section précédente, au cours du codage et de la saisie des données, les questionnaires et les journaux sont soumis à certaines vérifications manuelles et automatisées, et font l'objet de certaines imputations manuelles ou automatisées, y compris l'étiquetage des champs qui doivent être soumis au processus d'imputation automatisée.

Les champs qui sont étiquetés en vue de l'imputation sont principalement manquants. Parmi les exemples de champs manquants dans le questionnaire (FE2) figurent les dépenses dans les restaurants sans indication du nombre de repas achetés. Parmi les exemples de champs manquants dans le journal (FE3) figure la déclaration d'un achat de bœuf, sans indication du type de bœuf (haché, en cubes, etc.). On a recours à plus d'une méthode d'imputation automatisée, mais la méthode principale utilisée en 2001 est celle du plus proche voisin.

Cette technique repose sur la constitution de groupes de ménages similaires, selon certaines classifications (par ex., province de résidence, trimestre de la collecte des données). À l'intérieur de ces groupes, chaque ménage qui doit être imputé (receveur) est apparié à un ménage qui a rempli un questionnaire (donneur) et qui s'apparente le plus à l'autre du point de vue de certaines caractéristiques (p. ex., revenu, taille du ménage, etc.). Les données du donneur sont imputées au receveur si ces données permettent de satisfaire les règles de vérification de cohérence avec les données déclarées par le receveur. Par exemple, si le receveur a indiqué avoir acheté des aliments préparés, les dépenses imputées doivent être supérieures à zéro.

Si on ne peut trouver de donneurs pour certains receveurs lors de cette première étape, on a recours encore une fois à la méthode d'imputation par donneur, en supprimant l'exigence liée à la similitude entre les revenus des ménages, qui est considérée comme moins importante que les autres caractéristiques. Plusieurs étapes d'imputation par donneur sont nécessaires afin de pouvoir imputer tous les receveurs. Dans quelques cas, la condition la plus importante, à savoir le trimestre, est aussi abandonnée pour trouver un donneur. Pour plus de détails concernant les différentes étapes d'imputation par donneur, voir la référence [2].

Dans le cas du questionnaire (FE2), deux sections sont imputées : *Habitudes de dépense du ménage* (section B) et *Aliments et boissons consommés hors du lieu de résidence – pendant une absence d'une nuit ou plus au cours du mois précédent* (section C). Ces sections sont imputées séparément, étant donné qu'elles sont considérées comme relativement indépendantes l'une de l'autre. Pour la section B, on utilise l'approche du plus proche voisin, mais pour la section C, on utilise d'autres méthodes dans certains cas, en raison de la difficulté de trouver un donneur approprié. Par exemple, certains répondants ont déclaré des dépenses au restaurant, et n'ont pas déclaré le nombre de repas achetés, ou vice versa. Ces cas sont imputés grâce à la méthode d'imputation par la moyenne. La moyenne utilisée pour imputer est celle au niveau national.

Dans le cas du journal (FE3), pour chaque jour, on impute uniquement la section des *Aliments et boissons achetés dans un magasin*. Afin d'imputer cette section, on procède à des imputations indépendantes selon le regroupement des données manquantes qui doivent être imputées. De façon générale, l'imputation est effectuée au niveau de l'entrée, mais dans certains cas, elle se fait au niveau de la journée. Ces cas surviennent lorsque le répondant a indiqué uniquement les dépenses totales au titre des aliments achetés dans un magasin pour une journée donnée.

Pour que l'imputation au niveau de l'entrée soit possible, le répondant doit avoir fourni certains renseignements additionnels. Par exemple, le répondant indique avoir acheté du bœuf, sans préciser le type : haché, en cubes, etc. Dans ce cas, le prix à l'unité est considéré comme le critère d'appariement le plus important pour trouver un donneur. Si le prix à l'unité est manquant, on utilise la dépense totale de même que d'autres variables pour apparier.

Certaines descriptions d'articles déclarés sont imprécises : *Légumes*, par exemple. Pour la première fois, en 2001, ces descriptions sont codées et imputées. Quelques fois, on ne peut pas trouver de donneurs pour certains receveurs, du fait d'une dépense élevée correspondant à la description. Pour l'imputation, on additionne pour chaque donneur possible toutes les dépenses d'une journée correspondant aux entrées appropriées (tous les légumes par exemple). Lorsqu'un donneur est trouvé avec le total des dépenses appropriées correspondant à la dépense élevée du receveur, on attribue toutes les entrées appropriées d'aliments du donneur au receveur.

Lorsque le répondant a indiqué uniquement les dépenses totales au titre des aliments pour une journée donnée, l'imputation est effectuée au niveau de la journée. On apparie les dépenses d'une journée du receveur à celles des donneurs. Les journées qui peuvent servir de donneurs sont celles qui ne comportent pas d'imputations au niveau des entrées sous-jacentes. Toutes les entrées pour une journée d'un donneur sélectionné sont attribuées au receveur. Les dépenses totales du receveur sont remplacées par les dépenses totales du donneur dans les cas où les deux chiffres ne correspondent pas exactement.

Enfin, il convient de souligner que les ménages qui n'ont fourni qu'une semaine de données sur deux ne font pas l'objet d'une imputation dans le cas de la semaine pour laquelle ils n'ont pas déclaré de données.

6. Pondération et estimation

Les estimations sont fondées sur le principe que chaque ménage de l'échantillon représente un certain nombre de ménages de la population cible, telle qu'elle a été définie dans la section 2. On attribue donc, à chaque ménage répondant, un poids d'enquête qui indique combien de ménages de la population sont représentés par ce ménage. Ce poids d'enquête est généralement le produit de trois facteurs : un poids d'échantillonnage qui incorpore les données du plan d'échantillonnage ; un facteur d'ajustement pour la non-réponse qui vise à compenser pour les ménages non-répondants et finalement un facteur

d'ajustement pour rajuster les estimations de certaines caractéristiques de l'échantillon en fonction d'estimations provenant de sources externes considérées plus fiables.

6.1 Poids d'échantillonnage

Le poids d'échantillonnage d'un ménage correspond à l'inverse de la probabilité qu'il soit inclus dans l'échantillon. Comme l'EDA est une enquête probabiliste, chaque ménage de la population cible a une probabilité connue d'être choisi dans l'échantillon. Si par exemple la probabilité de sélection d'un ménage est 1 sur 200, ce ménage aura un poids de 200.

Le plan d'échantillonnage définissant, entre autres, la méthode de répartition de l'échantillon, permet de déterminer le poids d'échantillonnage. L'EDA utilise le plan d'échantillonnage de l'EPA qui est un plan autopondéré dans chaque strate c'est-à-dire que les poids d'échantillonnage établis lors de l'élaboration du plan sont égaux dans chacune des strates. Dans la mesure où le plan d'échantillonnage et la répartition de l'échantillon ne sont pas modifiés, les poids initiaux pourraient être utilisés. Toutefois, les étapes de stabilisation et de sous-échantillonnage décrites dans la section 3.3 modifient les probabilités de sélection initiales. Les poids d'échantillonnage de l'EPA sont donc ajustés pour tenir compte de ces modifications.

Comme l'échantillon de l'EDA correspond à un sous-ensemble des 6 groupes de renouvellement de l'EPA, on détermine les poids d'échantillonnage de l'enquête en ajustant les poids d'échantillonnage de l'EPA en fonction du nombre de groupes de renouvellement retenus. Ce dernier facteur peut varier d'une strate à l'autre puisque le nombre de groupes de renouvellement retenu dans chaque strate est déterminé de sorte à satisfaire les besoins spécifiques à l'EDA en termes de répartition de l'échantillon. Étant donné que la taille de l'échantillon de l'EDA est très petite, on a eu la plupart du temps à sous-échantillonner dans un groupe de renouvellement de l'EPA. Les poids sont donc également ajustés pour tenir compte de cette étape.

6.2 Ajustement pour la non-réponse totale

Lorsque le répondant a omis de répondre à certaines questions du questionnaire FE2 ou qu'il a omis certaines informations concernant les dépenses rapportées sur le journal, les données manquantes sont imputées selon les méthodes décrites dans la section 5.2. Pour compenser la non-réponse totale, c'est-à-dire lorsqu'il a été impossible de contacter le ménage ou lorsque celui-ci a refusé de répondre, on ajuste les poids.

L'ajustement des poids pour la non-réponse totale est fondé sur l'hypothèse que les caractéristiques des ménages non-répondants sont similaires à celles des ménages répondants et, par conséquent, ces derniers peuvent être utilisés pour représenter l'ensemble des ménages : les répondants et les non-répondants. Afin de procéder à ce rajustement, l'échantillon est d'abord subdivisé en classes de non-réponse définies de sorte à augmenter les chances que les répondants et les non-répondants possèdent des caractéristiques semblables. Ces

caractéristiques sont choisies en fonction de ce que l'enquête cherche à mesurer.

Les classes de non-réponse ont été définies par trimestre car la taille de l'échantillon ne permettait pas de définir des classes pour chacun des mois.

Les classes de non-réponse par trimestre correspondent à différents niveaux d'urbanisation dans chaque région. Les niveaux d'urbanisation sont généralement les suivants : les principales régions métropolitaines, les régions urbaines de plus d'un million d'habitants, les autres régions urbaines plus petites et, finalement, les régions rurales. Dans certaines régions, il peut être nécessaire de regrouper certains niveaux parce que l'échantillon est trop petit. Dans chacune des trois villes des territoires, il n'y a que quatre classes de non-réponse, chacune représentant un trimestre.

Les strates de ménage à revenu élevé constituent également des classes de non-réponse particulières dans chacune des régions où elles existent. On notera que les classes de non-réponse ne se chevauchent pas et que, réunies, elles couvrent l'ensemble de la population cible.

Pour chaque classe de non-réponse, un facteur de compensation pour la non-réponse est calculé et défini comme étant l'inverse du taux de réponse pondéré de la classe. Ce facteur est, en d'autres mots, le rapport du nombre de ménages échantillonnés pondérés, afin qu'ils représentent tous les ménages de la classe, sur le nombre de ménages répondants pondérés. Pour éviter que les facteurs d'ajustement pour la non-réponse ne soient trop élevés, on regroupe autant que possible certaines classes de non-réponse lorsque le facteur d'ajustement est supérieur à 2.

6.3 Ajustement à l'information auxiliaire

En principe, en multipliant le poids d'échantillonnage par le facteur d'ajustement pour la non-réponse, on peut produire des estimations au niveau de la population. Toutefois, il est possible d'utiliser des données auxiliaires sur la population cible pour améliorer les estimations de l'enquête. Si ces données auxiliaires sont corrélées avec les principales caractéristiques mesurées par l'enquête, des estimations plus fiables peuvent être produites. Par exemple, les dépenses alimentaires sont corrélées avec la taille du ménage. Une mauvaise répartition de l'échantillon par rapport à la taille du ménage aurait un impact sur les estimations de dépenses. À partir de données auxiliaires sur le nombre de ménages selon la taille, on peut ajuster les poids de sorte à obtenir une meilleure distribution du nombre de ménages.

Dans l'EDA, on utilise plusieurs sources de données auxiliaires pour ajuster les poids. D'abord, les estimations post-censitaires de population, produites par la Division de la démographie de Statistique Canada, fournissent des comptes du nombre de personnes selon différents groupes d'âge et selon le sexe par province ou territoire. Ces comptes correspondent à des estimations de la population à une période donnée établies à partir des données du recensement et de l'information provenant de dossiers administratifs tels que les naissances, décès, immigration, émigration, etc.

Ces estimations démographiques sont ajustées pour tenir compte de certaines exclusions spécifiques aux enquêtes auprès des ménages comme les personnes vivant en institution. Conceptuellement, elles diffèrent légèrement de la population cible de l'EDA 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.

Ces comptes sont ensuite ajustés pour refléter la population cible de l'EDA à la fin de l'année de référence de l'enquête. Cet ajustement comprend l'exclusion des régions éloignées puisqu'elles ne sont pas couvertes par l'enquête. Des comptes pour six groupes d'âge pour chacune des régions⁶ sont utilisés pour ajuster les poids. Les comptes démographiques du nombre de personnes de 18 ans ou plus et du nombre de personnes de moins de 18 ans dans certaines régions métropolitaines⁷ sont également utilisés.

Les estimations du nombre de ménages selon la taille (une, deux ou trois personnes ou plus) pour chaque région ainsi que du nombre de ménages selon certains types de ménages sont également utilisées pour ajuster la représentativité de l'échantillon dans ces groupes. Pour les types de ménage, on utilise plus spécifiquement le nombre de ménages constitués d'une famille monoparentale et le nombre de ménages composés de parents avec des enfants jamais mariés.

Pour la ville de Whitehorse, on utilise les estimations, dérivées⁸ à partir de celles du Yukon, du nombre de ménages selon la taille (une, deux ou trois personnes ou plus). Pour les villes de Yellowknife et Iqaluit, on utilise seulement l'estimation du nombre de ménages. Dans le cas de Yellowknife, l'estimation a été dérivée à partir de celle des Territoires du Nord-Ouest. Dans le cas d'Iqaluit, on a utilisé l'estimation préliminaire du nombre de ménages suite au recensement de 2001 en raison de l'expansion inhabituelle de cette ville suite à la création du territoire du Nunavut en 1999.

L'ajustement des poids pour tenir compte de tous les comptes décrits ci-dessus est effectué simultanément en utilisant une variante de l'estimateur de régression généralisé (ERG) fondée sur la méthode de pondération proposée par Lemaître et Dufour [3]. Cette méthode permet d'obtenir la concordance des estimations de l'enquête avec les estimations provenant des sources auxiliaires tout en assurant que tous les membres du ménage auront encore un poids identique après l'ajustement. Le facteur d'ajustement calculé par l'ERG est ensuite appliqué au

6. À Whitehorse, on utilise seulement deux groupes : moins de 18 ans et 18 ans ou plus. À Yellowknife et Iqaluit, il n'y a aucun ajustement de fait au niveau du nombre de personnes.

7. Il s'agit des régions métropolitaines suivantes : St. John's, Halifax, Saint John, Québec, Montréal, Ottawa, Toronto, Winnipeg, Regina, Saskatoon, Calgary, Edmonton, Vancouver et Victoria.

8. Les estimations d'une ville d'un territoire sont obtenues à partir des estimations équivalentes de son territoire multipliées par la proportion qu'occupait cette ville par rapport à celui-ci selon le recensement de 1996.

poids d'échantillonnage et au facteur d'ajustement pour la non-réponse pour produire le poids final du ménage.

Dans les cas de Yellowknife et Iqaluit, une légère variante à cette méthode a dû être apportée. Bien qu'on ne contrôlait que le nombre de ménages, les estimations du nombre de ménages par taille de ménage étaient inacceptables suite à l'application de la méthode telle quelle. Pour plus de détails, voir la référence [4].

6.4 Autres ajustements

Au cours de l'entrevue en personne sur le questionnaire (FE2), on a demandé aux répondants d'indiquer les dépenses du ménage pour les aliments achetés au magasin pendant une période de quatre semaines. L'estimation des dépenses obtenue par l'entremise du questionnaire (FE2) est supérieure de 8,3 % à celle obtenue avec les données des journaux (FE3). On considère que le fardeau imposé en demandant au répondant de remplir les deux journaux (FE3), qui est principalement de rapporter une grande variété d'articles sur une période de deux semaines, peut entraîner la fatigue du répondant et avoir un impact sur la qualité des réponses obtenues. Suivant l'hypothèse que l'estimation des aliments achetés au magasin selon le questionnaire est plus fiable que l'estimation correspondante du journal⁹ en raison du fardeau associé au journal, les valeurs de chaque ménage pour les aliments achetés au magasin localement et à l'occasion de déplacements d'une seule journée ont été multipliées par 1.083.

Afin de tenir compte du caractère saisonnier des dépenses alimentaires, la collecte des données est effectuée au cours des 12 mois de l'année. La stratégie adoptée afin de réaliser ceci est de diviser l'échantillon en 12 sous-ensembles égaux, soit un sous-ensemble par mois. Un autre ajustement potentiel à ce genre d'enquête serait un ajustement temporel. Par exemple, des problèmes de collecte rencontrés pour un mois en particulier pourraient justifier ce genre d'ajustement. Jusqu'à maintenant, avec les ajustements déjà présents, un ajustement temporel n'a pas été jugé nécessaire.

6.5 Estimations

L'EDA recueille auprès des répondants leurs dépenses alimentaires sur une période de deux semaines. Il arrive que certains répondants ne fournissent des dépenses que pour une semaine. Afin de calculer l'estimation de la dépense hebdomadaire moyenne par ménage pour un article, on calcule l'estimation de sa dépense totale divisée par l'estimation du nombre de journaux. De même, afin de calculer le pourcentage répondant d'un article, on calcule l'estimation du nombre de journaux ayant rapportés une dépense pour l'article divisée par l'estimation du nombre de journaux.

9. D'autres analyses appuient cette hypothèse. Pour plus de détails, voir la référence [5].

7. Estimation de l'erreur d'échantillonnage

Après avoir calculé les estimations, il faut en déterminer la fiabilité ou, en d'autres mots, estimer l'erreur d'échantillonnage associée à chaque estimation. L'erreur-type ou le coefficient de variation (CV), qui est simplement l'erreur-type exprimée en tant que pourcentage de l'estimation, est la mesure courante de l'erreur d'échantillonnage. Elle correspond au degré de variation que l'on observe dans les estimations suite au choix d'un échantillon particulier plutôt qu'un autre. Comme l'EDA est une enquête probabiliste, on peut estimer l'erreur-type des estimations.

Dans l'EDA, on utilise l'estimateur jackknife pour estimer l'erreur-type. Cette méthode consiste à créer des répliques d'échantillon à partir des données de l'EDA. On crée autant de répliques qu'il y a d'unités primaires d'échantillonnage (UPE)¹⁰ en enlevant à tour de rôle une UPE de l'échantillon. Chaque UPE fait partie d'une strate et lorsque l'UPE est enlevée, les poids d'échantillonnage des autres UPE de la strate sont ajustés pour compenser ce retrait. On recalcule ensuite les estimations finales en appliquant les ajustements aux données auxiliaires décrits à la section 6.3 aux répliques. En répétant cette opération pour chaque UPE de l'échantillon, on obtient autant d'estimations qu'il y a d'UPE. La variabilité entre ces estimations sert à estimer l'erreur-type de l'estimation provenant de l'échantillon complet. La formule détaillée se trouve à l'annexe A.

7.1 Modèle pour dériver une approximation du CV pour les estimations des domaines

Pour des raisons opérationnelles, il n'est pas possible de produire les CV pour tous les différents niveaux d'agrégation qui peuvent intéresser les utilisateurs, par exemple, par type de ménage, par niveau d'urbanisation ou pour certaines régions métropolitaines. L'approche qui est suggérée aux utilisateurs de l'EDA est de calculer une approximation du CV en utilisant une relation entre le nombre de journaux dans lesquels des dépenses pour un article ont été déclarées et le CV à un niveau agrégé (généralement au niveau national). Cette relation, basée sur la tendance du CV à croître proportionnellement à une diminution de la racine carrée du nombre de journaux ayant un montant différent de zéro, est illustrée dans l'annexe B.

7.2 Modèle pour dériver une approximation du CV à partir des fichiers de microdonnées

Les utilisateurs des fichiers de microdonnées peuvent se servir d'une autre approche pour dériver une approximation du CV d'une estimation, qui sera généralement plus performante que celle décrite ci-dessus. Cette approche, relativement simple à appliquer, est décrite plus en détails dans la référence [6]. 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.

10. Les UPE, dans le cas de l'EDA, sont les grappes. Pour plus de détails, voir la section 3.

8. Suppression de données et confidentialité

Certaines mesures sont prises pour assurer que les estimations produites à partir de l'EDA sont suffisamment fiables pour être publiées et que l'anonymat des ménages répondants est respecté.

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

Comme le coefficient de variation est un indicateur de la fiabilité des données, idéalement on l'utiliserait pour déterminer si les estimations devraient être publiées ou non. Les estimations dont le CV est estimé à plus de 33% ne sont pas suffisamment fiables pour être publiées.

Par contre, le grand nombre d'estimations produites à partir de l'EDA fait en sorte qu'il n'est pas possible de calculer les CV pour chacune de ces estimations. Une étude effectuée à partir de données des enquêtes 1996 et 2001 a démontré que les CV atteignent en général 33% lorsque le nombre de journaux ayant un montant différent de zéro est autour de 40. Cette règle est donc appliquée pour déterminer si les estimations de l'EDA peuvent être publiées ou non. Comme il s'agit d'une règle approximative, certaines estimations seront publiées même si le CV est supérieur à 33% et certaines estimations ne seront pas publiées malgré un CV inférieur à 33%. Une évaluation de la performance de cette règle se trouvent dans le rapport sur la qualité des données de 2001 [7].

On doit noter que même si les estimations pour une certaine catégorie d'aliments ne sont pas diffusées parce que moins de 40 journaux ont un montant différent de zéro pour cette catégorie, les données sont retenues dans les estimations des composantes agrégées.

8.2 Confidentialité des fichiers de microdonnées

Quoique des fichiers de microdonnées à grande diffusion soient produits à partir des données recueillies par l'EDA, ceux-ci sont différents des fichiers utilisés par Statistique Canada pour la diffusion des estimations. Ces différences proviennent principalement d'une série de mesures prises pour protéger l'anonymat des ménages qui ont répondu à l'enquête.

9. Changements dans la méthodologie de l'enquête

La méthodologie de l'enquête est demeurée semblable à celle de 1996 à l'exception des stratégies d'imputation et d'ajustement aux données auxiliaires dans la pondération. Le plan d'échantillonnage a également été modifié suite à la diminution de la taille d'échantillon pour des raisons budgétaires. On a maintenant comme objectif d'obtenir des estimations avec des variances raisonnables au niveau des cinq grandes régions canadiennes au lieu des dix provinces.

La stratégie de pondération a été modifiée dans le cadre d'un projet qui vise à harmoniser les ajustements aux données auxiliaires dans les enquêtes de Statistique Canada liées à la statistique du revenu. Pour l'EDA, les principaux changements ont été l'utilisation de plus de groupes d'âge pour les comptes démographiques et l'ajout de comptes au niveau des types de ménages. Afin d'assurer la comparabilité dans les analyses des tendances, l'EDA 1996 a été repondérée à l'aide de cette nouvelle stratégie et d'estimations démographiques basées sur le recensement de 1996 au lieu de celles basées sur le recensement de 1991 utilisées lors de la pondération originale.

Bibliographie

- [1] *Méthodologie de l'enquête sur la population active du Canada*, N° 71-526-XPB au catalogue.
- [2] Khan, K. (2002), *Imputation of the 2001 Food Expenditure Survey*, Division des méthodes d'enquêtes auprès des ménages, Document interne, Statistique Canada.
- [3] Lemaître et Dufour (1987), *Une méthode intégrée de pondération des personnes et des familles*, Technique d'enquête, Vol.13, n° 2, pp.211-220, Statistique Canada.
- [4] Arsenault, S. et Auger, S. (Février 2003), *Pondération de l'Enquête sur les dépenses alimentaires 2001*, Division des méthodes d'enquêtes auprès des ménages, Document interne, Statistique Canada.
- [5] *FAMEX Reference Paper (1995)*, Division de l'analyse des enquêtes auprès des ménages et sur le travail, Document interne, Statistique Canada.
- [6] Beaumont, J.-F. (2000), *Estimation de variance pour un fichier de microdonnées à grande diffusion provenant d'une enquête complexe*, Division des méthodes d'enquêtes auprès des ménages, Document de travail HSMD-2000-002F/A, Statistique Canada.
- [7] Auger, S., Khan, K., Nadon, S. et Tremblay, J. (2005), *Enquête sur les dépenses alimentaires de 2001 – Indicateurs de qualité des données*, N° 62F0026MIF-2005003 au catalogue, Division des méthodes d'enquêtes auprès des ménages, Statistique Canada.
- [8] *Dépenses alimentaires au Canada (2001)*, N° 62-554-XIF au catalogue.

Annexe A

Formule pour le calcul de la variance des estimations à partir du jackknife

Dans la procédure jackknife pour estimer la variance des estimations, on mesure la variabilité entre les estimations à l'aide de la formule suivante :

$$Var(\hat{Y}) = \sum_{h=1}^H \frac{n(h) - 1}{n(h)} \sum_{i=1}^{n(h)} (\hat{Y} - \hat{Y}_{(hi)})^2$$

où

$n(h)$ est le nombre d'UPE dans la strate h

$\hat{Y}_{(hi)}$ est l'estimation de Y quand l'UPE i de la strate h est enlevée.

L'erreur-type est la racine carrée de la variance.

Annexe B

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 dépense hebdomadaire 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 niveau d'urbanisation, ...) à partir de l'équation suivante :

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

où

n = le nombre de journaux ayant un montant différent de zéro pour cette caractéristique

= le nombre de journaux * l'estimation du pourcentage déclarant¹¹

100

n_d = le nombre de journaux ayant un montant différent de zéro pour cette caractéristique dans le domaine d

= le nombre de journaux dans le domaine d * l'estimation du pourcentage déclarant dans le domaine d

100

Généralement on utilise le CV, le nombre de journaux et le pourcentage déclarant 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 région (par ex. région métropolitaine), il est préférable d'utiliser ces valeurs à l'échelle régionale puisque les CV régionaux sont publiés pour l'EDA.

Vous trouverez un exemple de calcul à la page suivante.

11. Pourcentage de journaux contenant un montant différent de zéro pour cette caractéristique.

Exemple de calcul

Dans le tableau 2b de la publication de l'EDA 2001 [8], on retrouve différentes estimations selon les tranches de revenu, par contre les CVs associés à ces estimations ne se retrouvent pas dans la publication.

Le CV des dépenses hebdomadaires moyennes des ménages pour les aliments achetés au magasin dans le cas de tous les ménages de la tranche de revenu la plus basse, selon la formule ci-dessus et à l'aide des tableaux 2a, 2c et du tableau technique 2b, est

$$= 1,04\% \times \sqrt{10140/1756}$$

$$= 1,04\% \times 2,40$$

$$= 2,50\%$$

où n est le nombre de journaux ayant un montant différent de zéro

$$= (\text{nombre de journaux} \times \text{pourcentage déclarant}) / 100$$

$$= (11\ 034 \times 91,9) / 100$$

$$= 10\ 140$$

et n_d est le nombre de journaux des ménages de la tranche de revenu la plus basse ayant un montant différent de zéro

$$= (\text{nombre de journaux des ménages de la tranche de revenu la plus basse} \times \text{pourcentage déclarant pour les ménages de la tranche de revenu la plus basse}) / 100$$

$$= (2\ 011 \times 87,3) / 100$$

$$= 1\ 756.$$