



No. 89M0019GPF au catalogue

ENQUÊTE SUR LA DIVERSITÉ ETHNIQUE (2002)

Guide de l'utilisateur du fichier de microdonnées à grande diffusion



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 sociale et autochtone, Statistique Canada, Ottawa, Ontario, K1A 0T6 (téléphone : (613) 951-5979).

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 bibliothèques de dépôt	1 800 700-1033
Télécopieur pour le Programme des bibliothèques de dépôt	1 800 889-9734
Renseignements par courriel	infostats@statcan.ca
Site Web	www.statcan.ca

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.



Statistique Canada
Division de la statistique sociale et autochtone

Enquête sur la diversité ethnique (2002)

Guide de l'utilisateur du fichier de microdonnées à grande diffusion

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2005

Tous droits réservés. L'utilisation de ce produit est limitée au détenteur de licence et à ses employés. Le produit ne peut être reproduit et transmis à des personnes ou organisations à l'extérieur de l'organisme du détenteur de licence.

Des droits raisonnables d'utilisation du contenu de ce produit sont accordés seulement à des fins de recherche personnelle, organisationnelle ou de politique gouvernementale ou à des fins éducatives. Cette permission comprend l'utilisation du contenu dans des analyses et dans la communication de résultats et conclusions de ces analyses, y compris la citation de quantités limitées de renseignements complémentaires extraits du produit de données dans ces documents. Cette documentation doit servir à des fins non commerciales seulement. Si c'est le cas, la source des données doit être citée comme suit : Source (ou « Adapté de », s'il y a lieu) : Statistique Canada, nom du produit, numéro au catalogue, volume et numéro de l'issue, période de référence et page(s). Autrement, les utilisateurs doivent d'abord demander la permission écrite aux Services d'octroi de licences, Statistique Canada, Ottawa, Ontario, Canada, K1A 0T6.

Mai 2005

89M0019GPF au catalogue

Périodicité : Occasionnel

Ottawa

This publication is available in English (Catalogue no. 89M0019GPE)

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.

Signes conventionnels

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

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

Table des matières

1.0	Introduction.....	3
2.0	Méthodologie de l'enquête	3
2.1	BASE DE SONDAGE	3
2.2	POPULATION CIBLE	3
2.3	PÉRIODE DE RÉFÉRENCE ET COLLECTE DES DONNÉES	4
2.4	PLAN D'ÉCHANTILLONNAGE.....	4
2.5	TAILLE ET SÉLECTION DE L'ÉCHANTILLON	6
2.6	TRAITEMENT DES QUESTIONNAIRES	6
2.6.1	<i>Définition du statut de réponse.....</i>	7
2.6.2	<i>Vérification de la validité des réponses</i>	7
2.7	EXCLUSION DE LA POPULATION AUTOCHTONE	9
3.0	Estimation	10
3.1	PONDÉRATION	10
3.2	LIGNES DIRECTRICES CONCERNANT LA PONDÉRATION	12
3.3	TYPES D'ESTIMATION	12
3.3.1	<i>Estimations qualitatives.....</i>	12
3.3.2	<i>Estimations quantitatives</i>	12
3.4	LIGNES DIRECTRICES POUR L'ANALYSE.....	13
4.0	Lignes directrices concernant la diffusion et fiabilité des données	13
4.1	TAILLE MINIMALE D'ÉCHANTILLON POUR LA PRODUCTION D'ESTIMATIONS.....	14
4.2	LIGNES DIRECTRICES CONCERNANT LA VARIABILITÉ D'ÉCHANTILLONNAGE	14
4.2.1	<i>Erreurs non dues à l'échantillonnage</i>	14
4.2.2	<i>Erreurs d'échantillonnage</i>	15
4.3	ARRONDISSEMENT	16
4.3.1	<i>Lignes directrices concernant l'arrondissement</i>	16
4.3.2	<i>Méthode d'arrondissement classique</i>	17
5.0	Autre produits à partir des données de l'EDE.....	17

1.0 Introduction

L'Enquête sur la diversité ethnique (EDE) a été effectuée par Statistique Canada conjointement avec le ministère du Patrimoine canadien. Cette enquête poursuit deux objectifs. En premier lieu, les données nous aidera à mieux comprendre la façon dont les antécédents des personnes influencent leur participation à la vie sociale, économique et culturelle du Canada. En second lieu, les renseignements recueillis aideront à mieux comprendre comment les Canadiens de différentes origines ethniques interprètent et déclarent leur appartenance ethnique. Les renseignements recueillis au moyen de l'enquête serviront à Patrimoine Canada à éclairer l'élaboration de politiques et de programmes.

Le présent document a été élaboré afin de faciliter l'utilisation du fichier de microdonnées à grande diffusion (FMGD) de l'Enquête sur la diversité ethnique. Il documente la méthodologie de l'enquête, la qualité des données et d'autres questions relatives à l'analyse et à la diffusion des données.

Toutes questions au sujet du FMGD de l'EDE ou de son utilisation devraient être adressées à :

Services à la clientèle
Division de la statistique sociale et autochtone
Statistique Canada
Immeuble Jean Talon, 7e étage
Parc Tunney
Ottawa (Ontario)
K1A 0T6

Téléphone : (613) 951 5979
Télécopieur : (613) 951 0387
Courriel : sasd-dssea@statcan.ca

2.0 Méthodologie de l'enquête

Le présent chapitre présente une brève description des méthodes utilisées dans le cadre de l'Enquête sur la diversité ethnique (EDE). Il couvre également les principaux aspects de la qualité des données ainsi que les lignes directrices pour l'analyse et la diffusion des données. Ces renseignements visent à aider les utilisateurs à comprendre les forces et les limites des données de même que la façon de bien les utiliser.

2.1 Base de sondage

Une base de sondage permet d'avoir accès à la population devant être comprise dans une enquête. L'Enquête sur la diversité ethnique est une enquête post censitaire, ce qui signifie que l'échantillon a été obtenu à partir du recensement. Les répondants de l'EDE ont été sélectionnés d'après les réponses qu'ils ont fournies à certaines questions lors du dernier recensement qui a eu lieu le 15 mai 2001. Ainsi, la base de sondage de l'EDE était donc constituée de la liste des personnes qui ont répondu au questionnaire long (2B) du recensement de 2001.

2.2 Population cible

La population cible de l'EDE est constituée des personnes de 15 ans et plus vivant dans les logements privés des dix provinces canadiennes. Tout comme c'était le cas lors du recensement, les citoyens Canadiens, les immigrants reçus et les résidents non permanents (les détenteurs de permis d'études, de permis de travail ou de permis ministériels, les revendicateurs du statut de réfugié et les membres de leur famille qui demeurent au Canada avec elles) font partie de la population cible. Toutefois, les groupes suivants en sont exclus:

- les personnes de moins de 15 ans;
- les personnes vivant dans des logements collectifs (hôtels, maisons de repos, hôpitaux, camps militaires ou chantiers de travail, prisons, foyer pour personnes âgées, etc.);
- les réserves indiennes;
- les personnes ayant déclaré une origine ou une identité autochtone lors du Recensement de 2001;
- les territoires et les régions éloignées.

La population cible de l'EDE représente 23 092 643 personnes de la population canadienne. Parmi celle-ci, 57 242 individus ont été sélectionnés pour l'enquête. Au total, de ces individus, 42 476 personnes ont participé à l'EDE.

Pour assurer la confidentialité des répondants, un sous-échantillon a été tiré à partir des répondants de l'EDE pour le fichier de microdonnées. Le FMGD de l'EDE inclut 41 695 répondants représentant la même population cible de 23 092 643 personnes.

2.3 Période de référence et collecte des données

La période de référence de l'enquête correspond à celle de la collecte de données qui a eu lieu entre les mois d'avril et août 2002.

Les employés des bureaux régionaux de Statistique Canada ont recueilli les données de l'EDE au moyen du logiciel Blaise et de la méthode d'interview téléphonique assistée par ordinateur (ITAO). Cette technologie permet de recueillir des données de grande qualité à prix raisonnable dans le cas d'enquêtes ayant de longs et complexes questionnaires comme celui de l'EDE.

La durée moyenne des interviews était de 35 à 45 minutes, selon les caractéristiques du répondant. En plus du français et de l'anglais, les interviews ont été effectuées dans les sept langues non officielles suivantes : le mandarin, le cantonnais, l'italien, le punjabi, le portugais, le vietnamien et l'espagnol. Les entrevues menées dans ces langues étaient généralement plus longues que celles menées en anglais et en français.

Les réponses par procuration (ou par personne interposée) n'étaient pas admises. Ainsi, seule la personne sélectionnée devait être jointe et interviewée.

2.4 Plan d'échantillonnage

L'enquête sur la diversité ethnique est une enquête probabiliste, c'est-à-dire, qu'un échantillon aléatoire a été sélectionné pour représenter la population cible. Le plan d'échantillonnage utilisé pour l'EDE est un plan stratifié à deux phases. La phase I, le recensement de 2001, consistait à distribuer un questionnaire long à 1 ménage sur 5, en moyenne, au Canada. La phase II consistait à choisir un échantillon parmi les répondants de la phase I en fonction des réponses fournies aux questions sur l'origine ethnique, le lieu de naissance du répondant ainsi que le lieu de naissance de ses parents.

Pour atteindre les objectifs de l'enquête, il était fondamental de cibler deux groupes principaux et d'assurer une bonne représentativité de ces sous populations dans l'échantillon. Ces deux groupes principaux incluent les personnes ayant déclaré l'origine « Canadienne » au recensement de 2001 et les personnes qui ne sont pas d'origine canadienne, britannique ou française.

Les réponses à l'origine ethnique ont donc été ventilées pour former les deux principales catégories :

- **CBFA+**

où C = Canadien, B = Britannique, F = Français et A+ = Américain, Australien et/ou Néo-zélandais

- **Non-CBFA+**

Toutes les autres réponses comprenant au moins une origine autre que CBFA+

Ces deux catégories ont été subdivisées selon que les réponses comprenaient ou non l'origine « canadienne ». La catégorie des Non-CBFA+ a, de plus, été subdivisée en origines européennes (par exemple, Allemand, Italien, Hollandais, Portugais) et non européennes (par exemple, Chinois, Jamaïcain, Libanais, Iranien).

Finalement, les questions portant sur le lieu de naissance du répondant et celui de ses parents ont servi à former le statut générationnel du répondant. La **génération 1** représente les répondants nés à l'extérieur du Canada. La **génération 2** représente les répondants nés au Canada ayant au moins un parent né à l'extérieur du Canada. La **génération 3** représente les répondants nés au Canada dont les deux parents sont également nés au Canada. Le cas échéant, les groupes ou strates formés par les générations ont été regroupés afin d'obtenir un nombre suffisamment élevé de personnes dans chaque strate.

La population cible de l'EDE a donc été répartie entre les strates suivantes :

- **CBFA+**

- Canadiens seulement
 - Génération 1 et 2
 - Génération 3 et plus
- Canadiens et BFA+
 - Génération 1 et 2
 - Génération 3 et plus
- BFA+ seulement
 - Génération 1 et 2
 - Génération 3 et plus

- **Non-CBFA+**

- Autres Européens et Canadiens
 - Génération 1 et 2
 - Génération 3 et plus
- Autres non-Européens et Canadiens
 - Toutes les générations
- Autres Européens
 - Génération 1
 - Génération 2
 - Génération 3 et plus
- Autres non-Européens
 - Génération 1
 - Génération 2 et plus

(Les réponses multiples à l'origine ethnique contenant à la fois des origines européennes et non européennes furent classifiées avec les autres origines non européennes en raison de la dominance de la présence des minorités visibles.)

Cette stratification a contribué à simplifier grandement la stratégie d'échantillonnage de l'EDE car cette façon de faire éliminait le besoin d'avoir à classer plusieurs centaines d'origines ethniques simples et multiples ainsi que de nombreux lieux de naissance mentionnés dans le recensement.

2.5 Taille et sélection de l'échantillon

En raison des objectifs visés par l'enquête sur la diversité ethnique et du besoin en données pour certaines sous populations, notamment les personnes de première et de deuxième génération, la répartition de l'échantillon a été établie à 1/3 pour les origines CBFA+ et à 2/3 pour les origines non-CBFA+. Cette distribution nous assurait que les personnes qui ne sont pas d'origine canadienne, britannique ou française seraient bien représentées dans l'échantillon. Toutefois, l'enquête n'était pas une enquête sur les immigrants seulement et par conséquent, elle devait s'adresser à la population canadienne en générale.

Afin de permettre des analyses, l'enquête visait à obtenir au moins 40 000 répondants. Puisqu'un taux de réponse de 70 % était prévu au départ, un échantillon d'environ 57 200 personnes était nécessaire pour atteindre cet objectif. Le coefficient de variation (c.v) a été utilisé comme mesure de la fiabilité. Dans chaque strate, la taille initiale de l'échantillon a été déterminée en fonction d'une proportion minimale à estimer de 4 %, un c.v. maximum de 12,5 %, un effet de plan de 1,2 et des taux de réponses estimés à partir des résultats du test pilote de septembre 2001.

Puisque la stratification de l'échantillon a été effectuée à l'échelon du Canada, la taille initiale résultant de ces calculs était inférieure à celle souhaitée. La différence a donc été répartie proportionnellement à la taille des strates afin d'arriver à la taille requise.

Une fois les strates établies, la base de sondage a été triée selon la province, la circonscription électorale, le secteur de dénombrement et le ménage afin d'assurer une bonne répartition géographique de l'échantillon et réduire le nombre de personnes sélectionnées à l'intérieur d'un même ménage. Un échantillon systématique a ensuite été sélectionné indépendamment dans chaque strate. (L'échantillonnage systématique comprend la sélection d'unités à partir d'une liste, selon un intervalle de sélection établi au préalable.)

L'échantillon final de l'EDE était composé de 57 242 personnes. De ce nombre, 42 476 ont répondu à l'enquête, ce qui représente un taux de réponse global de 75,6 % si l'on tient compte des 1 057 personnes classées hors du champ de l'enquête. De ce taux, 73,1 % sont des réponses complètes et 3,2 % des réponses partielles. Cela signifie qu'en général, lorsqu'une personne commençait l'interview, elle finissait de répondre à toutes les questions de l'enquête. Les réponses complètes représentent donc 96,8 % de tous les répondants à l'enquête. La non-réponse totale représente 24,4 % de l'échantillon. Les taux de réponse par strate, définis précédemment, varient entre 72 % et 80 %. Comme on pouvait s'y attendre, les premières générations ont obtenu un taux de réponse moins élevé, soit 73 % comparativement à 77 % pour les deuxièmes et troisièmes générations et plus.

781 des 42 476 répondants de l'EDE n'ont pas été inclus dans le fichier de microdonnées en raison du risque de divulgation associé avec ceux-ci. Cependant, des ajustements ont été faits aux poids des individus inclus dans le FMGD pour assurer que ce fichier soit représentatif de la population cible de l'EDE. Ainsi, le fichier de microdonnées inclut 41 695 répondants, représentant la population cible.

2.6 Traitement des questionnaires

Une fois le codage des réponses de la catégorie « Autre – Précisez » terminé, des contrôles et des vérifications de toutes les unités échantillonnées ont été effectués. Ces contrôles et vérifications comprenaient principalement la lecture des remarques et des notes laissées par les interviewers dans les questionnaires, la vérification de chaque réponse afin de déterminer les entrées manquantes, non valables ou incohérentes ainsi que la vérification des codes finaux de chaque questionnaire attribués par les interviewers. Un fichier maître contenant les données épurées et les poids de l'enquête a ensuite été créé.

2.6.1 Définition du statut de réponse

Une des étapes préliminaires à la pondération est de vérifier les codes finaux des questionnaires dans le but d'attribuer un statut de réponse à chaque unité échantillonnée. Pour ce faire, il doit y avoir un enregistrement pour chaque personne sélectionnée dans l'échantillon. Puis, un des statuts suivants est attribué à chaque enregistrement :

- i) réponse complète
- ii) réponse partielle
- iii) non-réponse totale
- iv) hors du champ d'enquête

Il y a une **réponse complète** lorsqu'on recueille toutes ou presque toutes les données pour une unité échantillonnée. Dans l'EDE, un enregistrement recevait le statut de réponse complète quand on avait obtenu au moins **80 %** de réponses valides aux questions obligatoires (au moins 31 des 39 questions posées à tous les répondants) **et** au moins une réponse valide à l'une des trois questions suivantes : l'ascendance ethnique (ID_Q010), l'identité ethnique (ID_Q100) et le lieu de naissance (BK_Q010).

Il y a une **réponse partielle** lorsqu'on ne recueille que certaines données pour une unité échantillonnée. Dans l'EDE, un enregistrement recevait le statut de réponse partielle quand on avait obtenu au moins 31 % de réponses valides aux questions obligatoires (au moins 12 des 39 questions posées à tous les répondants) mais sans excéder 79,9 % des questions obligatoires ou quand on avait obtenu moins de 31 % de réponses valides aux questions obligatoires et une réponse valide à chacune des trois questions suivantes : l'ascendance ethnique (ID_Q010), l'identité ethnique (ID_Q100) et le lieu de naissance (BK_Q010). Une réponse valide excluait les réponses « Ne sait pas » et « Refus ».

Il y a une **non-réponse totale** quand aucune donnée ou presque n'est recueillie pour une unité échantillonnée. Dans l'EDE, un enregistrement recevait le statut de non-réponse totale quand on avait obtenu moins de **31 %** de réponses valides aux questions obligatoires **et** aucune réponse valide aux trois questions suivantes : l'ascendance ethnique (ID_Q010), l'identité ethnique (ID_Q100) et le lieu de naissance (BK_Q010).

Une unité est **hors du champ d'enquête** lorsqu'elle figure dans la base de sondage mais que, selon les informations recueillies à l'enquête, elle ne fait pas partie de la population cible. Dans le cas de l'EDE, les personnes hors du champ d'enquête étaient les personnes décédées, les personnes de moins de 15 ans, les personnes ayant déménagé dans un des trois territoires, les personnes habitant une réserve indienne, un logement collectif, etc.

Les unités ayant reçu les statuts de « non-réponse totale » et de « hors du champ d'enquête » furent enlevées du fichier final de l'EDE. Les répondants ayant reçu les statuts de « réponse complète » ou de « réponse partielle » sont inclus dans le fichier maître de l'enquête.

2.6.2 Vérification de la validité des réponses

Une autre partie du traitement des questionnaires de l'EDE consistait à vérifier la validité de chaque réponse pour tous les répondants à l'enquête. Des vérifications de la cohérence et de l'enchaînement de questions ont été effectuées. Le but premier de ces vérifications était de déceler les erreurs, les lacunes et les incohérences dans les données de l'enquête afin de les corriger. Par exemple, une langue parlée au travail qui n'aurait pas été mentionnée auparavant à la question sur la connaissance des langues,

était considérée comme une réponse non valable ou incohérente. Dans la mesure du possible, ces incohérences étaient corrigées à partir des données supplémentaires et des notes des interviewers.

Dans le cadre de la vérification de la validité des réponses, chaque réponse fut classifiée dans l'une ou l'autre des catégories décrites ci-dessous :

- **Réponse valide**

Le répondant a donné une réponse à une question à laquelle il était supposé répondre. Une réponse valide est différente des réponses « Ne sait pas » et « Refus ».

Il faut mentionner que certaines réponses valides ont, lors du traitement, été attribuées à une réponse « impossible à coder ». Les réponses « impossible à coder » sont des réponses qui ne pouvaient pas être attribuées à un code existant, par les interviewers durant la collecte ou par les experts durant le codage, pour une des trois raisons suivantes :

(1) la réponse ne pouvait pas être codée. Par exemple, une origine ethnique telle que « xyzlocan » serait attribuée à une réponse « impossible à coder ».

(2) la réponse pouvait être attribuée à plus d'un code et il n'était pas possible de déterminer quel était le bon code. Par exemple, un lieu de naissance telle que « Albertville » serait attribué à une réponse « impossible à coder » s'il n'était pas possible de déterminer si cette réponse correspond à une ville du Canada, de la France ou des États-Unis.

(3) la réponse donnée à l'ascendance ou à l'identité ethnique indique que le répondant a compris la question, dans une certaine mesure, mais la réponse fournie n'est pas considérée comme un type de réponse appropriée pour cette question. Par exemple, les réponses « impossible à coder » pour l'ascendance ethnique incluent des réponses qui indiquent le nom de famille du répondant, son statut d'immigrant ou un commentaire général à propos de son histoire familiale.

- **Ne sait pas**

Le répondant a indiqué qu'il ne connaissait pas la réponse à la question. Dans le fichier de l'EDE, on attribue le code « 9 » à la réponse « ne sait pas ». Dans le cas d'une variable à deux chiffres, on attribue le code « 99 » et d'une variable à trois chiffres, le code « 999 », etc.

- **Refus**

Le répondant a refusé de répondre à la question. Dans le fichier de l'EDE, on attribue le code « 8 » à un refus. Dans le cas d'une variable à deux chiffres, on attribue le code « 98 » et d'une variable à trois chiffres, le code « 998 », etc.

- **Ne s'applique pas**

Le répondant n'a pas répondu à la question car il n'était pas supposé le faire à cause d'une réponse particulière qu'il a donné à la question filtre correspondante. Une question filtre, est une question située en tête d'un groupe de questions, qui sert à éliminer les répondants pour lesquels la question ou les questions subséquentes ne s'appliquent pas. Dans le fichier de l'EDE, on attribue le code « 7 » à la réponse « Ne s'applique pas ». Dans le cas d'une variable à deux chiffres, on attribue le code « 97 » et d'une variable à trois chiffres, le code « 997 », etc.

- **Question non posée**

Dans le fichier de l'EDE, deux types de réponse correspondent à une « question non posée ». Ces deux types de réponse sont « question non posée » et «cheminement inconnu ». On attribue le code « 6 » à une question non posée. Dans le cas d'une variable à deux chiffres, on attribue le code « 96 » et d'une variable à trois chiffres, le code « 996 », etc.

Question non posée (ou question non demandée) : Le répondant était supposé répondre à la question mais cette question n'a pas été posée. Cette situation est habituellement le résultat d'une circonstance spéciale ou d'une erreur de système. Par exemple, si un répondant refusait d'identifier les autres membres de son ménage dans le module d'entrée de l'EDE, mais qu'il mentionnait plus tard qu'il avait une épouse et que ce fait était inscrit par l'interviewer dans la section des notes de l'enquête, l'état matrimonial « marié » était imputée à ce répondant. Cependant, étant donné que cette imputation a été faite durant le traitement (après la collecte des données) aucune question se rapportant à son épouse n'a été posée à ce répondant durant l'enquête. Ainsi, le code correspondant à une « question non posée » était attribué à toutes les questions se rapportant à son épouse.

- **Cheminement inconnu :** Il n'était pas possible de savoir si le répondant devait ou non répondre à la question et la question filtre correspondante était aussi sans réponse. Cette situation est habituellement le résultat d'une question filtre qui n'a pas été posée. Dans le fichier de l'EDE, on d'abord attribué le code « -1 » à une question dont le cheminement était inconnu mais ce code fut regroupé avec les codes « 6 », « 96 », 996 », etc. dans le fichier final.

Durant les analyses, les utilisateurs auront besoin de définir leur domaine d'estimation (la population totale) pour chaque variable. Il sera important de déterminer s'ils doivent inclure ou exclure certains des codes « Ne sait pas », « Refus », « Ne s'applique pas » et « Question non posée ». L'inclusion ou l'exclusion de chaque code dépend du but de l'analyse. Toutefois, les utilisateurs qui désirent tenir compte en partie de la non-réponse partielle dans les analyses, devraient inclure dans le domaine de chaque variable les codes « Ne sait pas » et « Refus » mais exclure les cas correspondant aux codes « Ne s'applique pas » et « Question non posée ».

Les analystes qui désirent obtenir des résultats correspondants à ceux déjà publiés par Statistique Canada dans le rapport « Enquête sur la diversité ethnique : portrait d'une société multiculturelle », diffusé en septembre 2003, doivent généralement exclure les codes « Ne sait pas », « Refus », « Ne s'applique pas » et « Question non posée » de leur totaux. Bien que la façon de déterminer les domaines d'estimation peut varier selon les questions à l'étude, les pourcentages publiés dans ce rapport étaient habituellement calculés à partir d'un dénominateur basé sur les réponses valides seulement.

2.7 Exclusion de la population autochtone

Les personnes ayant déclaré une ascendance ou une identité autochtone au Recensement de 2001 ne faisaient pas partie de la population visée par l'Enquête sur la diversité ethnique. Ces personnes ont donc été exclues de la population cible de l'EDE. Les principales raisons de ces exclusions étaient le fardeau de réponse du répondant (ces personnes sont couvertes par une autre enquête post-censitaire : l'Enquête auprès des peuples autochtones) et la méthode de collecte des données (par entrevue téléphonique qui ne serait pas approprié dans la plupart des réserves). (Pour plus d'information sur l'Enquête auprès des peuples autochtones, veuillez vous référer au lien suivant : http://www.statcan.ca/cgi-bin/imdb/p2SV_f.pl?Function=getSurvey&SDDS=3250&lang=en&db=IMDB&dbg=f&adm=8&dis=2 ou contacter le service à la clientèle en utilisant l'information fournie à la section 1.0 de ce guide).

Cependant, malgré ces exclusions, un certain nombre de répondants à l'EDE (810 dans le fichier complet et 793 dans le FMGD) ont déclaré une ascendance autochtone aux questions ID_10 et ID_20 et/ou une identité autochtone à la question ID_100. Les raisons qui pourraient expliquer cette situation incluent les réponses par procuration au recensement, la déclaration de l'origine « canadienne » seulement au recensement et la fluidité du concept de l'ethnicité.

Avant la collecte, il avait été décidé que les répondants qui déclareraient une ascendance ou une identité autochtone à l'EDE seraient considérés comme étant hors du champ de l'enquête. L'application Blaise avait été programmée pour considérer comme hors du champ de l'enquête les enregistrements ayant une réponse associée à un groupe ethnique autochtone. Lorsqu'une réponse faisant partie de la liste prédéterminée des ascendances ou des identités autochtones était déclarée, l'application passait directement à la dernière question.

Néanmoins, au cours du traitement des données, il fut décidé que les répondants ayant déclaré une ascendance ou une identité autochtone seraient conservés dans le fichier des données de l'EDE. Plus de la moitié de ces personnes avaient répondu à toutes les questions de l'EDE parce que leurs réponses n'avaient pas été identifiées par la liste prédéterminée de Blaise. Il était adéquat de conserver ces répondants dans l'enquête puisqu'ils n'étaient pas couverts par l'Enquête auprès des peuples autochtones. De plus, des analyses précises portant sur cette population ont été planifiées.

Lors de la diffusion des données de l'Enquête sur la diversité ethnique du mois de septembre 2003, les répondants ayant déclaré une ascendance ou une identité autochtone, étaient exclus des analyses car cette population ne faisait pas partie à l'origine de la population cible de l'EDE. Cependant, aucune variable d'origine et d'identité autochtone n'a été incluse dans le fichier de microdonnées de l'EDE car de telles variables auraient représenté un risque de divulgation. Ainsi, les résultats d'analyses produites à partir du FMGD peuvent être légèrement différents que ceux présentés lors de la diffusion des données.

3.0 Estimation

Le principe sur lequel s'appuie l'estimation dans le cas d'un échantillon probabiliste comme celui de l'EDE veut que, outre elle-même, chaque personne faisant partie de l'échantillon en représente plusieurs autres qui ne font pas partie de l'échantillon. Par exemple, dans un Échantillon aléatoire simple de 2 % de la population, chaque personne représente 50 membres de la population (elle-même et 49 autres). Le nombre de personnes représentées par un répondant donné est ce que l'on appelle le poids ou le facteur de pondération de la personne.

Un facteur de pondération est inclus dans le fichier de microdonnées de l'EDE :

WGT_PUMF : Il s'agit du poids de base pour l'analyse au niveau des personnes, c.-à-d. pour calculer les estimations du nombre de personnes (faisant partie de la population cible) présentant une ou plusieurs caractéristiques données. WGT_PUMF devrait être utilisé pour calculer toutes les estimations. Par exemple, pour estimer le nombre de personnes noires, on fait la somme des valeurs de WGT_PUMF sur tous les enregistrements présentant cette caractéristique (VISMIND=4).

3.1 Pondération

Rappelons que l'EDE 2002 est une enquête visant les personnes et que le fichier de microdonnées contient les réponses au questionnaire et les renseignements connexes fournis par 41 695 répondants. Le calcul du poids pour le FMGD comptait quatre étapes :

1) Calcul du poids initial

La première étape consistait à attribuer un poids initial fondé sur le plan d'échantillonnage. Le poids initial correspond à l'inverse de la probabilité d'inclusion dans l'échantillon. Dans le cas de l'EDE de 2002, le

poids initial était le produit de deux éléments : le poids de recensement 2001 et l'inverse de la probabilité de sélection à la deuxième phase (poids d'échantillonnage). Suivant ce calcul, on a tenu compte des gens sélectionnés par erreur et de ceux omis durant la sélection de l'échantillon et appliqué au poids initial des rajustements de poids appropriés.

2) Correction pour la non-réponse

La deuxième étape du processus de pondération était le rajustement pour tenir compte de la non-réponse. Ce stade consistait à appliquer aux poids initiaux un facteur de correction pour compenser pour les effets de la non-réponse. La méthode utilisée à l'EDE est la méthode de propension à répondre (response propensity model). Cette méthode consiste à prédire la probabilité de répondre à l'enquête à partir d'un modèle de régression logistique en fonction d'un ensemble de variables indépendantes.

L'EDE disposait d'un nombre considérable de variables pour appliquer une telle méthode, soient les variables du questionnaire long du recensement. Avec cette méthode, les probabilités de réponses prédites résultant du modèle sont utilisées pour regrouper les individus en classes de repondération de taille à peu près égale de façon à ce que les individus ayant des probabilités prédites semblables soient dans la même classe. L'inverse du taux de réponse pondéré de chaque groupe est utilisé comme facteur d'ajustement pour ce groupe. Les poids initiaux des répondants sont multipliés par ce facteur d'ajustement.

Des modèles différents ont ainsi été développés successivement pour les personnes non contactées et les personnes contactées qui n'ont pas répondu. On a obtenu approximativement dix classes en gros de la même taille pour chaque modèle de régression logistique.

3) Stratification a posteriori

La troisième étape de la pondération est la stratification a posteriori, aussi appelée post-stratification. Cette méthode permet de s'assurer que la somme des poids des répondants correspond aux totaux du Recensement de 2001 pour les variables utilisées. Dans le cas de l'EDE, la post-stratification s'est faite à partir du croisement de certaines variables (région, strate, groupe d'âge et sexe).

De plus, afin d'assurer que les estimations finales, fondées sur l'échantillon, d'un certain nombre d'autres variables concordent avec leurs distributions connues, les poids ont été ajustés pour différents niveaux géographiques selon la méthode itérative du quotient. Cet ajustement a été effectué séparément pour la religion, les générations, les langues maternelles et les minorités visibles.

Les poids corrigés pour tenir compte des non réponses sont ainsi ajustés à l'aide du rapport du chiffre du recensement au chiffre de l'échantillon pour chaque strate a posteriori. Les poids obtenus après cette étape ont été utilisés pour produire toutes les estimations des différentes diffusions de l'EDE.

4) Ajustements supplémentaires des poids pour le FMGD

Puisque le FMGD est un sous-échantillon de répondants de l'EDE, une quatrième étape de pondération était nécessaire. En effet, des ajustements supplémentaires ont dû être effectués sur les poids des unités du FMGD afin de tenir compte des unités retirées. Pour ce faire, on a d'abord effectué le produit du poids obtenu à l'étape 3) précédente par le poids de sous-échantillonnage. Ensuite, un nouveau ajustement de stratification a posteriori a été effectué afin de s'ajuster aux comptes du recensement.

3.2 Lignes directrices concernant la pondération

Ainsi, le poids final associé à chaque répondant a subi de nombreux ajustements afin de représenter le mieux possible la population cible. Grâce à la pondération des données, l'échantillon de l'EDE est représentatif de la population cible, et ce, même si la fraction de sondage diffère largement d'un individu à l'autre. **L'utilisation des poids est donc essentielle pour toutes les analyses faites à partir des données de l'enquête.**

Les utilisateurs ne devraient diffuser aucune totalisation non pondérée ni faire des analyses fondées sur des résultats d'enquête non pondérés. Les taux d'échantillonnage et les taux de non-réponse varient considérablement d'une strate à l'autre et les taux de non-réponse varient en outre selon les caractéristiques démographiques. Il est donc évident que les dénombrements non pondérés d'échantillon ne peuvent être considérés comme étant représentatifs de la population visée par l'enquête.

3.3 Types d'estimation

On peut calculer deux types d'estimations « simples » d'après les données de l'EDE, à savoir des estimations qualitatives (estimations de nombres ou de proportions de personnes possédant certaines qualités ou caractéristiques) et des estimations quantitatives, c'est-à-dire des quantités ou des moyennes. La section 3.4 traite des estimations et des analyses plus complexes.

3.3.1 Estimations qualitatives

Les estimations qualitatives sont des estimations du nombre ou du pourcentage de personnes de la population visée par l'enquête qui possèdent une certaine caractéristique ou qui font partie d'une catégorie définie. Les valeurs de ces variables représentent une qualité plutôt qu'une quantité. Le nombre ou la proportion de personnes qui ont déclaré comme niveau le plus élevé de scolarité atteint un « Diplôme d'études secondaires » est un exemple d'estimations qualitatives.

Les estimations qualitatives peuvent être obtenues en additionnant les poids finaux de tous les enregistrements possédant la ou les caractéristiques qui nous intéressent. On obtient des proportions et des rapports de la forme \hat{Y}/\hat{W} en suivant les étapes suivantes :

- (i) additionnez les poids finaux des enregistrements possédant la caractéristique d'intérêt \hat{Y} ;
- (ii) additionnez les poids finaux des enregistrements possédant la caractéristique d'intérêt \hat{W} ;
- (iii) divisez le résultat obtenu en (i) par celui obtenu en (ii), soit \hat{Y}/\hat{W} .

3.3.2 Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes ou d'autres mesures de tendance centrale représentant des quantités. Le nombre de semaines ou le nombre d'heures travaillées sont des exemples d'estimations quantitatives.

Ce type d'estimation peut être obtenu en multipliant la valeur de la variable qui nous intéresse par le poids final de l'enregistrement correspondant et en additionnant cette quantité pour tous les enregistrements visés.

Pour obtenir une moyenne pondérée de la forme \hat{Y}/\hat{W} , on calcule le numérateur (\hat{Y}) de la même façon qu'une estimation quantitative et on calcule le dénominateur (\hat{W}) de la même façon qu'une estimation qualitative. Par exemple, pour estimer le nombre moyen d'heures travaillées par les répondants, procédez de la façon suivante :

- (i) estimez le nombre total d'heures travaillées par les répondants (\hat{Y}) en multipliant le nombre d'heures travaillées par chaque répondant par son poids final correspondant, puis additionnez cette valeur pour tous les répondants;
- (ii) estimez le nombre de répondants (\hat{W}) en additionnant les poids finals de tous les enregistrements correspondant à un répondant;
- (iii) divisez (i) par (ii), soit \hat{Y}/\hat{W} .

3.4 Lignes directrices pour l'analyse

Comme on l'explique en détail à la section 2.4, les répondants de l'EDE ne forment pas un échantillon aléatoire simple de la population cible. En effet, l'enquête se fonde sur un plan de sondage complexe. Par conséquent, la sélection des répondants s'est faite selon des probabilités inégales.

Les poids de sondage doivent donc être utilisés pour le calcul des estimations ou les analyses afin de tenir compte dans la mesure du possible de la sur ou sous-représentation de certains groupes dans le fichier non pondéré. L'utilisation des données provenant d'une enquête aussi complexe peut poser des difficultés aux analystes, car le choix des méthodes d'estimation et de calcul de la variance dépend du plan de sondage et des probabilités de sélection. Nombre de méthodes d'analyse intégrées aux progiciels statistiques permettent d'utiliser des poids, mais la signification et la définition de ces poids diffèrent souvent de celles applicables dans le contexte d'une enquête par sondage. Par conséquent, si les estimations calculées au moyen de ces progiciels sont souvent exactes, les variances n'ont, quant à elles, pratiquement aucune signification.

Dans le cas de nombreuses méthodes d'analyse (par exemple la régression linéaire, la régression logistique, l'estimation de taux ou de proportions et l'analyse de variance), on peut rendre l'application des progiciels courants plus significative en rééchantillonnant les poids qui figurent dans les enregistrements de façon à ce que le poids moyen soit égal à un (1). Les résultats produits par les progiciels classiques sont ainsi plus raisonnables puisque, même s'ils ne reflètent toujours pas la stratification et la mise en grappes du plan de sondage, ils tiennent compte de la sélection avec probabilités inégales. On peut effectuer ce rééchantillonnage en divisant chaque poids par le poids moyen global avant de procéder à l'analyse.

Par exemple, pour une analyse de tous les répondants noirs d'après la variable de minorité visible, la marche à suivre est la suivante :

- sélectionner tous les répondants du fichier qui vivaient en union libre (VISMIND=4);
- calculer la valeur moyenne de WGT_PUMF pour ces enregistrements;
- pour chacun de ces répondants, calculer un poids « de travail » égale à WGT_PUMF / poids moyen;
- faire l'analyse pour ces répondants en se servant du poids « de travail ».

La section 4 donne une description plus détaillée de la variabilité d'échantillonnage et de la fiabilité des données, et l'annexe –Estimation de la variance contient les règles d'obtention de la variance approximative pour estimer la variabilité d'échantillonnage d'un grand nombre d'estimations qualitatives de proportions.

4.0 Lignes directrices concernant la diffusion et fiabilité des données

Il est important que l'utilisateur se familiarise avec le contenu de la présente section avant de publier ou de diffuser autrement toute estimation calculée d'après les fichiers de microdonnées de l'EDE.

La présente section du document donne les lignes directrices que doit suivre l'utilisateur du fichier de microdonnées. Il sera ainsi capable d'obtenir des chiffres qui concordent avec ceux produits par Statistique Canada et qui sont conformes aux lignes directrices établies pour l'arrondissement et la diffusion. Les lignes directrices se répartissent en trois grands groupes, à savoir la taille minimale

d'échantillon pour le calcul d'estimations, les lignes directrices concernant la variabilité d'échantillonnage et les lignes directrices concernant l'arrondissement.

Il doit être noté que les résultats obtenus à partir du fichier de microdonnées de l'EDE peuvent être légèrement différents de ceux obtenus à partir du fichier analytique. Ces différences sont dues au sous-échantillonnage et aux autres méthodes appliquées afin d'assurer la confidentialité des données.

4.1 Taille minimale d'échantillon pour la production d'estimations

L'utilisateur doit déterminer le nombre d'enregistrements du fichier de microdonnées qui ont fourni les données entrant dans le calcul d'une estimation particulière. Si ce nombre est inférieur ou égal à 10, l'estimation pondérée ne doit généralement pas être diffusée, quelle que soit la valeur de son coefficient de variation approximatif. Si l'estimation est malgré tout diffusée, elle doit l'être avec beaucoup de prudence et le nombre insuffisant d'enregistrements sur lesquels elle est fondée doit être indiqué clairement. Veuillez noter que les données supprimées doivent être incluses dans les totaux (si les totaux contiennent plus de 10 individus).

4.2 Lignes directrices concernant la variabilité d'échantillonnage

Les estimations tirées de l'enquête sont fondées sur un échantillon de personnes. On aurait pu obtenir des chiffres différents si on avait procédé à un recensement complet sans changer le questionnaire, les intervieweurs, les superviseurs, les méthodes de traitement des données, etc. La différence entre une estimation produite d'après l'échantillon et celle tirée d'un dénombrement complet effectué dans des conditions semblables est appelée erreur d'échantillonnage de l'estimation.

Des erreurs qui ne se rapportent pas à l'échantillonnage peuvent se produire à presque toutes les étapes d'une enquête. Il se peut que les intervieweurs comprennent mal les instructions, que les répondants se trompent en répondant aux questions, que des réponses soient mal entrées dans l'ordinateur ou que des erreurs se produisent au moment du traitement et de la totalisation des données. Ces erreurs sont des exemples d'erreur non due à l'échantillonnage.

Si l'on considère un grand nombre d'observations, les erreurs aléatoires ont peu d'effets sur les estimations tirées des données de l'enquête. Toutefois les erreurs qui se produisent systématiquement les biaisent. Beaucoup de temps et d'efforts ont été consacrés à la réduction des erreurs non dues à l'échantillonnage. Des mesures d'assurance de la qualité ont été appliquées à chaque étape du cycle de collecte et de traitement des données afin de contrôler la qualité des données. On a notamment fait appel à des intervieweurs hautement qualifiés, donné à ceux-ci une formation poussée sur les méthodes d'enquête et le questionnaire, observé les intervieweurs afin de déceler les problèmes causés par la conception du questionnaire ou l'incompréhension des instructions, adopté des procédures pour veiller à réduire au minimum les erreurs de saisie des données et procédé à des contrôles du codage et des vérifications pour confirmer la logique de traitement.

4.2.1 Erreurs non dues à l'échantillonnage

L'effet de la non-réponse sur les résultats de l'enquête constitue une source importante d'erreur non due à l'échantillonnage. L'ampleur de la non-réponse varie d'une non-réponse partielle (le fait de ne pas répondre à une ou à plusieurs questions) à une non-réponse totale. Il y a non-réponse totale lorsque l'intervieweur est incapable de dépister le répondant, que le répondant ne peut fournir les renseignements voulus (peut-être à cause d'un problème de langue) ou qu'il refuse de participer à l'enquête. On traite les cas de non-réponse totale en corrigeant le poids appliqué aux personnes qui ont répondu à l'enquête afin de compenser pour ceux qui n'ont pas répondu.

Dans la plupart des cas, il y a eu non-réponse partielle à l'enquête lorsque le répondant n'a pas compris ou a mal interprété une question, ou qu'il n'a pu se souvenir du renseignement demandé. Dans le cas de l'EDE, on n'a imputé aucune réponse pour compenser les non-réponses partielles et on a attribué à la question le code de réponse « non déclaré ».

4.2.2 Erreurs d'échantillonnage

Puisque les estimations d'une enquête par sondage comportent inévitablement une erreur d'échantillonnage, de bonnes méthodes statistiques exigent que les chercheurs renseignent les utilisateurs sur l'ordre de grandeur de cette erreur d'échantillonnage.

Bien qu'on ne puisse mesurer exactement l'erreur d'échantillonnage d'une estimation telle que définie plus haut d'après des données d'échantillon uniquement, on peut estimer une mesure statistique de cette erreur, c'est-à-dire l'erreur-type, d'après ces données. À partir de l'erreur-type, on peut obtenir les intervalles de confiance des estimations (en ne tenant pas compte des effets des erreurs non dues à l'échantillonnage) sous l'hypothèse que la distribution des estimations autour de la valeur réelle de population est normale. Dans ces conditions, les chances que l'écart entre une estimation fondée sur l'échantillon et la valeur réelle de population soit inférieur à un écart-type sont de 68 sur 100, celles qu'il soit inférieur à deux écarts-types sont d'environ 95 sur 100, et il est virtuellement certain qu'il soit inférieur à trois écarts-types.

Puisque la grandeur absolue de l'erreur d'échantillonnage d'une estimation est souvent moins importante que sa grandeur relative (comparativement à l'estimation proprement dite), l'erreur-type n'est pas toujours la meilleure mesure de l'erreur d'échantillonnage. Par exemple, une erreur-type de 10 pour une estimation de 20 indiquerait généralement que la qualité de l'estimation est médiocre, tandis que la même erreur-type pour une estimation de 1 000 indiquerait généralement que l'estimation est bonne. Par conséquent, on exprime souvent la grandeur de l'erreur d'échantillonnage relativement à la taille de l'estimation, sous forme de coefficient de variation (c.v.). On obtient le coefficient de variation d'une estimation en divisant l'erreur-type de l'estimation par l'estimation proprement dite et en exprimant la fraction résultante en pourcentage. Dans l'exemple qui précède, le c.v. de la première estimation est de 50 % (10/20), tandis que celui de la deuxième est de 1 % (10/1 000).

Lignes directrices pour la diffusion des estimations

Avant de diffuser et (ou) de publier des estimations tirées du fichier de microdonnées, l'utilisateur devrait calculer le coefficient de variation associé à l'estimation, consulter le tableau ci-dessous et se conformer aux lignes directrices correspondant à la valeur du coefficient de variation de l'estimation.

Pour plus d'information concernant la méthode d'auto-amorçage (bootstrap) et comment calculer des coefficients de variation à partir des données de l'EDE, veuillez consulter l'Annexe – Estimation de la variance.

Coefficient de variation - Lignes directrices pour la diffusion

Niveau de qualité de l'estimateur	Symbole à utiliser	Condition	Lignes directrices
Acceptable	Ne requiert aucun symbole	$0,0 \% \leq \text{c.v.} \leq 16,5\%$	L'estimation peut être utilisée sans restriction
Médiocre	E	$16,6\% \leq \text{c.v.} \leq 33,3\%$	L'estimation doit être utilisée avec prudence car un niveau d'erreur élevé y est associé
Inacceptable	F	$\text{c.v.} \geq 33,4 \%$	Cette information ne devrait pas être diffusée. Cependant, si l'utilisateur choisit de le faire, il doit diffuser l'information avec la mise en garde suivante : « Nous informons l'utilisateur que <précisez la donnée> ne répond pas aux normes de qualité de Statistique Canada. Les conclusions tirées de cette donnée ne sauraient être fiables. »

Nota : Les lignes directrices concernant la variabilité d'échantillonnage devraient être appliquées aux estimations arrondies. Pour de plus amples renseignements, consultez la publication intitulée *Lignes directrices sur le niveau de qualité de Statistique Canada*, (No 12-539-XIF dans le catalogue de Statistique Canada).

4.3 Arrondissement

Afin que les estimations produites d'après les fichiers de microdonnées de l'EDE correspondent à celles produites par Statistique Canada, il est vivement conseillé à l'utilisateur de se conformer aux lignes directrices qui suivent concernant leur arrondissement. La diffusion d'estimations non arrondies risque d'induire en erreur, car ces estimations pourraient donner l'impression d'être plus précises qu'elles ne le sont en réalité.

4.3.1 Lignes directrices concernant l'arrondissement

- 1) Les estimations de totaux qui figurent dans le corps d'un tableau statistique doivent être arrondies à la dizaine près par la méthode d'arrondissement classique (voir la définition à la section 4.3.2).
- 2) Les totaux partiels et généraux de marge des tableaux statistiques doivent être calculés à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la dizaine près par la méthode d'arrondissement classique.
- 3) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir d'éléments non arrondis, puis arrondis à une décimale par la méthode d'arrondissement classique.
- 4) Les sommes et les différences d'agrégats ou de rapports doivent être calculées à partir de leurs éléments correspondants non arrondis, puis arrondies à leur tour à la dizaine près ou à la décimale près selon la méthode d'arrondissement classique.

5) Si, en raison de contraintes d'ordre technique ou autre, on applique une autre méthode que l'arrondissement classique, si bien que les estimations obtenues différeront des estimations correspondantes produites par Statistique Canada, il est vivement conseillé à l'utilisateur d'indiquer la raison de ces divergences dans le document diffusé.

4.3.2 Méthode d'arrondissement classique

Selon la méthode d'arrondissement classique, si le premier ou le seul chiffre à supprimer est compris entre 0 et 4, le dernier chiffre retenu ne change pas. Si le premier ou le seul chiffre à supprimer est compris entre 5 et 9, on augmente d'une unité (1) la valeur du dernier chiffre retenu.

Par exemple, le nombre 8 499 arrondi au millier près serait 8 000 et le nombre 8 500 arrondi au millier près serait 9 000.

5.0 Autre produits à partir des données de l'EDE

De l'information additionnelle sur l'Enquête sur la Diversité Ethnique peut être obtenue sur le site web www.statcan.ca de Statistique Canada.

Plus précisément, de l'information générale sur l'enquête (tel que ce guide de l'utilisateur) est disponible à partir du lien :

http://www.statcan.ca/cgi-bin/imdb/p2SV_f.pl?Function=getSurvey&SDDS=4508&lang=en&db=IMDB&dbq=f&adm=8&dis=2

Certains résultats de l'enquête déjà publiés peuvent être accédés à des liens suivants :

<http://dissemination.statcan.ca/Daily/Francais/030929/q030929a.htm>

<http://www.statcan.ca:8096/bsolc/francais/bsolc?catno=89-593-X&CHROPG=1>

Les utilisateurs peut également s'ils le désirent appliquer pour avoir accès au fichier analytique de l'EDE qui est un fichier de microdonnées avec beaucoup plus de détails que le FMGD de l'EDE. Le fichier analytique de l'EDE contient tout le contenu de l'enquête (données brutes et variables dérivées), incluant des identificateurs détaillés de la géographie et certaines informations provenant du Recensement 2001 pour les répondants de l'EDE. L'accès à ce fichier est disponible seulement à partir des Centres de Données de Recherche (CDR) de Statistique Canada et sont situés dans plusieurs universités canadiennes (plus de détails sont disponibles sur le site web

http://www.statcan.ca/francais/rdc/index_f.htm). Pour accéder au fichier analytique de l'EDE, les chercheurs doivent soumettre une proposition de projet à un comité d'examen relevant du Conseil de recherches en sciences humaines (CRSH) et de Statistique Canada en utilisant le formulaire suivant : http://www.sshrc.ca/web/apply/program_descriptions/ciss_reseach_data_f.asp

Des tableaux personnalisés à partir des données de l'EDE sont disponibles à Statistique Canada à un prix qui reflète les ressources utilisées pour les produire. Pour en faire l'achat ou pour plus d'information sur le FMGD de l'EDE ou sur d'autres produits, veuillez contacter :

Services à la clientèle
Division de la statistique sociale et autochtone
Statistique Canada
Immeuble Jean Talon, 7e étage
Parc Tunney
Ottawa (Ontario)
K1A 0T6

Téléphone : (613) 951 5979
Télécopieur : (613) 951 0387
Courriel : sasd-dssea@statcan.ca