

Enquête sociale générale

Cycle 30 : Les Canadiens au travail et à la maison fichier de microdonnées à grande diffusion, documentation et guide de l'utilisateur

N° 12M0030X au catalogue



Juin 2018
Also available in English

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 (courriel : statcan.sasdcclientservices-dsseaserviceaclientele.statcan@canada.ca).

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

Centre de contact national de Statistique Canada

Numéros sans frais (Canada et États-Unis) :

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

Appels locaux ou internationaux :

Service de renseignements	1-514-283-8300
Télécopieur	1-613-951-0581

Programme des services de dépôt

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

Renseignements pour accéder au produit ou le commander

Ce produit N° 12M0030X au catalogue est aussi disponible en version imprimée.

Les frais de livraison supplémentaires suivants s'appliquent aux envois à l'extérieur du Canada :

La version imprimée peut être commandée par :

- Téléphone (Canada et États-Unis) 1-800-267-6677
- Télécopieur (Canada et États-Unis) 1-877-287-4369
- Courriel STATCAN.infostats-infostats.STATCAN@canada.ca
- Poste
Statistique Canada
Immeuble R.-H.-Coats, 6^e étage
150, promenade Tunney's Pasture
Ottawa (Ontario) K1A 0T6

- En personne auprès des agents et librairies autorisés.

Lorsque vous signalez un changement d'adresse, veuillez nous fournir l'ancienne et la nouvelle adresse.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui sont 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 sur le site www.statcan.ca sous À propos de nous > Offrir des services aux Canadiens.

Enquête sociale générale

Cycle 30: Les Canadiens au travail et à la maison

Fichier de microdonnées à grande diffusion,
documentation et guide de l'utilisateur

Par Anna Kemeny

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2018

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

Juin 2018

N° 12M0030X au catalogue

Périodicité : hors série

Ottawa

This publication is available in English upon request (catalogue no. 12M0030X)

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.

ESG de 2016 : Les Canadiens au travail et à la maison

Fichier de microdonnées à grande diffusion (FMGD) —Guide de l'utilisateur

Table of Contents

1.	Introduction	3
2.	Objectifs de l'enquête sociale générale.....	3
3.	Contenu de l'ESG de 2016.....	3
3.1	Questions générales.....	4
3.2	La vie au travail	4
3.3	La vie à la maison	4
3.4	Équilibre entre le travail et la vie personnelle	5
3.5	Santé, bien-être et résilience	5
3.6	Variables sociodémographiques	5
4.	Nouveau contenu et changements apportés au contenu existant précédemment.....	5
4.1	Sommaire des principaux changements	6
4.2	Comparabilité des estimations	8
5.	Plan d'enquête et d'échantillonnage	8
5.1	Population cible	8
5.2	Stratification.....	8
5.3	Base de sondage	9
5.4	Stratégie d'échantillonnage	9
5.5	Taille de l'échantillon et répartition.....	10
6.	Collecte et taux de réponse	10
6.1	Collecte	10
6.2	Taux de réponse.....	11
7.	Traitement des données	11

7.1	Saisie des données	11
7.2	Codage	12
7.3	Vérification et imputation.....	12
7.4	Création de variables combinées et dérivées	12
8.	Estimation	13
8.1	Pondération des personnes	13
8.2	Politique de pondération	17
8.3	Types d'estimations	17
8.4	Lignes directrices pour l'analyse	18
8.5	Estimation du nombre de personnes à l'aide de WGHT_PER au fichier principal.....	19
9.	Lignes directrices concernant la diffusion et fiabilité des données.....	19
9.1	Taille minimale d'échantillon pour le calcul d'estimations.....	20
9.2	Lignes directrices concernant la variabilité d'échantillonnage.....	20
9.3	Estimation de la variance au moyen des poids bootstrap	22
9.4	Arrondissement.....	22
10.	Autres renseignements	23
	Annexe A – Codes de citoyenneté	24
	Annexe B – Codes de pays de naissance.....	29
	Annexe C – Codes de religion	36
	Annexe D – Codes de sport.....	39
	Annexe E – Codes de principal domaine d'études.....	41
	Annexe F – Conseils sur l'utilisation des poids bootstrap standards de l'ESG.....	42

Introduction

Le présent guide a été préparé à l'intention des utilisateurs du fichier de microdonnées à grande diffusion (FMGD) de l'Enquête sociale générale (ESG) de 2016, *Les Canadiens au travail et à la maison*. Ses objectifs sont de fournir des renseignements contextuels et généraux, de familiariser les utilisateurs avec le contenu de l'enquête et de décrire les procédures et les concepts liés à la qualité des données, à l'estimation, à la collecte, au traitement et à la méthodologie.

L'ESG de 2016, réalisée du 2 août au 23 décembre 2016, est une enquête par sondage dont la conception est transversale. La population visée par l'enquête comprend toute personne âgée de 15 ans et plus ne vivant pas en institution, qui réside dans l'une des 10 provinces canadiennes. L'enquête utilise une nouvelle base de sondage, créée en 2013, qui combine des numéros de téléphone (lignes terrestres et téléphones cellulaires) avec le Registre des adresses de Statistique Canada, et recueille des données par voie électronique et par téléphone. Les données sont sujettes à des erreurs d'échantillonnage et à des erreurs non dues à l'échantillonnage.

Les renseignements figurant dans les sections suivantes devraient être utilisés afin de bien comprendre les concepts fondamentaux qui définissent les données fournies dans le FMGD du cycle 30 de l'ESG, la méthodologie sous-jacente de l'enquête et les principaux aspects de la qualité des données. Ces renseignements permettront de mieux comprendre les points forts et les limites des données de même que la façon de les utiliser et de les analyser efficacement. Les renseignements peuvent être particulièrement utiles lors de comparaisons avec des données provenant d'autres enquêtes ou d'autres sources d'information et lors de la formulation de conclusions concernant les changements au fil du temps ou les différences entre sous-groupes de la population cible.

1. Objectifs de l'enquête sociale générale

Le programme de l'ESG, qui a débuté en 1985, procède à des enquêtes téléphoniques dans les 10 provinces. L'ESG est reconnue comme étant un outil de collecte de données transversales permettant d'analyser les tendances et offrant la capacité d'élaborer et de mettre à l'essai de nouveaux concepts s'appliquant aux questions d'actualité.

Les deux principaux objectifs de l'Enquête sociale générale sont les suivants :

- a) rassembler des données sur les tendances sociales, de manière à suivre l'évolution des modes de vie et du bien-être des Canadiens, et
- b) fournir des renseignements sur des questions de politique sociale précises qui suscitent déjà ou qui susciteront de l'intérêt.

Afin de réaliser les objectifs susmentionnés, l'ESG comprend la collecte de données sur des variables entrant dans deux catégories, à savoir les variables de classification et la thématique principale. Les variables de classification permettent de définir des groupes démographiques qui servent à analyser les données de la thématique principale. L'âge, le sexe, le niveau de scolarité et le revenu sont des exemples de variables de classification. Le contenu de la thématique principale sert à mesurer l'évolution de la société sur le plan des conditions de vie et du bien-être ainsi qu'à fournir des données destinées à éclairer des débats bien précis de politique publique.

2. Contenu de l'ESG de 2016

Le profil démographique du Canada, qui évolue rapidement, ainsi que les enjeux sociaux et économiques connexes ont entraîné beaucoup de discussion sur les relations entre le travail, le style de vie et le bien-être. L'évaluation de la qualité de la vie au travail peut contribuer à mieux cerner les enjeux relatifs à la productivité, au moral, à l'efficacité et à l'équité. La détermination des tendances relatives aux activités

de loisir et à celles menées à la maison permet de prendre le pouls de la culture canadienne. La combinaison de ces deux aspects fournira des éclaircissements sur la santé et le bien-être des Canadiens par rapport aux défis de la vie quotidienne d'aujourd'hui et de demain.

Le nouveau cycle du programme de l'Enquête sociale générale, *Les Canadiens au travail et à la maison*, jette un regard exhaustif sur la façon dont les Canadiens vivent en intégrant les domaines du travail, de la maison, des loisirs et du bien-être général en une seule enquête. Les utilisateurs des données ont indiqué qu'ils souhaitaient vraiment en savoir plus sur les habitudes de vie des Canadiens qui peuvent avoir une incidence sur leur santé et leur bien-être, tant dans le milieu de travail qu'à la maison. Le point fort de cette enquête est sa capacité de prendre les renseignements divers que fournissent les Canadiens sur diverses facettes de la vie et de les combiner d'une manière qui n'était pas possible antérieurement avec des enquêtes qui ne couvraient qu'un seul sujet principal.

3.1 Questions générales

Après que la liste des membres du ménage a été dressée et qu'un répondant a été sélectionné (à l'aide de l'outil automatisé de l'application), la date de naissance du répondant et son état matrimonial sont consignés. Cette opération est suivie par l'établissement du lien des membres du ménage avec le répondant, du lieu de naissance du répondant et de son statut d'immigrant. De plus, le répondant doit répondre à un ensemble de questions sur son activité principale, son historique de travail et diverses caractéristiques d'emploi.

3.2 La vie au travail

Cette partie de l'enquête permet de dresser le portrait des sentiments des Canadiens à l'égard de leur milieu de travail à l'aide de questions sur leurs rapports avec leur travail, leur équipe, leurs collègues, superviseurs et subordonnés. En outre, l'enquête explore des sujets tels que l'éthique de travail, l'intensité et la répartition du travail, la rémunération et les avantages sociaux, la satisfaction au travail et la signification du travail, les relations interculturelles en milieu de travail, et la discrimination et la compétition en milieu de travail. En plus du fait que ces variables sont intéressantes, nombre d'entre elles sont corrélés avec la productivité, la motivation, l'absentéisme et la ponctualité, les accidents, la santé mentale ou physique et la satisfaction à l'égard de la vie en général, ce qui les rend essentielles lors de l'examen des enjeux de niveau micro et macro des personnes et de l'économie. En plus de couvrir les Canadiens ayant un emploi rémunéré, certains sous-groupes, tels que les travailleurs autonomes, les personnes cherchant un emploi, les retraités et les étudiants répondent à une série distincte de questions se rapportant à leurs situations particulières.

3.3 La vie à la maison

Afin d'explorer la diversité des familles et des modalités de vie d'aujourd'hui, des questions portent sur la satisfaction à l'égard de la quantité et de la qualité du temps passé ensemble en famille et la répartition des tâches dans le ménage. Parmi les autres sujets à la rubrique « La vie à la maison », mentionnons la pratique d'activités en plein air et de sports, la participation à des activités culturelles, l'utilisation des technologies et la sensibilisation à l'alimentation. En vue de lutter contre un mode de vie de plus en plus sédentaire, les taux croissants d'obésité et les problèmes de santé qui en découlent, l'intérêt pour l'activité physique sous toutes ses formes s'intensifie. Les réponses aux questions portant sur les activités en plein air et les sports informeront à quel point les Canadiens sont actifs aujourd'hui. Dans le domaine de la culture, nous remarquons que les technologies jouent un rôle de transformation de plus en plus grand en influant sur la façon dont nous consommons, produisons et distribuons les produits culturels. L'enquête explore les modes traditionnels de participation et l'expression culturelle au moyen d'appareils technologiques.

3.4 Équilibre entre le travail et la vie personnelle

Des éléments fondamentaux de longue date du programme de l'ESG, le temps et le stress constituent d'importants piliers d'organisation des différents cycles de l'enquête. Des questions figurant dans le questionnaire de l'ESG de 2016 sondent la fréquence à laquelle le travail rémunéré empiète sur le temps passé en famille et les responsabilités familiales interfèrent avec le travail rémunéré, ainsi que les niveaux de satisfaction à l'égard de l'équilibre entre le travail rémunéré et la vie à la maison. Les questions sur la quantité et la source de stress que les Canadiens connaissent dans leur vie quotidienne sont étroitement liées à l'équilibre entre le travail et la vie personnelle, quoiqu'elles ne s'y rapportent pas exclusivement.

3.5 Santé, bien-être et résilience

L'état de santé autodéclaré et le bien-être subjectif sont des questions posées depuis longtemps dans l'ESG et ont été utilisées au fil des années pour évaluer la santé et le bien-être des répondants. Bien que l'état de santé autodéclaré soit devenu un élément essentiel dans les enquêtes sociales et qu'il suive très bien d'autres mesures objectives de santé, le bien-être subjectif sert traditionnellement à mesurer la satisfaction générale au regard de la vie. L'ESG de 2016 les complète par un certain nombre de questions qui permettront de dresser un portrait plus détaillé de la façon dont les Canadiens évaluent leur vie. Un ensemble de questions sur le degré de satisfaction donne une répartition détaillée de la satisfaction dans différentes sphères de la vie, telles que le niveau de vie, la santé et les réalisations, pour n'en nommer que quelques-unes. Un module sur les possibilités de vie porte sur les différences en ce qui concerne les possibilités générationnelles et vise à mettre en lumière les préoccupations quant à la mobilité sociale, à la cohésion sociale et à la société inclusive. Les questions sur les perspectives visent à savoir si les personnes estiment que leur vie en général et leur vie financière seront plus faciles ou plus difficiles dans l'avenir. Enfin, un ensemble de questions sur la résilience touche la capacité des personnes à rebondir après des crises ou l'adversité, ce qui correspond à la définition la plus courante de la résilience.

3.6 Variables sociodémographiques

Cette section offre une variété de mesures sociodémographiques dont beaucoup sont répétées chaque année dans l'ESG, concernant le répondant, leur conjoint ou partenaire et leurs parents afin de soutenir l'analyse des familles canadiennes ainsi que celle des individus. La section renferme également des questions sur le lieu de naissance, l'identité autochtone et l'appartenance à une minorité visible, la religion, la langue, l'orientation sexuelle ainsi que sur le revenu personnel et celui du ménage.

Les questions sur le revenu personnel du répondant ainsi que celui du ménage ont été retirées du questionnaire et ces données sont maintenant le résultat d'un couplage d'enregistrements entre les fichiers fiscaux des répondants et l'ESG.

3. Nouveau contenu et changements apportés au contenu existant précédemment

L'ESG de 2016, *Les Canadiens au travail et à la maison*, est le premier cycle sur ce sujet. Bien que les renseignements de certains modules aient été recueillis dans des enquêtes précédentes, la plupart ne l'ont pas été. Cette section exposera tous les changements, non seulement ceux apportés aux questions qui existaient auparavant, mais aussi ceux apportés aux processus liés à l'élaboration et à la méthodologie de l'enquête ainsi qu'à la collecte et au traitement des données.

4.1 Sommaire des principaux changements

1) Contenu de base

Tout le nouveau contenu et les révisions de certaines des questions existantes ont été mis à l'essai par le Centre de ressources en conception de questionnaires (CRCQ) de Statistique Canada dans le cadre d'un processus à deux phases. À l'automne 2014, le contenu a été mis à l'essai à l'aide d'un questionnaire papier. Ensuite, au printemps 2015, l'application électronique a été mise à l'essai par les répondants pour vérifier sa convivialité. De février à la mi-mars 2016, un essai pilote à grande échelle a été mené pour le questionnaire électronique auto-administré (QE) et l'interview téléphonique assistée par ordinateur (ITAO).

2) Lien avec le répondant sélectionné

En 2015, la matrice de composition du ménage, qui était auparavant remplie dans le module d'entrée, a été remplacée au complet par le module « Lien avec le répondant sélectionné » (LRS), afin de réduire le temps d'interview. Cette liste unidirectionnelle a été utilisée dans l'ESG de 2016.

3) Classification sociodémographique

En outre, en 2014, de nombreuses questions sociodémographiques propres à l'enquête ont été remplacées par le contenu harmonisé de Statistique Canada (c.-à-d. des modules de questionnaire uniformisés pour les variables d'enquête auprès des ménages, comme l'éducation, la situation vis-à-vis de l'activité, l'identité autochtone, le lieu de naissance et la citoyenneté, la santé auto évaluée et la religion). Les modules de contenu harmonisé comprennent des concepts, des définitions, une classification et un libellé standard pour plusieurs modes de collecte. Ce nouveau contenu uniformisé s'apparente pour une large part aux concepts utilisés précédemment par l'ESG sur l'Emploi du temps, mais dans certains cas, les variables dérivées traditionnelles ont dû être rajustées.

4) Revenu

En 2015, les questions sur le revenu personnel et le revenu du ménage n'ont pas été demandées dans le cadre de l'enquête. Les données sur le revenu ont plutôt été obtenues grâce au couplage des données fiscales administratives pour les répondants qui ne s'objectaient pas à ce couplage. Les répondants ont été informés du couplage prévu, avant et pendant l'enquête. Tous les répondants qui s'objectaient au couplage de leurs données ont fait consigner leurs objections, et aucun couplage de leurs données fiscales administratives n'a eu lieu. Ces cas ont été imputés à l'aide de donneurs d'imputation. Le couplage avec les données fiscales administratives réduit le fardeau pour les répondants et augmente la qualité des données, tant du point de vue de l'exactitude que des taux de réponses. En outre, l'ESG de 2016 est le premier cycle de l'ESG à utiliser le revenu famille (obtenu grâce au couplage des données fiscales administratives) plutôt que le revenu ménage utilisé par le passé. (Voir la section 7.3 pour plus de détails.)

5) Base de sondage

L'ESG de 2016 sur les Canadiens au travail et à la maison a été mise en œuvre au moyen de la base remaniée de l'ESG, qui intègre des données de sources de numéros de téléphone (fixes et cellulaires) disponibles à Statistique Canada et du Registre des adresses (RA). Cette nouvelle base de sondage comprend les ménages « dotés d'un téléphone cellulaire seulement », une population en croissance qui n'est pas couverte par la CA. L'unité d'échantillonnage est aussi différente et est maintenant définie comme des regroupements de numéros de téléphone. Voir la section 5.3 pour plus de détails.

6) Codage

On a utilisé le Système de classification des industries de l'Amérique du Nord (SCIAN) de 2012 et la Classification nationale des professions (CNP) de 2016 pour le codage des industries et des professions.

7) Traitement

La plupart des étapes courantes de traitement des données sont standard, y compris les contrôles de cohérence et de famille. Deux aspects du traitement sont cependant toujours relativement nouveaux et justifient une description plus détaillée.

L'un d'eux est les outils communs, utilisés dans l'ensemble du Secteur de la statistique sociale, de la santé et du travail à Statistique Canada. Du début de l'élaboration du questionnaire jusqu'au traitement et à la diffusion, ces nouveaux outils communs sont conçus pour simplifier les spécifications du questionnaire et les étapes de traitement en vue d'améliorer l'efficacité, la cohérence et l'uniformité de toutes les enquêtes. La majorité des nouvelles procédures sont invisibles pour les utilisateurs, sauf celles qui se rapportent au dictionnaire de données. Toutes les enquêtes traitées à l'aide des outils communs ont des noms de variable composés de huit caractères ou moins et les codes de réserve suivants :

- 6 – Saut valide;
- 7 – Ne sait pas;
- 8 – Refus;
- 9 – Non déclaré.

En plus des procédures d'outils communs courantes, certaines solutions de rechange ont été créées pour l'ESG de 2016 afin de tenir compte du fait que les outils communs utilisés à ce moment-là n'étaient pas en mesure de traiter les données des QE. La majorité des solutions de rechange ont été codées manuellement en SAS et ont été exécutées séparément. Au cours du dernier stade de traitement, les données du QE et des ITAO ont été fusionnées en un seul ensemble de données.

Il est très important de noter qu'il n'existe aucune façon de séparer les catégories « Ne sait pas », « Refus » et « Non déclaré » dans les données du QE. Par conséquent, une fois que les données sont fusionnées (ce qui est le cas pour le fichier principal et le fichier de microdonnées à grande diffusion (FMGD)), les catégories « Ne sait pas », « Refus » et « Non déclaré » ne peuvent pas être considérées comme distinctes. Elles doivent être fusionnées et traitées comme s'il s'agissait de la catégorie « Non déclaré ». En revanche, la catégorie « Saut valide » a sa propre raison d'être puisqu'elle renvoie aux répondants à qui une question donnée ne leur a pas été posée en raison de leurs caractéristiques.

Couplage des données fiscales : Le couplage avec les dossiers fiscaux administratifs, a été couronné de succès, et des données fiscales administratives étaient disponibles pour 91,9% des répondants à l'ESG de 2016 qui ne se sont pas objectés au couplage.

8) Variables dérivées

Dans des cycles passés, certaines variables dérivées ont été créées à partir d'une seule variable à l'enquête. La variable était renommée dans le FMGD. Au cycle 29, nous avons mis fin à cette pratique ce qui signifie que certaines variables dérivées présentes dans les cycles passés sont simplement représentées par la variable à l'enquête. À titre d'exemple, en 2010 la variable DWELLING était dérivée de la variable DOR_Q110. En 2015, DWELLING a été retirée et seul l'acronyme et le numéro de la question à l'enquête ont été gardés. Ce changement ne diminue pas l'apport de la base de données quant à l'information disponible à l'utilisateur.

En plus des variables dérivées standard, sept autres variables dérivées ont été créées pour ce cycle :

1. « WORKSIZE » (détermine si le lieu de travail du répondant est une petite, une moyenne ou une grande entreprise);

2. « IMMSTAT » (détermine si le répondant est un immigrant);
3. « RECIMM » (détermine si le répondant est un immigrant récent, c.-à-d. s'il est arrivé au Canada au cours des dix dernières années);
4. « CUREMPLO » (détermine si le répondant est actuellement employé en tant que travailleur rémunéré ou travailleur autonome);
5. « RETIRED » (détermine si le répondant est retraité);
6. « RETYEARS » (détermine le nombre d'années depuis la retraite la plus récente);
7. « BULLIED » (détermine si le répondant a déjà été victime d'intimidation (de discrimination, de traitement injuste) en milieu de travail.

9) Poids

En 2013, l'utilisation d'une nouvelle base d'échantillonnage et d'une nouvelle définition pour notre unité d'échantillonnage a entraîné une nouvelle stratégie de pondération pour l'ESG (voir la section 8.1). Par ailleurs, on a modifié les poids bootstrap pour passer de poids bootstrap moyens à des **poids bootstrap standards** (voir l'annexe F pour plus de renseignements sur la façon d'utiliser les poids bootstrap standards).

4.2 Comparabilité des estimations

Même si la plus grande partie du contenu de cette enquête est toute nouvelle, il y a certains modules qui ont figuré dans des cycles précédents de l'ESG (par exemple les sports, la culture, certaines variables liées au travail, la question sur la satisfaction à l'égard de la vie, le stress dans la vie quotidienne et certains autres). Pour ceux-ci, il est important de signaler que tout changement significatif à la méthodologie d'enquête (comme il a été indiqué précédemment) peut avoir des répercussions sur la comparabilité des données au fil du temps. Il est impossible de déterminer avec certitude si, et dans quelle mesure, les différences dans une variable sont attribuables à un changement réel survenu au sein de la population ou à la modification de la méthodologie d'enquête.

Par conséquent, à chaque étape du traitement, de la vérification et de la diffusion, des efforts considérables ont été déployés pour produire les données les plus précises possible en ce qui a trait à leur niveau de détail et pour veiller à ce que les estimations publiées soient de bonne qualité, en respectant les normes de Statistique Canada.

4. Plan d'enquête et d'échantillonnage

Les données de l'Enquête sociale générale (ESG) de 2016, *Les Canadiens au travail et à la maison*, ont été recueillies du 2 août au 23 décembre 2016. Veuillez consulter les sections suivantes pour obtenir des descriptions de la population cible, de la stratification, de la base de sondage, de la stratégie d'échantillonnage, de la taille de l'échantillon et de la répartition de l'échantillon.

4.1 Population cible

La population visée par l'ESG de 2016 incluait toutes les personnes de 15 ans et plus au Canada, sauf :

1. les habitants du Yukon, des Territoires du Nord-Ouest et du Nunavut;
2. les résidents à plein temps des établissements institutionnels.

4.2 Stratification

À des fins d'échantillonnage, chacune des dix provinces a été divisée en strates (c.-à-d. en régions géographiques). Nombre de régions métropolitaines de recensement¹ (RMR) ont été considérées individuellement comme des strates distinctes. C'est le cas pour St. John's, Halifax, Saint John, Montréal, Québec, Toronto, Ottawa, Hamilton, Winnipeg, Regina, Saskatoon, Calgary, Edmonton et Vancouver.

Toutes les RMR qui ne figurent pas dans cette liste sont situées au Québec, en Ontario et en Colombie-Britannique, sauf pour Moncton. Trois autres strates ont été constituées en regroupant les RMR qui restent (sauf Moncton) au Québec, en Ontario et en Colombie-Britannique. Enfin, on a regroupé les régions autres que des RMR dans chacune des dix provinces pour former dix autres strates, pour un total de 27 strates. Moncton a été ajouté à la strate de régions autres qu'une RMR du Nouveau-Brunswick.

4.3 Base de sondage

La base de sondage a été créée à partir de deux composantes différentes :

- Liste des numéros de téléphone utilisés (fixes et cellulaires) dont disposait Statistique Canada à partir de diverses sources (compagnies de téléphone, recensement de la population, etc.) ;
- Registre des adresses (RA) : liste de tous les logements dans les dix provinces.

Le Registre des adresses (RA) a servi à regrouper tous les numéros de téléphone associés à la même adresse valide. Environ 86 % des numéros de téléphone disponibles ont été reliés au RA. Les enregistrements découlant de cet appariement pouvaient comporter plus d'un numéro de téléphone (regroupés selon l'adresse). Les autres numéros de téléphone (14%) non couplés au RA ont aussi été inclus dans la base de sondage². La combinaison de ces deux composantes a pour résultat la base de sondage. La justification du recours à tous les numéros de téléphone (couplés et non couplés) visait à assurer une bonne couverture de tous les ménages ayant un numéro de téléphone.

Lorsque plus d'un numéro de téléphone était lié à un enregistrement, ces numéros ont été triés selon la source et le type de numéro de téléphone. Le premier numéro de téléphone a été considéré comme le meilleur numéro de téléphone disponible pour joindre le ménage.

À noter que pour les sections restantes du document, le mot « enregistrement » se rapporte au regroupement de numéros de téléphone qui constituent notre unité d'échantillonnage dans la base de sondage.

4.4 Stratégie d'échantillonnage

Dans chaque base de sondage, chaque enregistrement a été attribué à une strate de la province à laquelle il appartenait. Un échantillon aléatoire simple sans remise a été sélectionné par la suite dans chaque strate.

La base de sondage a été créée en utilisant plusieurs sources couplées, tel que le recensement, des données administratives et des fichiers de facturation. La couverture a été améliorée (même si une sur-représentation ou une sous-représentation peut toujours exister) comparativement à la composition aléatoire utilisée par le passé. Tous les répondants des dix provinces ont été contactés par téléphone et interviewé par téléphone ou en complétant un questionnaire électronique. Les ménages sans téléphones ont par conséquent été exclus de la population cible. Les estimations d'enquête ont été ajustées (pondérées) pour représenter toutes les personnes de la population cible incluant celles non couvertes par la base de sondage.

¹ Basé sur la géographie du Recensement 2011.

² Environ 9% de ces numéros ont été regroupés en utilisant les adresses disponibles sur les sources administratives. Chacun des numéros restants constitue un enregistrement distinct sur la base.

Pour l'ESG sur l'Emploi du temps de 2015, 89,8% des numéros de téléphone composés ont permis de joindre des ménages admissibles. Un ménage était admissible s'il comptait au moins une personne de 15 ans et plus. Au cours de la collecte, pour les ménages ne rencontrant pas les critères d'éligibilité, les interviews étaient terminées après une série de questions permettant d'établir si le ménage rencontrait ou non ces critères.

Un répondant de chaque ménage a ensuite été sélectionné aléatoirement pour qu'il remplisse le questionnaire électronique ou qu'il réponde à une interview téléphonique.

4.5 Taille de l'échantillon et répartition

La taille de l'échantillon cible (c.-à-d. le nombre de répondants ciblé) pour le cycle 30 sur les Canadiens au travail et à la maison était de 20 000 tandis que le nombre réel de répondants était de 19 609. Pour chaque province, des tailles minimales d'échantillon ont été déterminées, afin de veiller à ce que certaines estimations aient une variabilité d'échantillonnage acceptable au niveau de la strate. Lorsque la taille de l'échantillon des strates ciblée a été atteinte, le reste de l'échantillon a été réparti entre les strates de manière à équilibrer les exigences de précision pour les estimations au niveau national et au niveau des strates.

5. Collecte et taux de réponse

5.1 Collecte

Les données de l'ESG de 2016 ont été recueillies par voie électronique au moyen d'un questionnaire autoadministré (QE) et d'interviews téléphoniques assistées par ordinateur (ITAO). Les répondants ont été interviewés dans la langue officielle de leur choix. L'interview par personne interposée n'était pas permise.

Toutes les interviews ont été faites au moyen d'installations téléphoniques regroupées à cette fin dans cinq bureaux régionaux de Statistique Canada, et les appels ont été effectués entre 9 h et 21 h 30 environ, du lundi au vendredi inclusivement. Les interviews ont été aussi menées de 10 h à 17 h les samedis et de 13 h à 21 h, les dimanches. Les cinq bureaux régionaux étaient : Halifax, Sherbrooke, Sturgeon Falls, Winnipeg et Edmonton. Les intervieweurs ont reçu de la formation du personnel de Statistique Canada concernant les techniques d'enquêtes téléphoniques utilisant l'ITAO, de même que les concepts et procédures d'enquête. La majorité des intervieweurs avaient travaillé aux cycles précédents de l'ESG.

On a demandé aux intervieweurs de faire toute tentative raisonnable pour obtenir une interview complète avec le membre du ménage sélectionné de façon aléatoire. On a communiqué de nouveau à au plus deux reprises avec les personnes qui avaient d'abord refusé de participer, afin de leur expliquer l'importance de l'enquête et de les inciter à y participer. Dans les cas où l'intervieweur avait téléphoné à un mauvais moment, un rendez-vous a été pris en vue de rappeler à un moment plus opportun. Dans les cas où il n'y avait personne à la maison, on a rappelé à de nombreuses reprises.

Les manuels d'enquête ne sont pas inclus dans la présente trousse de documentation, mais ils peuvent être obtenus en communiquant avec Statistique Canada (voir la section 11).

Les données de l'ESG de 2016 ont été recueillies du 2 août au 23 décembre 2016. Le questionnaire électronique autoadministré (QE) a été offert jusqu'au 30 septembre, moment auquel le portail du QE a été fermé et tous les cas incomplets ont été transférés pour être traités par ITAO. L'échantillon était composé d'une seule vague.

Les étapes suivantes résument le processus de collecte et le mouvement des cas du QE à l'ITAO. Une fois dans l'ITAO, les cas ne pouvaient pas être transférés de nouveau au QE.

1. Des lettres de présentation papier ont été envoyées aux ménages dont le numéro de téléphone avait pu être apparié à une adresse. Cette situation correspondait à environ 80 % des cas. Dans la lettre figurait un code d'accès sécurisé qu'un membre du ménage âgé de 15 ans ou plus devait utiliser pour ouvrir une session en vue de répondre à l'enquête.
2. Une fois dans l'application de l'enquête, ce membre du ménage, c'est-à-dire la personne devant dresser la liste des membres du ménage, était prié de confirmer qu'il avait 15 ans ou plus, de vérifier que le principal numéro de téléphone à l'écran appartenait à un membre du ménage, et d'entrer le nombre de lignes terrestres et de téléphones cellulaires appartenant au ménage. Si le numéro de téléphone indiqué n'appartenait à aucun membre du ménage, le cas était immédiatement transféré à l'ITAO. De même, si le numéro de téléphone indiqué était un numéro de téléphone d'entreprise, mais qu'il y avait d'autres numéros de téléphone associés au ménage, le cas était immédiatement transféré à l'ITAO.
3. Enfin, la personne en question était priée de dresser la liste des noms de toutes les personnes résidant dans le ménage, et d'indiquer leur âge et leur sexe.
4. L'application choisissait ensuite aléatoirement une personne du ménage comme répondant à qui on demanderait de remplir la partie contenu de l'enquête. Si le répondant sélectionné était la personne qui avait dressé la liste, il pouvait simplement poursuivre avec la partie contenu de l'enquête.
5. Si le répondant sélectionné n'était pas la personne qui avait dressé la liste (cette personne est appelée « répondant secondaire »), une adresse de courriel pouvant servir à communiquer avec le répondant secondaire était demandée. Si une adresse de courriel n'était pas fournie, le cas était transféré à l'ITAO pour qu'un intervieweur tente de communiquer avec lui.
6. À l'aide d'une adresse de courriel valide, une invitation était envoyée par courriel avec un nouveau code d'accès sécurisé pour l'ouverture de session. Si le répondant secondaire ne remplissait pas l'enquête en ligne avant la date limite, le cas était également transféré à l'ITAO.

6.2 Taux de réponse

Le taux global de réponse de l'ESG de 2016 a été de 50,8 %.

L'échantillon de 2016 a été sélectionné au moyen de la nouvelle base de sondage de l'ESG, ce qui a nécessité certains ajustements à la méthode utilisée pour calculer le taux de réponse. Depuis 2013, la nouvelle base de sondage comprend les ménages « dotés d'un téléphone cellulaire seulement », une population qui n'était pas couverte dans la base de sondage par composition aléatoire (CA) précédente. L'ajout de ces ménages à la base de sondage était essentiel puisque cette population représente une partie sans cesse croissante de la population et que la couverture était constamment en baisse avec la base de sondage précédente. Même si l'ajout de ces ménages est nécessaire pour assurer la couverture de la population canadienne, cette population est plus difficile à joindre. D'autres facteurs ont une incidence sur la comparabilité du taux de réponse au fil du temps : la façon dont le statut (dans le champ, hors champ) est déterminé dans le cadre du nouveau plan et l'utilisation de deux modes différents (par voie électronique ou par téléphone) pour recueillir les données.

6. Traitement des données

7.1 Saisie des données

À l'aide de l'application d'ITAO, les réponses aux questions ont été saisies directement sur ordinateur à mesure que se déroulait l'interview. Le programme de saisie des données de l'application d'ITAO comportait, pour chaque question, la spécification, l'intervalle de valeurs permises et l'intégration de règles de vérification, et suivait automatiquement le cheminement du questionnaire. Les données ont été

transmises électroniquement à Ottawa. Pour le questionnaire électronique, les répondants inscrivaient leurs réponses directement dans le questionnaire. Des règles de vérification et des cheminements automatiques ont été programmées pour s'assurer que les deux applications étaient identiques. Les données d'enquête étaient transmises à Statistique Canada par l'entremise d'un portail sécurisé.

7.2 Codage

Plusieurs questions permettaient des réponses écrites. Ces dernières ont été codées à des catégories existantes, lorsqu'il y avait une correspondance, sinon à de nouvelles catégories ou laissées dans la catégorie 'autre-spécifier' (si une correspondance avec une catégorie existante n'était pas possible ou si le nombre n'était pas suffisant pour créer une nouvelle catégorie). Dans la mesure du possible (p. ex., pour la profession, les types d'industries, la langue, le niveau de scolarité, le pays de naissance, la religion), le codage a été fait conformément aux systèmes de classification type utilisés pour le recensement de la population et le Programme du contenu harmonisé de Statistique Canada.

7.3 Vérification et imputation

Tous les enregistrements de l'enquête ont été soumis à des vérifications informatiques au cours de l'interview. Le système ITAO identifiait les valeurs « non valides » à mesure qu'elles étaient entrées. Ceci permettait à l'intervieweur de résoudre de tels problèmes immédiatement avec le répondant. Lorsqu'il ne pouvait pas résoudre correctement les erreurs décelées, l'intervieweur avait la possibilité de passer outre le problème et de transmettre les données au Bureau central pour qu'il le résolve. Lors de la vérification des données, les employés du Bureau central ont passé en revue tous les commentaires des intervieweurs et en ont tenu compte.

Le Bureau central a effectué les mêmes vérifications que l'application d'ITAO, ainsi que des vérifications plus poussées. Dans un petit nombre de cas, des réponses manquantes ou incorrectes étaient complétées, corrigées de manière déterministe, ou imputées à partir de renseignements obtenus ailleurs dans l'enquête.

La non-réponse n'était pas permise pour les questions requises pour la pondération. On a recouru à l'imputation dans les rares cas où les renseignements manquaient sur le sexe du répondant. Pour procéder à l'imputation, on a examiné en détail les données et tenu compte de toute donnée utile, comme l'âge ou le sexe d'autres membres du ménage, et des commentaires de l'intervieweur.

En 2016, les questions sur le revenu personnel n'ont pas été posées dans l'enquête. Les informations sur le revenu ont été obtenues à l'aide du couplage aux données fiscales des répondants qui ne s'étaient pas objecté audit couplage. Les informations sur le revenu ont été obtenues à partir du fichier T1FF de 2015 pour 90,6 % des répondants. Les informations manquantes pour les autres répondants ont été imputées. Contrairement à l'ESG de 2015, le **revenu familial** (soit le couplage direct avec une variable du fichier T1FF qui correspond au revenu familial du recensement) a été utilisé pour l'ESG de 2016 au lieu du revenu du ménage. Au total, la valeur du revenu familial a été obtenue pour 89,9 % des ménages. Les informations manquantes pour les autres répondants ont été imputées.

7.4 Création de variables combinées et dérivées

Un certain nombre de variables ont été dérivées en utilisant des éléments d'information du questionnaire. Dans certains cas, les variables dérivées sont simples et comprennent le regroupement de catégories. Dans d'autres, deux variables ou plus ont été combinées afin d'en créer une nouvelle. Le dictionnaire des données précise quelles variables sont dérivées et la façon dont elles l'ont été.

7. Estimation

Lorsqu'on utilise un échantillon probabiliste comme dans l'ESG, le principe qui sous-tend la démarche d'estimation est que chaque personne échantillonnée représente (outre elle-même) plusieurs personnes non échantillonnées. Ainsi, dans un échantillon aléatoire simple de 2 % d'une population de 1 000, chaque unité représente 50 éléments de la population. Le nombre de personnes représentées par une personne donnée dans l'échantillon est ce que l'on appelle habituellement le poids ou le facteur de pondération de la personne échantillonnée.

Le FMGD renferme les réponses et les renseignements connexes reçus de 19 609 répondants.

Trois facteurs de pondération ont été placés dans le FMGD. En voici l'énumération et l'explication :

WGHT_PER : Il s'agit du poids de base pour l'analyse au niveau des personnes, c'est-à-dire pour faire l'estimation du nombre de personnes (âgées de 15 ans et plus et autres que les pensionnaires d'établissement) ayant une ou plusieurs caractéristiques données. On doit utiliser WGHT_PER pour toutes les estimations au niveau des personnes. Ainsi, pour estimer le nombre de gens qui ont indiqué que l'utilisation de la technologie leur fait gagner du temps la valeur de WGHT_PER doit être additionnée sur tous les enregistrements ayant cette caractéristique.

En outre, les répondants ont été répartis (aléatoirement) en deux sous-échantillons de taille à peu près égale. La moitié des répondants se sont vu poser les questions du module sur la participation à des activités culturelles et l'autre moitié, les questions du module sur les activités sportives.

À la suite du fractionnement de l'échantillon, les ensembles de poids suivants ont été créés.

WGHT_CSP : Il s'agit du facteur de pondération pour l'analyse au niveau des personnes créé à l'aide de l'échantillon de personnes qui ont répondu aux questions du module sur les activités culturelles. Par exemple, on devrait faire la sommation de WGHT_CSP sur tous les enregistrements ayant cette caractéristique pour estimer le nombre de personnes qui, au cours des 12 derniers mois, ont assisté à des festivals culturels ou artistiques. Ce facteur de pondération est nul pour les répondants qui ont répondu aux questions du module sur la participation à des activités sportives.

WGHT_SNT : Il s'agit du facteur de pondération pour l'analyse au niveau des personnes créé à l'aide de l'échantillon de personnes qui ont répondu aux questions du module sur les activités sportives. Par exemple, pour estimer le nombre de personnes qui, au cours des 12 derniers mois, ont participé régulièrement à des activités sportives, on devrait faire la sommation de WGHT_SNT sur tous les enregistrements ayant cette caractéristique. Ce facteur de pondération est nul pour les répondants qui ont répondu aux questions du module sur la participation à des activités culturelles.

En plus des estimations de poids, des poids bootstrap ont été créés pour estimer la variance basée sur le plan de sondage³.

7.1 Pondération des personnes

Comme il a été mentionné précédemment, les enregistrements de la base de sondage sont des groupes de numéros de téléphone. Un échantillon aléatoire simple de ces enregistrements a été sélectionné dans chacune des strates. Chaque enregistrement dans une strate a donc des probabilités égales de sélection.

³ Trois ensembles de 500 poids bootstrap standards sont disponibles pour l'ESG de 2016 sur les Canadiens au travail et à la maison: WTBS_001 à WTBS_500 pour les 19 609 répondants; WSBS_001 à WSBS_500 pour les répondants aux questions du module sur les activités sportives; WCBS_001 à WCBS_500 pour les répondants aux questions du module sur les activités culturelles.

Cette probabilité correspond à :

$$\frac{\text{Nombre d'enregistrements échantillonnés dans la strate}}{\text{Nombre d'enregistrements dans la strate de la base de sondage}}$$

1) Calcul du poids initial

Certains ménages de la base de sondage étaient susceptibles d'être contactés au moyen de plus d'un enregistrement, en raison des regroupements de numéros de téléphone qui peuvent être sujets à erreur.

Comme il a été mentionné précédemment, les numéros de téléphone appartenant à la même adresse valide ont été regroupés dans la base de sondage. Toutefois, dans certains cas, le regroupement des numéros de téléphone pouvait être erroné (c.-à-d. que tous les numéros de téléphone regroupés n'appartenaient pas au même ménage). En outre, les numéros de téléphone qui n'ont pas pu être couplés à des adresses ont aussi été inclus dans la base de sondage. Il est possible que certains de ces numéros de téléphone aient correspondu à des ménages déjà couverts par les numéros de téléphone couplés à des adresses.

Par conséquent, une série de questions a été ajoutée à l'enquête, afin de déterminer la prévalence de ces situations. Plusieurs ajustements ont été apportés à la probabilité initiale de sélection, afin de tenir compte du fait que ces ménages avaient une probabilité plus élevée d'être sélectionnés (c.-à-d. qu'ils pouvaient être contactés par l'entremise de plus d'un regroupement de numéros de téléphone). Par conséquent, le poids initial représente l'inverse de cette probabilité de sélection corrigée. Le poids initial résultant est un poids de ménage.

2) Suppression des enregistrements hors du champ de l'enquête

Les numéros de téléphone associés à des entreprises, des établissements ou d'autres logements hors du champ de l'enquête, de même que les numéros hors service, sont tous des exemples de cas hors du champ de l'enquête pour la base téléphonique. Les enregistrements dont les numéros de téléphone sont tous hors du champ de l'enquête sont simplement supprimés du processus, ce qui laisse uniquement les enregistrements faisant partie du champ de l'enquête dans l'échantillon. Ces enregistrements dans le champ de l'enquête conservent le même poids initial que celui décrit à l'étape précédente.

3) Ajustement en trois étapes pour tenir compte de la non-réponse

On a corrigé les poids des numéros de téléphone avec entrée en communication pour représenter les numéros de téléphone sans entrée en communication.

Nous avons catégorisé dans trois groupes les numéros de téléphone sans entrée en communication; il s'agit des catégories avec information auxiliaire (particulièrement avec la liste complète des membres du ménage), ceux avec des informations auxiliaires provenant de sources disponibles à Statistique Canada et ceux sans information auxiliaire.

Cette correction de non-réponse s'est faite en trois étapes. Il y a d'abord eu une correction de non-réponse complète (ménages pour lesquels nous ne disposions d'aucune information auxiliaire). La correction a été faite indépendamment pour chaque strate. Dans la seconde étape, une correction de non-réponse a été faite à partir de sources disponibles à Statistique Canada. Ces ménages avaient certaines informations auxiliaires permettant de modéliser la propension à répondre. À la troisième étape, des corrections ont été faites à des cas de non-réponse partielle. Ces ménages avaient certaines informations auxiliaires permettant de modéliser la propension à répondre. Les deux derniers

ajustements ont été effectués de façon indépendante dans chaque région⁴. La combinaison de ces trois corrections constitue ce qu'on appelle le facteur 1.

On a ensuite éliminé les numéros de téléphone pour lesquels aucun contact n'a été établi.

4) Calcul du poids de la personne

On a ensuite calculé un poids de la personne pour le répondant en multipliant le poids du ménage par le nombre de membres du ménage admissibles à l'échantillonnage aux fins de l'enquête (nombre de personnes de 15 ans et plus).

Cette étape crée le poids personne = Poids initial du ménage x Facteur 1 x Nombre membre du ménage éligible

5) Ajustement des poids de la personne d'après les totaux externes

Les poids de la personne ont été corrigés plusieurs fois au moyen d'une méthode itérative du quotient. Cette méthode fait en sorte que les estimations produites à partir de l'échantillon correspondent à certains totaux de référence externes. Deux ensembles de valeurs externes de référence ont été employés. Il s'agissait dans les deux cas de totaux de population : pour les strates (géographiques) et les groupes d'âge et sexe par province.

Il convient de souligner que les personnes vivant dans des ménages sans service téléphonique (ou ayant un service téléphonique non couvert par la base de sondage) sont incluses dans les références externes, même si elles n'ont pas été échantillonnées.

5a) Ajustement strate

On a corrigé les poids de la personne dans les enregistrements de chaque strate (géographique), de sorte que les estimations de population concordent avec les projections de la population dans ce même domaine. Pour cela, on a multiplié le poids de la personne de chaque enregistrement dans une strate donnée par le rapport suivant :

$$\frac{\text{Projection de la population pour la strate}}{\text{Somme des poids de la personne de la strate}}$$

5b) Ajustement pour le domaine province - âge - sexe

L'étape suivante de la pondération consiste à corriger les poids de sorte qu'ils concordent avec la distribution projetée de la population selon le domaine province-âge-sexe. Des projections de la population pour les hommes et les femmes ont été obtenues pour les seize groupes d'âge suivants :

15-19	20-24	25-29	30-34
35-39	40-44	45-49	50-54
55-59	60-64	65-69	70-74
75-79	80-84	85-89	90 +

Pour chaque domaine province-âge-sexe résultant, le poids de la personne de chaque enregistrement dans ce domaine a été multiplié par le rapport suivant :

$$\frac{\text{Projection de la population pour le domaine province - âge - sexe}}{\text{Somme des poids de la personne pour le domaine province - âge - sexe}}$$

⁴ Les cinq régions géographiques sont l'Atlantique, le Québec, l'Ontario, les Prairies et la Colombie-Britannique.

Dans les cas où la taille d'échantillon était petite, on a combiné deux domaines d'âge ou plus de la même province et du même sexe avant de procéder à l'ajustement.

5c) Ajustement par la méthode itérative du quotient

Les poids calculés pour chaque répondant ont été ajustés plusieurs fois à par la méthode itérative du quotient pour s'assurer que les estimations des totaux produites par strate et par groupe province-âge-sexe concordent avec les totaux de référence externes. Cette correction a été faite en répétant les étapes 5a) et 5b) de la pondération jusqu'à ce que la dernière répétition de chaque étape produise une correction négligeable des poids.

6) Poids final de la personne

Le poids produit à la fin de l'étape 5) est le poids final de la personne WGHT_PER qui figure dans le fichier principal.

7) Poids de la personne fondé sur l'échantillon fractionné

Afin de prendre en compte le plan d'échantillonnage fractionné du questionnaire, deux autres poids affectés à la personne, WGHT_CSP et WGHT_SNT, ont été créés à l'aide de la même approche que celle utilisée pour le poids affecté à la personne. WGHT_CSP est nul pour les répondants à qui on ne posait pas les questions du module sur la participation aux activités culturelles, et WGHT_SNT est nul pour les répondants à qui on ne posait pas les questions du module sur les activités sportives.

7.2 Politique de pondération

Les utilisateurs ne devraient diffuser aucune totalisation non pondérée, ni faire des analyses fondées sur des résultats d'enquête non pondérés. Comme il a été expliqué à la section 8.1, plusieurs ajustements des poids ont été faits en fonction de la province, de la strate, de l'âge et du sexe du répondant. Les taux d'échantillonnage et les taux de non-réponses variaient considérablement d'une province à l'autre, et les taux de non-réponse variaient selon les caractéristiques démographiques. Par exemple, les non-répondants étaient souvent plus susceptibles d'être des hommes et d'être jeunes. Dans l'échantillon des répondants, 2,0 % étaient des hommes âgés de 20 à 24 ans, tandis que dans la population totale, environ 4,1 % étaient des hommes âgés de 20 à 24 ans. Par conséquent, il est évident que les dénombrements non pondérés de l'échantillon ne peuvent pas être considérés comme représentatifs de la population visée par l'enquête.

Le nombre total de ménages de l'échantillon dans le champ de l'enquête a été estimé à 38 569. Parmi ces ménages, 19 609 réponses utilisables ont été obtenues. Il en résulte un taux de réponse global de 50,8 %. La répartition des catégories de non-réponse et de réponse est présentée dans le tableau suivant :

Source	Nombre	%
1. Non-réponse des ménages	14 482	37,5
2. Refus de la personne sélectionnée	1 165	3,0
3. Autre non-réponse de la personne sélectionnée	3 313	8,6
4. Réponse (ITAO)	13 520	35,1
5. Réponse (QE)	6 089	15,8
Total des ménages	38 569	100,0

Au total, le nombre de cas de non-réponse est estimé à 18 960 cas. Les catégories 2 et 3 correspondent à la non-réponse se produisant après la sélection du répondant dans le ménage. La catégorie « autre non réponse » (3) comprend les cas où une réponse n'a pu être obtenue à cause de difficultés linguistiques ou d'autres problèmes. Les réponses obtenues au moyen du questionnaire électronique (QE) représentent 31,1 % des 19 609 réponses obtenues.

7.3 Types d'estimations

Deux types d'estimations 'simple' sont rendus possibles à partir des résultats de l'Enquête sociale générale. Il y a les estimations qualitatives (estimations du nombre ou de la proportion de cette population possédant certaines caractéristiques) ainsi que les estimations quantitatives qui représentent des quantités ou des moyennes. Les estimations et analyses plus complexes sont traitées à la section 8.5.

7.3.1 Estimations qualitatives

La population visée par l'ESG est formée des personnes de 15 ans et plus, hors institutions, vivant dans l'une des dix provinces. Les estimations qualitatives sont des estimations du nombre ou de la proportion de cette population possédant certaines caractéristiques. Le nombre de personnes (4 657 522) qui se disent en excellente santé (SRH_110 = 1) est un exemple d'estimation de ce genre. Ces estimations s'obtiennent facilement par sommation des poids de la personne (WGHT_PER) des enregistrements où

figure la caractéristique étudiée. Toutefois, cette estimation ne contient aucune correction pour la non réponse à la question.

Si nous supposons que les personnes qui ont refusé de répondre à la question ou qui ont répondu «ne sait pas» sont caractérisées par la même distribution que celles qui ont répondu, on peut rajuster l'estimation. C'est pourquoi nous estimons la proportion de la population cible possédant cette caractéristique en ne tenant pas compte des enquêtés ayant inscrit « non déclaré » ou « ne sait pas » comme réponse à la question SRH_110 et en calculant le rapport de sommation des poids entre les répondants qui se sont dits en excellente santé (SRH_110=1) et tous les répondants qui ont répondu à la question (SRH_110=1, 2, 3, 4, ou 5). Ensuite multipliée par la taille de la population visée, la proportion en question donne l'estimation finale (à noter que cette correction n'est pas obligatoire, mais qu'elle peut être apportée au besoin) :

$$4,680,203 = 30,078,789 \times \frac{4,657,522}{29,933,021}$$

30 078 789 est l'estimation du nombre de personnes de 15 ans et plus dans la population (population visée). 29 933 021 correspond à la sommation des poids de tous les répondants qui ont répondu à la question SRH_110 (c.-à-d. SRH_110 = 1, 2, 3, 4, ou 5). Si la proportion de réponses « ne sait pas » ou « refus » est forte, l'écart entre les deux estimations est important.

7.3.2 Estimations quantitatives

Certaines variables de FMGD de l'ESG sont quantitatives (p. ex., l'âge, le nombre de semaines de travail au cours des 12 derniers mois). Pour ces variables, il est possible d'obtenir des estimations comme le nombre moyen de semaines de travail au cours des 12 derniers mois. Ces estimations quantitatives sont de la forme :

$$\text{Estimée (moyenne)} = \frac{X}{Y}$$

Le numérateur (X) est une estimation quantitative du total de la variable étudiée (par exemple, le nombre de semaines de travail au cours des 12 derniers mois) pour une sous-population donnée (par exemple, les hommes qui ont travaillé au cours des 12 derniers mois). Dans cet exemple, on calculerait X en multipliant le poids de la personne (WGHT_PER) par la valeur de la variable étudiée (WET_110) si on la connaît, 1 <= WET_110 <= 52, (c.-à-d. ne correspondant pas aux codes « 96 », « 97 », « 98 » ou « 99 ») et en faisant la somme de ce produit sur tous les enregistrements d'hommes qui avaient un emploi, c.-à-d. SEX=1 et (1 <= WET_110 <= 52), ce qui donne 482 951 168.

Le dénominateur (Y) est l'estimation qualitative du nombre de membres de cette sous population (répondants de sexe masculin qui ont travaillé dans les 12 derniers mois). Dans le présent exemple, on calculerait Y par sommation du poids de la personne (WGHT_PER) sur tous les répondants de sexe masculin pour lesquels 1 <= WET_110 <= 52, ce qui donnerait 10 975 964.

On calcule les estimations X et Y indépendamment, puis on divise la première par la seconde pour obtenir l'estimation quantitative. Le nombre moyen de semaines se calcule donc comme suit:

$$\frac{482,951,168}{10,975,964} = 44.0$$

7.4 Lignes directrices pour l'analyse

Comme l'explique en détail la section 5, les répondants à l'ESG ne forment pas un échantillon aléatoire simple de la population cible. En effet, l'enquête se fonde sur un plan de sondage complexe qui prévoit une stratification et un échantillonnage à plusieurs degrés, ainsi que la sélection des répondants avec

probabilités inégales. L'utilisation des données provenant d'une enquête aussi complexe pose 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.

Le plan de sondage de l'ESG est un plan stratifié où la fraction d'échantillonnage varie considérablement selon la strate. Aussi certaines régions sont-elles surreprésentées dans l'échantillon (comparativement à leur population) et d'autres, relativement sous-représentées; donc, l'échantillon non pondéré n'est pas représentatif de la population cible, même s'il n'y a aucune non-réponse. Le taux de non-réponse peut varier selon le groupe démographique, ce qui rend l'échantillon non pondéré encore moins représentatif.

Les poids de sondage doivent être utilisés pour le calcul des estimations ou les analyses afin de tenir compte dans la mesure du possible de la sur et sous-représentation géographique ou de la sous ou surreprésentation des groupes âge-sexe ou des mois de l'année dans le fichier non pondéré. Bien que de nombreuses méthodes d'analyse qui font partie des progiciels statistiques permettent d'utiliser des poids, le poids diffère souvent sous l'angle de la définition ou du sens de celui qui convient à une enquête par sondage, de sorte que si les estimations faites au moyen de ces progiciels sont exactes dans bien des cas, les variances calculées n'ont 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 ré échantillonnage en divisant chaque poids par le poids moyen global avant de procéder à l'analyse. La section 9 donne une description plus détaillée de la variabilité d'échantillonnage et de la fiabilité des données.

7.5 Estimation du nombre de personnes à l'aide de WGHT_PER au fichier principal

Comme il est mentionné plus haut, on a attribué un poids de base à chaque personne échantillonnée et, tel que décrit à la section 7.1, on a ajusté ces poids de sorte qu'ils reflètent la composition par âge et sexe des populations provinciales telles qu'estimées par Statistique Canada pour chaque mois couvert par le cycle 30.

$$\sum_{i=1}^{n=19,609} WGHT_{PER} = 30,078,789^1$$

¹ Estimation du nombre de personnes de 15 ans et plus dans la population.

En règle générale, si l'estimation vise la personne comme unité d'observation, on doit utiliser le fichier principal et WGHT_PER. Comme exemples, citons le nombre moyen de semaines travaillées par les personnes de 25 à 29 ans, le pourcentage de gens pour qui l'activité principale au cours des 12 derniers mois était d'aller à l'école et le nombre de personnes de 25 à 44 ans qui sont présentement à l'école, collège, CEGEP ou université.

Le dernier cas se calculerait ainsi : on ferait la sommation de WGHT_PER sur tous les enregistrements au fichier principal avec $2 \leq AGEGR10 \leq 3$ et $ESC1_01 = 1$ pour ainsi constater que 954 136 personnes de 25 à 44 ans sont présentement à l'école, collège, CEGEP ou université.

8. 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 d'une autre manière toute estimation calculée d'après le FMGD de l'ESG.

La présente section du document donne les lignes directrices que doit suivre l'utilisateur du FMGD. 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 quatre grands groupes, à savoir la taille minimale d'échantillon pour le calcul d'estimations, les lignes directrices concernant la variabilité d'échantillonnage, l'estimation de la variabilité d'échantillonnage et les lignes directrices concernant l'arrondissement.

8.1 Taille minimale d'échantillon pour le calcul d'estimations

L'utilisateur doit déterminer le nombre d'enregistrements du FMGD qui ont fourni les données entrant dans le calcul d'une estimation. Ce nombre devrait être d'au moins 15 dans le cas des personnes ou des ménages. Si le nombre d'enregistrements contribuant à l'établissement de l'estimation pondérée est de moins de 15, celle-ci 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.

8.2 Lignes directrices concernant la variabilité d'échantillonnage

Les estimations tirées de l'enquête sont fondées sur un échantillon de personnes. Les estimations obtenues seraient peut-être sensiblement différentes si on réalisait un recensement exhaustif en reprenant par exemple le même questionnaire, les mêmes intervieweurs et superviseurs et les mêmes méthodes de traitement que pour l'enquête elle-même. 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 le système ITAO ou QE 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 et procédé à des contrôles du codage et des vérifications pour confirmer la logique de traitement.

8.2.1 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 quand l'intervieweur n'a pu communiquer avec le répondant, lorsqu'aucun membre du ménage ne pouvait fournir les renseignements (peut-être à cause d'un problème de langue) ou quand le répondant a refusé de participer à l'enquête. On traite les cas de non-réponse totale en corrigeant le poids appliqué aux ménages 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 question, a refusé de répondre ou qu'il n'a pu se souvenir du renseignement demandé.

8.2.2 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 de l'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 la population est normale. Il y a environ 68 chances sur 100 que l'écart entre une estimation de l'échantillon et la valeur vraie pour la population soit inférieur à une erreur-type, environ 95 chances sur 100 que l'écart soit inférieur à deux erreurs-types et environ 99 chances sur 100 que l'écart soit inférieur à trois erreurs-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 (CV). On obtient le CV 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 CV de la première estimation est de 50 % (10/20), tandis que celui de la deuxième est de 1 % (10/1 000).

Le choix entre utiliser l'erreur-type de l'estimation ou le coefficient de variation comme mesure de la variabilité d'échantillonnage est un choix laissé à l'utilisateur basé sur ses analyses spécifiques. Des directions pour la publication des estimations utilisant le CV sont données dans la prochaine section.

Lorsqu'il existe un nombre suffisant d'observations, l'utilisateur peut calculer les erreurs-type et les coefficients de variables associés aux estimations à l'aide des poids bootstrap inclus avec les données (voir la section 9.2.3 pour les lignes directrices à suivre lorsqu'un coefficient de variation est utilisé et la section 9.3 pour plus de détails sur les progiciels qui permettent l'utilisation de poids bootstrap).

8.2.3 Lignes directrices pour la diffusion des estimations

Avant de diffuser *et/ou* de publier des estimations tirées du FMGD, l'utilisateur devrait consulter le tableau ci-dessous et se conformer aux lignes directrices correspondant à la valeur du CV de l'estimation.

Type d'estimation	Coefficient de variation	Énoncé de la politique
1. Variabilité d'échantillonnage moyenne	De 0,0 % à 16,5 %	On peut envisager une diffusion générale non restreinte des estimations. Aucune annotation particulière n'est nécessaire.
2. Forte variabilité d'échantillonnage	De 16,6 % à 33,3 %	On peut envisager une diffusion générale non restreinte des

		estimations, avec une mise en garde quant à la forte variabilité d'échantillonnage qui y est associée.
3. Très forte variabilité d'échantillonnage	33,4 % et plus	On ne devrait généralement pas diffuser les estimations, mais si c'est le cas, les estimations devraient être diffusées avec beaucoup de prudence et être accompagnées d'une note bien visible mentionnant la très forte variabilité d'échantillonnage qui y est associée.

8.3 Estimation de la variance au moyen des poids bootstrap

Pour juger de la qualité d'une estimation et en calculer le c.v., on doit d'abord établir l'erreur-type. On a aussi besoin de l'erreur-type de l'estimation pour les intervalles de confiance. Dans l'ESG, on utilise des plans d'échantillonnage et d'étalonnage à plusieurs degrés et, par conséquent, il n'y a pas de formule simple d'estimation de la variance qu'on puisse appliquer. Il faut donc procéder par approximation. On se sert de la méthode bootstrap afin de prendre en compte l'échantillonnage et l'étalonnage au moment d'estimer la variance. Grâce au logiciel disponible pour calculer les variances avec l'aide des poids bootstrap (dont il est question dans la sous-section suivante), la méthode est relativement facile pour les utilisateurs.

Cette technique nécessite la division des enregistrements du fichier de microdonnées en sous-groupes (ou répétitions) et à déterminer la variation dans les estimations, de répétition en répétition. Les répétitions sont le résultat de la sélection indépendante dans chaque strate d'un échantillon aléatoire simple avec remplacement de (n-1) des n unités de l'échantillon. À noter que, comme la sélection s'opère avec remise, une unité peut être choisie plusieurs fois. Un poids bootstrap fondé sur l'échantillon bootstrap est calculé pour chaque unité échantillonnée dans la strate. Ce processus (sélection d'échantillons aléatoires simples, nouveau calcul de poids pour chaque strate) est répété à B reprises, B étant une valeur élevée, ce qui donne à B différents poids bootstrap initiaux. Dans l'ESG, on utilise habituellement B=500 pour produire 500 poids bootstrap.

Ces poids sont par la suite corrigés selon le même processus de pondération que les poids réguliers de personne : correction pour la non-réponse, étalonnage, etc. On obtient 500 poids bootstrap finaux pour chaque unité de l'échantillon. La variation obtenue entre les 500 estimations possibles produites à partir des 500 poids bootstrap correspond à une estimation de la variance de l'estimateur basé sur les poids originaux.

8.4 Arrondissement

Afin que les estimations produites d'après les fichiers de microdonnées de l'Enquête sociale générale 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.

8.4.1 Lignes directrices concernant l'arrondissement

- 1) Les estimations de totaux qui figurent dans le corps d'un tableau statistique doivent être arrondies au millier près par la méthode d'arrondissement classique (voir la définition à la section 9.4.2).
- 2) Les totaux partiels marginaux et les totaux marginaux des tableaux statistiques doivent être calculés à partir de leurs composantes correspondantes non arrondies, puis arrondis à leur tour au millier près à l'aide de la technique 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 au millier 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.

8.4.2 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.

9. Autres renseignements

D'autres renseignements sur l'enquête peuvent être obtenus auprès des personnes mentionnées ci-dessous.

Gestionnaire d'enquête

Anna Kemeny

Division de la statistique sociale et autochtone

(613) 854-8420

Anna.kemeny@canada.ca

Les données de l'enquête sont présentées sous forme de rapports publiés, de totalisations spéciales et de fichier de microdonnées. Le fichier peut être obtenu auprès de la Division de la statistique sociale et autochtone de Statistique Canada. Les totalisations sont fournies à un prix qui tient compte des ressources nécessaires pour les produire.

Pour obtenir une copie du fichier de microdonnées ou pour commander des tableaux spéciaux, veuillez contacter :

Service à la clientèle

Division de la statistique sociale et autochtone

Statistique Canada

statcan.sasdcclientservices-dsseaserviceaclientele.statcan@canada.ca

Annexe A – Codes de citoyenneté

CNTRYTEXT	CTZCODE
Autre - Précisez	001
Angola	024
États-Unis (É.-U.)	103
Belize	201
Costa Rica	202
El Salvador	203
Guatemala	204
Honduras	205
Mexique	206
Nicaragua	207
Antigua-et-Barbuda	302
Bahamas	304
Barbade	305
Cuba	308
Dominique	309
Dominicaine, République	310
Grenade	311
Haïti	313
Jamaïque	314
Saint-Kitts-et-Nevis	319
Sainte-Lucie	320
Saint-Vincent-et-les Grenadines	321
Trinité-et-Tobago	322
Argentine	401
Bolivie	402
Brésil	403
Chili	404
Colombie	405
Équateur	406
Guyana	409
Paraguay	410
Pérou	411
Suriname	412
Uruguay	413
Venezuela	414
Autriche	501
Belgique	502
France et dépendances	503

Allemagne	505
Liechtenstein	506
Luxembourg	507
Monaco	508
Pays-Bas et dépendances	509
Suisse	511
Bulgarie	521
Tchécoslovaquie	522
Tchèque, République	523
Estonie	524
Hongrie	525
Lettonie	526
Lituanie	527
Roumanie	529
Slovaquie	530
Union des Républiques Socialistes Soviétiques (URSS)	531
Bélarus	533
Moldova, République de	534
Russie, Fédération de	535
Ukraine	536
Irlande, République d'	541
Irlande (Eire)	542
Britannique	543
Royaume-Uni, citoyens britanniques	544
Danemark	546
Finlande	547
Islande	548
Norvège	549
Suède	550
Guernesey	553
Jersey	555
Albanie	561
Andorre	562
Bosnie-Herzégovine	563
Croatie	564
Grèce	566
Italie	567
Macédoine, République de	568
Malte	569
Portugal	571
Saint-Marin	572

Yugoslavie	573
Slovénie	574
Espagne	575
Saint-Siège (État de la Cité du Vatican)	576
Serbie et Montenegro	578
Kosovo	583
Panama	591
Bénin	601
Burkina Faso	602
Cap-Vert	603
Côte d'Ivoire	604
Gambie	605
Ghana	606
Guinée	607
Guinée-Bissau	608
Libéria	609
Mali	610
Mauritanie	611
Niger	612
Nigéria	613
Sénégal	615
Sierra Leone	616
Togo	617
Burundi	621
Comores	622
Djibouti	623
Érythrée	624
Éthiopie	625
Kenya	626
Madagascar	627
Malawi	628
Maurice	629
Mozambique	631
Rwanda	633
Seychelles	634
Somalie	635
Tanzanie, République unie de	636
Ouganda	637
Zambie	638
Zimbabwe	639
Algérie	651

Égypte	652
Libye	653
Maroc	654
Soudan	655
Tunisie	656
Sahara occidental	657
Cameroun	662
Centrafricaine, République	663
Tchad	664
Congo, République du	665
Guinée équatoriale	666
Gabon	667
Sao Tomé-et-Principe	668
Congo, République démocratique du (Zaire)	669
Botswana	681
Lesotho	682
Namibie	683
Afrique du Sud, République d'	684
Swaziland	685
Afghanistan	701
Chypre	702
Iran	703
Turquie	704
Bahreïn	705
Iraq	706
Israël	707
Jordanie	708
Koweït	709
Liban	710
Oman	711
Palestine	712
Qatar	713
Arabie saoudite	714
Syrie	715
Émirats arabes unis	716
Cisjordanie et bande de Gaza (Palestine)	717
Yémen	718
Arménie	720
Azerbaïdjan	721
Géorgie, République de	722
Kazakhstan	723

Kirghizistan	724
Tadjikistan	725
Turkménistan	726
Ouzbékistan	727
Chine, République populaire de	732
Japon	734
Corée du Nord	735
Corée du Sud	736
Mongolie	739
Taïwan	740
Brunéi Darussalam	751
Cambodge	752
Indonésie	753
Laos	754
Malaisie	755
Birmanie (Myanmar)	756
Myanmar (Birmanie)	756
Philippines	757
Singapour	758
Thaïlande	759
Viet Nam	760
Bangladesh	771
Bhoutan	772
Inde	773
Maldives	774
Népal	775
Pakistan	776
Sri Lanka	777
Timor oriental (Timor leste)	781
Kurdistan	782
Bande de Gaza	784
Australie	802
Fidji	804
Kiribati	807
Marshall, îles	808
Micronésie, États fédérés de	809
Nauru	810
Nouvelle-Zélande	812
Palaos	813
Papouasie-Nouvelle-Guinée	814
Salomon, îles	816

Tonga	817
Tuvalu	818
Vanuatu	820
Samoa	822
Pologne	825
Puerto Rico	831
Îles mineures éloignées des États-Unis	832
Apatrides	906
Royaume-Uni, territoires dépendants	908
Canada	991
111 - Aucun autre pays	995

Annexe B – Codes de pays de naissance

CNTRYTEXT	CNTRYCODE
Canada	11124
Groenland	11304
Saint-Pierre-et-Miquelon	11666
États-Unis (É.-U.)	11840
Belize	12084
Costa Rica	12188
El Salvador	12222
Guatemala	12320
Honduras	12340
Mexique	12484
Nicaragua	12558
Panama	12591
Antilles néerlandaises	13000
Antigua et Barbuda	13028
Bahamas	13044
Barbade	13052
Bermudes	13060
Vierges, Îles (britanniques)	13092
Caïmanes, Îles	13136
Cuba	13192
Dominique	13212
Dominicaine, République	13214
Grenade	13308
Guadeloupe	13312
Haïti	13332
Jamaïque	13388

Martinique	13474
Montserrat	13500
Curaçao	13531
Aruba	13533
Saint-Martin (partie néerlandaise)	13534
Bonaire, Saint-Eustache and Saba	13535
Porto Rico	13630
Saint-Barthélemy	13652
Saint-Kitts-et-Nevis	13659
Anguilla	13660
Sainte-Lucie	13662
Saint-Martin (partie française)	13663
Saint-Vincent-et-les Grenadines	13670
Trinité-et-Tobago	13780
Turks et Caïques, Îles	13796
Vierges, Îles (États-Unis)	13850
Argentine	14032
Bolivie	14068
Brésil	14076
Chili	14152
Colombie	14170
Équateur	14218
Falkland, Îles (Malvinas)	14238
Géorgie du Sud et les Îles Sandwich du Sud	14239
Guyane française	14254
Guyana	14328
Paraguay	14600
Pérou	14604
Suriname	14740
Uruguay	14858
Venezuela	14862
Autriche	21040
Belgique	21056
France	21250
Allemagne	21276
Liechtenstein	21438
Luxembourg	21442
Monaco	21492
Pays-Bas	21528
Suisse	21756
Autriche-Hongrie	22000

Tchecoslovaquie	22010
Union des Republiques Socialistes Sovietiques	22020
Bulgarie	22100
Bélarus	22112
Tchèque, République	22203
Estonie	22233
Hongrie	22348
Lettonie	22428
Lituanie	22440
Moldova	22498
Pologne	22616
Roumanie	22642
Russie, Fédération de	22643
Slovaquie	22703
Ukraine	22804
Danemark	23208
Féroé, îles	23234
Finlande	23246
Åland, îles	23248
Islande	23352
Irlande	23372
Irlande, République d'	23373
Norvège	23578
Svalbard et île Jan Mayen	23744
Suède	23752
La Grande-Bretagne	23825
Angleterre	23826
l'Irlande du Nord	23827
Écosse	23828
Royaume-Uni	23829
Le pays de Galles	23830
Guernesey	23831
Jersey	23832
Île de Man	23833
Yugoslavie	24000
Albanie	24008
Andorre	24020
Bosnie-Herzégovine	24070
Croatie	24191
Gibraltar	24292
Grèce	24300

Saint-Siège (État de la Cité du Vatican)	24336
Italie	24380
Malte	24470
Monténégro	24499
Portugal	24620
Saint-Marin	24674
Serbie	24688
Slovénie	24705
Espagne	24724
Macédoine (Region)	24807
Macédoine, République de	24808
Kosovo	24983
Cap-Vert	31132
Bénin	31204
Gambie	31270
Ghana	31288
Guinée	31324
Côte d'Ivoire	31384
Libéria	31430
Mali	31466
Mauritanie	31478
Niger	31562
Nigéria	31566
Guinée-Bissau	31624
Sainte-Hélène-et-Ascension	31654
Sénégal	31686
Sierra Leone	31694
Togo	31768
Burkina Faso	31854
Burundi	32108
Comores	32174
Mayotte	32175
Éthiopie	32231
Érythrée	32232
Djibouti	32262
Kenya	32404
Madagascar	32450
Malawi	32454
Maurice	32480
Mozambique	32508
Réunion	32638

Rwanda	32646
Seychelles	32690
Somalie	32706
Zimbabwe	32716
Ouganda	32800
Tanzanie	32834
Zambie	32894
Algérie	33012
Libye	33434
Maroc	33504
Soudan	33736
Tunisie	33788
Égypte	33818
Angola	34024
Cameroun	34120
Centrafricaine, République	34140
Tchad	34148
Congo, République du	34178
Congo, République démocratique du (Zaire)	34180
Congo, non spécifié	34181
Guinée équatoriale	34226
Gabon	34266
Sao Tomé-et-Principe	34678
Botswana	35072
Lesotho	35426
Namibie	35516
Afrique du Sud, République d'	35710
Swaziland	35748
Kurdistan	41000
Afghanistan	41004
Azerbaïdjan	41031
Bahreïn	41048
Arménie	41051
Chypre	41196
Géorgie, République de	41268
Georgie	41268
Bande de Gaza	41275
Palestine	41276
Cisjordanie et bande de Gaza (Palestine)	41277
Iran	41364
Iraq	41368

Israël	41376
Kazakhstan	41398
Jordanie	41400
Koweït	41414
Kirghizistan	41417
Liban	41422
Oman	41512
Qatar	41634
Arabie saoudite	41682
Syrie	41760
Tadjikistan	41762
Émirats arabes unis	41784
Turquie	41792
Turkménistan	41795
Ouzbékistan	41860
Yémen	41887
Corée	42000
Chine	42156
Taïwan	42158
Hong-Kong région administrative spéciale	42344
Japon	42392
Corée du Nord	42408
Corée du Sud	42410
Macao région administrative spéciale	42446
Mongolie	42496
Brunéi Darussalam	43096
Birmanie (Myanmar)	43104
Cambodge	43116
Indonésie	43360
Laos	43418
Malaisie	43458
Philippines	43608
Timor-Leste (Timor oriental)	43626
Singapour	43702
Viet Nam	43704
Thaïlande	43764
Bangladesh	44050
Bhoutan	44064
Océan Indien, Territoire britannique de l'	44086
Sri Lanka	44144
Inde	44356

Maldives	44462
Népal	44524
Pakistan	44586
Océanie	50000
Samoa américaines	51016
Australie	51036
Salomon, îles	51090
Christmas, île	51162
Cocos (Keeling), îles	51166
Cook, îles	51184
Fidji	51242
Polynésie française	51258
Kiribati	51296
Guam	51316
Nauru	51520
Nouvelle-Calédonie	51540
Vanuatu	51548
Nouvelle-Zélande	51554
Niué	51570
Norfolk, île	51574
Mariannes du Nord, îles	51580
Îles mineures éloignées des États-Unis	51581
Micronésie, États fédérés de	51583
Marshall, îles	51584
Palaos	51585
Papouasie-Nouvelle-Guinée	51598
Pitcairn	51612
Tokelau	51772
Tonga	51776
Tuvalu	51798
Wallis et Futuna	51876
Samoa	51882
Îles adjacentes à l'antarctique	60000
Antarctique	61010
Bouvet, île	61074
Terres australes françaises	61260
Heard, île et McDonald, îles	61334
Ne(e) en mer	80000
Autre - Précisez	90000

Annexe C – Codes de religion

Religion	Code
Adventiste du septième jour	102010
Agnostique	301000
Ahmadiyya	105051
Alliance chrétienne et missionnaire	102100
Amish	102270
Anabaptiste	102271
Anglicane	102020
Anglo-Catholique	102021
Animisme	201100
Apostolique	102381
Apostolique arménienne	103021
Apostolique nouvelle	102340
Armée du Salut	102490
Athée	302000
Autre - Précisez	900000
Autre catholique	101010
Autres chrétiennes	102380
Autres orthodoxes	103055
Baha'i	105010
Baptiste	102050
Baptiste évangélique	102051
Baptiste méridional	102052
Baptiste spirituel	102053
Bouddhiste	105020
Catholique	101000
Catholique arménienne	101011
Catholique chaldéenne	101012
Catholique grecque ou byzantine	101015
Catholique orientale	101013
Catholique rite oriental	101014
Catholique romaine	101030
Catholique syrienne	101018
Catholique ukrainienne	101040
Centre de renaissance	102391
Chamanisme	206002
Chiite	105054
Chrétien intersectaire	102240
Chrétien non confessionnel	102360

Chrétien orthodoxe d'Antioche	103010
Chrétienne	102130
Chrétienne régénérée	102131
Christadelphe	102090
Communauté du Christ	102481
Communion internationale dans la grâce	102551
Congrégation chrétienne	102120
Cri	201002
Culte d'ancêtre	201101
Déné	201003
Doukhobor	102037
Druide	206001
Druzes	105052
Eckankar	105001
Église autochtone	102389
Église catholique nationale polonaise	101020
Église chrétienne apostolique	102030
Église chrétienne des Disciples du Christ	102180
Église chrétienne réformée	102140
Église covenant	102432
Église covenant évangélique	102382
Église de Dieu	102150
Église de Jésus-Christ des saints des derniers jours	102160
Église de l'Ecosse/Presbytétien écossais	102431
Église du Nazaréen	102170
Église évangélique libre	102211
Église évangile, four square	102401
Église libre	102434
Église libre réformée	102070
Église métropolitaine de la Communauté	102388
Église missionnaire évangélique	102210
Église réformée hollandaise	102190
Église syrienne Mar Thoma/Marthomite	102386
Église unie	102530
Églises réformée du Canada et d'Amérique	102071
Évangélique	102383
Évangile	102404
Évangile de l'union	102040
Éxtremisme chrétien	102384
Fourth Way	209001
Free Thinker	303000

Frères dans le Christ	102060
Frères de Plymouth ou Frères chrétiens	102420
Frères mennonites	102273
Frères moraves	102330
Full Gospel	102402
Glad Tidings	102403
Gnostique	210000
Hindoue	105040
Humaniste	304000
Huttérite	102230
Iglesia ni Cristo	102385
Ismaélienne	105053
Jaïnisme	105060
Jedi	211001
Juif de messie/Chrétien juif	102387
Juive	104000
La science de l'esprit/La science religieuse	201102
Longue maison	201004
Luthérienne	102260
Luthérienne allemand	102261
Luthérienne Laestadian	102262
Manitou	201005
Maronite	101016
Melkite	101017
Mennonite	102272
Mennonite reinlander	102274
Mennonite sommerfeld	102275
Méthodiste	102280
Méthodiste libre	102220
Midewin	201006
Mission de l'Esprit Saint	102310
Mormone	102480
Mouvement raelian	211002
Mouvement rastafari	207000
Musulmane	105050
Nouvel Âge	204000
Orthodoxe	103000
Orthodoxe arménienne	103020
Orthodoxe bulgare	103051
Orthodoxe chrétienne	103052
Orthodoxe copte	103030

Orthodoxe éthiopienne	103053
Orthodoxe grecque	103040
Orthodoxe juive	104001
Orthodoxe macédonienne	103054
Orthodoxe roumaine	103060
Orthodoxe russe	103070
Orthodoxe serbe	103080
Orthodoxe ukrainienne	103090
Païenne	206000
Pentecôtiste	102400
Pieds-Noirs	201001
Presbytérienne	102430
Presbytérienne réformée	102435
Protestante	102000
Quaker	102450
Réformée	102436
Religion orientales	105000
Renouveau charismatique	102080
Salle d'évangile	102406
Satanisme	208000
Scientologie	209000
Shaker	102392
Shintoïsme	105080
Sikh	105090
Spiritualiste	102500
Spiritualité autochtone	201000
Standard Church	102510
Sunnite	105055
Swedenborgian (Nouvelle église)	102350
Taoïste	105100
Témoin de Jéhovah	102250
Unitarienne	102520
Unité - Nouvelle Pensée - Panthéiste	102511
Vineyard Christian Fellowship	102393
Wesleyenne	102540
Wicca	206003
Worldwide Church of God	102550
Zoroastrienne	105071

Annexe D – Codes de sport

Sport	Code
un sport adapté ou parasport	1000
le raid aventure	1005
le tir à l'arc	1010
les jeux d'hiver de l'Arctique	1015
le badminton	1020
le hockey-balle	1025
le baseball	1030
le basketball	1035
le BMX	1040
le bobsleigh	1045
le jeu des cinq-quilles	1050
le jeu des dix-quilles	1055
la boxe	1060
le ballon-balai	1065
le canoë-kayak	1070
meneur(se) de claques	1075
l'épreuve combinée d'athlétisme (triathlon, pentathlon, décathlon, etc.)	1080
l'épreuve combinée de sports d'hiver (biathlon, combiné nordique, combiné alpin)	1085
l'haltérophilie	1090
le cricket	1095
le ski de fond	1100
le curling	1105
le cyclisme	1110
le plongeon	1115
le ski alpin/ de descente	1120
un sport équestre	1125
l'escrime	1130
le hockey sur gazon	1135
le patinage artistique	1140
le football	1145
le ski acrobatique	1150
le golf	1155
la gymnastique	1160
le handball (par équipe, quatre murs)	1165
le hockey sur glace	1170
le hockey à patin à roues alignées	1175
le patin à roues alignées	1180
le kickboxing et arts martiaux mixtes	1185
la crosse	1190
les boules sur pelouse pas la pétanque	1195

la luge	1200
marathon	1205
les arts martiaux (karaté, jiu-jitsu, judo, aikido, tae kwon do, kung fu, capoeira, etc.)	1210
le vélo de montagne	1215
la planche tout-terrain	1220
le netball	1225
la marche athlétique	1230
le racquetball	1235
la ringuette	1240
l'aviron	1245
le rugby	1250
la course sur route	1255
la voile/le bateau	1260
le tir	1265
la planche à roulettes	1270
le skeleton	1275
le saut à ski	1280
la planche à neige	1285
la raquette	1290
le soccer	1295
le softball	1300
le patinage de vitesse	1305
le squash	1310
la natation	1315
la nage synchronisée	1320
le tennis de table/ping pong	1325
le tennis	1330
l'athlétisme	1335
le frisbee extrême	1340
le volleyball	1345
le water-polo	1350
le ski nautique/planche nautique	1355
la planche à voile/véloplanchisme	1360
la lutte	1365

Annexe E – Codes de principal domaine d'études

Principal domaine d'études	Codes
Aucun certificat postsecondaires, diplôme ou grade	01
Éducation	02
Arts visuels et d'interprétation, et technologie des communications	03

Sciences humaines	04
Sciences sociales et de comportements, et droit	05
Commerce, gestion et administration publique	06
Sciences physiques et de la vie, et technologies	07
Mathématiques, informatique et sciences de l'information	08
Architecture, génie et services connexes	09
Agriculture, ressources naturelles et conservation	10
Santé et domaines connexes	11
Services personnels, de protection et de transport	12
Autres	13

Annexe F – Conseils sur l'utilisation des poids bootstrap standards de l'ESG

Une variable du poids d'enquête ayant une série correspondante de 500 variables de poids bootstrap standards⁵ accompagne de nombreux fichiers de microdonnées de l'ESG afin qu'une approche entièrement fondée sur le plan d'enquête puisse être adoptée pour l'analyse des données.

L'approche de l'analyse fondée sur le plan d'enquête suppose d'abord l'utilisation de la variable du poids d'enquête en vue d'obtenir les estimations pondérées des quantités d'intérêt. On utilise ensuite des renseignements additionnels sur le plan d'enquête pour estimer les variances⁶ (et covariances) des quantités estimées. Dans le cas des nombreux fichiers de microdonnées de l'ESG, ces renseignements additionnels comprennent 500 variables de poids bootstrap de l'enquête. On peut alors utiliser les estimations et estimations des variances fondées sur le plan d'enquête pour tirer les conclusions nécessaires dans le cadre de l'analyse.

On peut décrire brièvement la création d'une estimation de variance de poids bootstrap de la façon suivante :

Supposons que $\hat{\beta}$ soit l'estimation pondérée de la quantité d'intérêt, β , calculée au moyen de la variable du poids de l'enquête W , et que $\hat{\beta}^{(b)}$ soit une estimation obtenue exactement de la même façon, à l'exception de la substitution de la variable de poids d'enquête W par la variable de poids bootstrap $w^{(b)}$, où $b=1,2,\dots,500$. Cela donne les estimations bootstrap $\hat{\beta}^{(1)}, \dots, \hat{\beta}^{(500)}$.

L'estimation bootstrap de la variance de $\hat{\beta}$ est donc

$$\hat{V}_B(\hat{\beta}) = \frac{1}{500} \sum_{b=1}^{500} (\hat{\beta}^{(b)} - \hat{\beta})^2. \quad (1)$$

⁵Contrairement aux années précédentes de l'ESG, nous utilisons maintenant la méthode des poids bootstrap standard. Une attention particulière devrait être portée à la formule (1) puisqu'elle est différente de la formule avec la méthode des poids bootstrap moyens.

⁶ La variance qui est estimée dans une approche fondée sur le plan d'enquête est la variabilité d'une estimation en raison du rééchantillonnage effectué exactement selon le même plan à partir de la même population limitée.

Si $\hat{\beta}$ est un vecteur plutôt qu'une valeur unique, comme si $\hat{\beta}$ était l'ensemble des coefficients d'un modèle, alors la matrice des estimations des variances et covariances des éléments de $\hat{\beta}$ est $\hat{V}_B(\hat{\beta}) = \frac{1}{500} \sum_{b=1}^{500} (\hat{\beta}^{(b)} - \hat{\beta})(\hat{\beta}^{(b)} - \hat{\beta})'$. (La valeur « 500 » figurant dans la formule provient du fait que nous avons 500 poids bootstrap différents. Si le nombre d'échantillons bootstrap venait à changer, la valeur « 500 » figurant dans la formule (1) devraient être modifiée.)

La méthode bootstrap n'est qu'une méthode de répétition parmi d'autres que l'on peut utiliser pour obtenir des estimations de variances fondées sur le plan d'enquête à partir des données d'enquête. Même si plusieurs progiciels commerciaux d'analyse fondée sur le plan d'enquête offrent des méthodes de répétition pour l'estimation de variances, ils ne désignent généralement pas la méthode bootstrap comme faisant partie de leurs méthodes. Afin d'obtenir des estimations de variances avec les poids bootstrap, on peut utiliser des programmes pouvant effectuer des estimations de variances au moyen de la méthode de réplication particulière appelée BRR (répliques répétées équilibrées) avec les poids de répétition fournis par l'utilisateur⁷. En effet, la forme de l'estimation des variances avec le bootstrap est similaire à celle avec la méthode BRR ce qui en fait un substitut adéquat. Ainsi, dans ces logiciels, les 500 poids bootstrap fournis dans les fichiers microdonnées de l'ESG doivent être désignés comme 500 poids BRR.

Dans les sections ci-dessous, des directives seront données sur la mise en œuvre de l'estimation de la variance des poids bootstrap à partir des microdonnées de l'ESG, au moyen de trois progiciels commerciaux différents qui peuvent effectuer certaines analyses fondées sur le plan d'enquête pour le BRR: Stata 9 ou 10, SUDAAN et WesVar. Dans tous les cycles de l'ESG où des poids bootstrap sont fournis, les noms donnés à ces variables bootstrap dans la documentation de l'utilisateur sont wtbs_001 à wtbs_500⁸. Le nom de la variable du poids d'enquête est normalement wght_per.

Stata 9 ou 10

À partir de la version 9, le progiciel commercial Stata contient des méthodes de répétition additionnelles permettant d'effectuer des estimations de variances fondées sur le plan d'enquête dans ses commandes d'analyse d'enquêtes. La méthode BRR fait partie des méthodes de répétition offertes; c'est cette méthode qui serait précisée dans l'analyse des données de l'ESG.

Pour préciser cette méthode, voici la marche à suivre recommandée :

1. Avant d'utiliser les commandes d'analyse d'enquêtes, utiliser l'énoncé « svyset » pour déclarer que les données sont des données d'enquête, pour désigner les variables qui contiennent des renseignements sur le plan d'enquête et pour préciser la méthode d'estimation des variances. Les paramètres établis par l'énoncé « svyset » sont sauvegardés avec un ensemble de données lorsque (ou si) un ensemble de données est sauvegardé. La forme de l'énoncé « svyset » à utiliser avec l'ensemble de données d'une analyse de l'ESG serait la suivante :

svyset [pweight=wght_per], vce(brr) fay(0) brrweight(wtbs_001-wtbs_500) mse

L'énoncé **pweight=wght_per** indique à Stata que le poids d'enquête (souvent appelé le poids de probabilité) est la variable wght_per.

L'option **vce(brr)** déclare que la méthode d'estimation de la variance à utiliser est BRR.

⁷ Pour obtenir une description plus détaillée, voir Phillips (2004).

⁸ Veuillez noter que pour les cycles précédents de l'ESG (cycle 26 et antérieur), les variables wtbs_001 à wtbs_500 avaient été créées à partir de la méthode du bootstrap moyen. À partir du cycle 27 de l'ESG (2013), les variables wtbs_001 à wtbs_500 sont créées à partir de la méthode du bootstrap standard.

L'option **fay(0)** déclare qu'il n'est pas nécessaire d'utiliser un ajustement de Fay avec la méthode d'estimation de la variance BRR. Cet ajustement est seulement nécessaire lorsque la méthode des poids bootstrap moyens est utilisée, ce qui n'est plus le cas depuis le cycle 27 de l'ESG avec l'utilisation des poids bootstrap standards.

L'option **brrweight(wtbs_001-wtbs_500)** indique que le nom des variables de poids BRR sont **wtbs_001, wtbs_002, ..., wtbs_500**. Cette option peut aussi être appelée **brrweight(wtbs_*)** à condition qu'il n'y ait pas de variable autre que les variables du poids bootstrap dont le nom commence par « wtbs_ ».

Enfin, l'option **mse** indique à Stata de calculer la variance au moyen des différences au carré entre les estimations bootstrap et l'estimation de l'échantillon complet des variables, comme le montre l'équation (1). Si cette option n'est pas incluse, Stata utilise les différences au carré entre chaque estimation bootstrap et la moyenne de toutes les estimations bootstrap. Les deux méthodes devraient produire à peu près le même résultat.

2. Stata contient une liste complète de commandes d'analyse d'enquêtes qui utilise une méthode fondée sur le plan d'enquête dans ses calculs. Ces commandes, décrites dans les documents Stata, sont mises en œuvre par l'utilisation du préfixe « svy » associé au nom des autres estimateurs. Par exemple, **svy: mean** est la commande servant à estimer les moyennes et les estimations de variabilité de la population et des sous-populations grâce à une méthode fondée sur le plan d'enquête. Lorsque l'énoncé **svyset** précède toutes les commandes de l'enquête, il n'est pas nécessaire que celles-ci contiennent des renseignements sur le plan d'enquête adopté. Il convient de noter que la plupart des commandes qui permettent le préfixe « svy » correspondent également au nom des commandes relatives aux données autres que les données d'enquête, à ce qui est estimé, aux options offertes et à ce qui peut être fait par le changement apporté après l'estimation lorsque le préfixe « svy » est ajouté.

SUDAAN

SUDAAN est un progiciel commercial expressément élaboré par le Research Triangle Institute pour l'analyse de données tirées d'enquêtes-échantillons complexes et d'autres études par observation et expérimentales comprenant des données corrélées par grappes. La version exécutable SAS du logiciel est particulièrement utile aux personnes qui connaissent le SAS.

La spécification de la méthode d'estimation des variances à utiliser par SUDAAN est apportée dans la déclaration de procédure pour une procédure particulière. D'autres déclarations sur les plans d'échantillonnage fournissent des renseignements supplémentaires requis par le programme. En particulier, pour utiliser les poids bootstrap des données de l'ESG, il faut suivre les étapes suivantes :

- préciser **DESIGN=BRR** dans la déclaration de procédure;
- inclure l'énoncé **WEIGHT** suivant pour définir la variable du poids d'enquête :

WEIGHT wght_per;

- inclure l'énoncé **REPWGT** pour indiquer le nom des variables de poids bootstrap sur le fichier de données. Plus particulièrement, pour les fichiers de microdonnées de l'ESG, cet énoncé **REPWGT** aurait la forme suivante :

REPWGT wtbs_001-wtbs_500;

WesVar

WesVar est un progiciel produit par Westat qui effectue diverses analyses de données d'enquête en se fondant exclusivement sur des méthodes de répétition pour l'estimation des variances. L'une des méthodes offertes est la méthode BRR. Largement tirée de Phillips (2004), dans WesVar, la méthode d'estimation des variances est mentionnée lors de la création d'un nouveau fichier de données WesVar.

Le fichier qui en résulte sert à définir des classeurs lorsque l'on procède à des demandes de tableaux et de régressions. Pour définir un fichier de données WesVar avec un poids bootstrap :

- déplacer les variables de poids répétées (c.-à-d. wtbs-001 à wtbs_500) dans la boîte *Replicates (Répétitions)*;
- déplacer la variable du poids d'enquête (c.-à-d. wght_per) dans la boîte *Full sample (Échantillon complet)*;
- pour le bootstrap, inscrire BRR pour la méthode (*Method*);
- déplacer les variables de l'analyse dans la boîte *Variables (Variables)*, un identificateur unique de la boîte d'identité (facultatif) et sauvegarder le fichier.

Références

PHILLIPS, Owen, *Comment utiliser les poids « bootstrap » avec Wes Var et SUDAAN*, 2004. Le Bulletin technique et d'information des Centres de données de recherche. (Automne) 1(2):1-10. Statistique Canada n° 12-002-XIF au catalogue. <http://www.statcan.ca/bsolc/francais/bsolc?catno=12-002-X20040027032>.