

Guide de l'utilisateur des microdonnées

Enquête canadienne sur les capacités financières

2009



Statistics
Canada

Statistique
Canada

Canada

Table des matières

1.0	Introduction	5
2.0	Contexte	7
3.0	Objectifs	9
4.0	Concepts et définitions	11
5.0	Méthodologie de l'enquête	15
5.1	Population visée.....	15
5.2	Stratification	15
5.3	Plan d'échantillonnage et répartition de l'échantillon.....	15
5.4	Tirage de l'échantillon	16
5.5	Taille de l'échantillon.....	16
5.6	Structure du questionnaire.....	16
6.0	Collecte des données	19
6.1	Conception du questionnaire	19
6.2	Interview.....	19
6.3	Supervision et contrôle de qualité.....	19
7.0	Traitement des données	21
7.1	Saisie des données.....	21
7.2	Vérification	21
7.3	Codage des questions ouvertes	22
7.4	Imputation	22
7.5	Création de variables dérivées	22
7.6	Pondération.....	22
7.7	Suppression de renseignements confidentiels	23
8.0	Qualité des données	25
8.1	Taux de réponse	25
8.2	Erreurs relatives à l'enquête	25
8.2.1	Collecte des données	26
8.2.2	Traitement des données	26
8.2.3	Non-réponse	26
8.2.4	Mesure de l'erreur d'échantillonnage.....	27
9.0	Lignes directrices pour la totalisation, l'analyse et la diffusion de données	29
9.1	Lignes directrices pour l'arrondissement d'estimations	29
9.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation	30
9.3	Définitions de types d'estimations : catégoriques et quantitatives	30
9.3.1	Estimations catégoriques.....	30
9.3.2	Estimations quantitatives	30
9.3.3	Totalisation d'estimations catégoriques.....	31
9.3.4	Totalisation d'estimations quantitatives	31
9.4	Lignes directrices pour l'analyse statistique	32
9.5	Lignes directrices pour la diffusion de coefficients de variation.....	32
9.6	Seuils pour la diffusion des estimations.....	34

10.0	Tables de variabilité d'échantillonnage approximative	35
10.1	Comment utiliser les tables de coefficients de variation pour des estimations catégoriques.....	36
10.1.1	Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques	37
10.2	Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance	44
10.2.1	Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance	45
10.3	Comment utiliser les tables de coefficients de variation pour effectuer un test t.....	45
10.3.1	Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t	45
10.4	Coefficients de variation pour des estimations quantitatives	46
10.5	Tables de coefficients de variation.....	46
10.6	Méthode bootstrap moyenne pour estimer la variance.....	46
10.7	Progiciels statistiques pour estimer la variance	47
11.0	Ponderation	49
11.1	Procédures de pondération.....	49
12.0	Questionnaires	53
13.0	Cliché d'enregistrement à valeurs univariées	55

1.0 Introduction

Le présent document a pour but de faciliter la consultation et la manipulation du fichier de microdonnées de l'Enquête canadienne sur les capacités financières, qui a été menée par Statistique Canada en 2009. L'enquête s'est tenue avec la collaboration et le soutien de Ressources humaines et Développement des compétences Canada, Ministère des Finances Canada et l'Agence de la consommation en matière financière du Canada. Ce guide comprend des renseignements sur les objectifs, la méthodologie et les méthodes d'estimation, ainsi que sur les règles de diffusion des estimations fondées sur les données de l'enquête. Il comprend en outre le questionnaire ainsi que des tables de variance approximative accompagnées d'exemples de la façon de les utiliser.

Toutes les questions concernant l'ensemble de données ou son utilisation devraient être adressées à :

Statistique Canada

Services à la clientèle

Division des enquêtes spéciales

N° de Téléphone : 613-951-3321 ou appelez sans frais : 1-800-461-9050

N° de télécopieur : 613-951-4527

Courriel : des@statcan.gc.ca

2.0 Contexte

Il s'agit de la première Enquête canadienne sur les capacités financières (ECCF). Cette enquête est devenue nécessaire du fait de l'évolution de la situation économique, de la variété et de la complexité des produits financiers disponibles et de la nécessité d'établir des données de base. L'ECCF a été tenue entre février et mai 2009, avec comme objectif d'élaborer une base de données statistiques comportant des estimations au sujet des capacités financières. Le but visé par l'enquête est de recueillir des données qui montreront le degré de connaissances des Canadiens en matière de décisions financières. De façon plus particulière, l'enquête jettera de la lumière sur les connaissances, les capacités et les comportements des Canadiens en ce qui a trait à la prise de décisions financières; autrement dit, la façon dont les Canadiens comprennent leur situation financière, les services financiers qui sont mis à leur disposition et leurs plans pour l'avenir. L'enquête est conçue pour recueillir des données concernant les approches des répondants en ce qui a trait à la gestion de l'argent au quotidien et à l'établissement d'un budget, à la gestion de l'argent à plus long terme et à la planification financière en général. Les données de l'enquête ont été recueillies auprès de Canadiens de 18 ans et plus dans les dix provinces.

3.0 Objectifs

L'objectif fondamental de l' Enquête canadienne sur les capacités financières (ECCF) est de mieux comprendre les connaissances, les préférences et les besoins financiers des Canadiens. De façon plus particulière, l'enquête permettra de recueillir des données sur les connaissances et la compréhension des Canadiens en matière financière, leurs capacités financières (capacité d'appliquer les connaissances et de prendre des décisions financières) et leurs responsabilités financières (comportement dans le domaine des finances).

Les renseignements obtenus à partir de l'ECCF aideront les gouvernements et l'industrie à mieux comprendre les connaissances et les comportements des Canadiens en ce qui a trait à leur participation aux marchés de services financiers et aux divers programmes gouvernementaux conçus pour faciliter la planification financière pour les études et la retraite, par exemple, les régimes enregistrés d'épargne-études (REEE), les régimes enregistrés d'épargne-retraite (REER) et les programmes comme le Supplément de revenu garanti (SRG) et le Régime de pensions du Canada (RPC) ou la Régie des rentes du Québec (RRQ).

4.0 Concepts et définitions

Le présent chapitre énonce les concepts et les définitions qui intéressent les utilisateurs des données de l'Enquête canadienne sur les capacités financières. Les utilisateurs doivent se reporter au chapitre 12.0 du présent document pour une copie du questionnaire de l'enquête.

Actifs – Bien ayant une valeur monétaire qui est possédé par une personne (ou une entreprise). Les biens immobiliers, les actions, les obligations et l'argent proprement dit sont tous considérés comme des actifs.

Actions – Actions ordinaires et privilégiées des sociétés. Termes connexes : actions émises dans le public, actions ordinaires, actions privilégiées, actions.

Caisse de crédit – Institution financière coopérative qui appartient à ses membres et qui fonctionne pour le bénéfice de ses membres en acceptant des dépôts et en faisant des prêts, y compris des prêts hypothécaires, et en fournissant d'autres services, comme des services de chèques et de cartes de crédit.

Carte bancaire – Carte délivrée par une banque qui permet aux détenteurs d'effectuer des paiements électroniques à partir d'un terminal de point de vente et d'effectuer des transactions bancaires à partir d'un guichet automatique.

Carte de crédit – Méthode de paiement de biens et de services qui permet à l'acheteur de reporter le paiement et de rembourser le principal et les intérêts en versements au fil du temps. Cela comprend les cartes de crédit de type bancaire (p. ex., VISA, Mastercard), ainsi que les cartes de crédit de magasins et de stations-service.

Carte de débit – voir Carte bancaire ci-dessus

Certificat de placement garanti (CPG) – Outil d'épargne dont la durée va généralement d'un à cinq ans, période pendant laquelle le taux d'intérêt est garanti et l'argent est habituellement bloqué jusqu'à l'échéance.

Chômage – Les personnes en chômage sont celles qui, pendant la semaine de référence :

- a) avaient été mises à pied temporairement, mais s'attendaient à être rappelées au travail et étaient disponibles pour travailler; ou
- b) étaient sans emploi, avaient activement cherché un emploi au cours des quatre dernières semaines et étaient disponibles pour travailler¹; ou
- c) devaient commencer un nouvel emploi dans quatre semaines ou moins à compter de la semaine de référence et étaient disponibles pour travailler.

Compte d'épargne libre d'impôt ou régime d'épargne libre d'impôt (CELI) – Les résidents du Canada de 18 ans et plus peuvent y verser 5 000 \$ par année sans que les revenus de placement ou les gains en capital soient imposés.

Coopératives / coop – voir Caisse de crédit ci-dessus

¹ Les personnes sont considérées comme disponibles pour travailler si elles :

- i) ont déclaré qu'elles auraient pu travailler si on leur avait offert un emploi convenable; ou encore si la raison pour laquelle elles ne pouvaient occuper un emploi était de nature temporaire, notamment du fait d'une maladie ou d'une incapacité, d'obligations personnelles ou familiales, parce qu'elles avaient déjà un emploi devant bientôt débiter, ou qu'elles étaient en vacances (avant 1997, les personnes en vacances n'étaient pas considérées comme étant disponibles);
- ii) étaient des étudiants à temps plein à la recherche de travail à temps partiel qui respectait également la condition i) ci-dessus. Les étudiants à temps plein qui sont actuellement aux études et qui cherchent un travail à temps plein ne sont pas considérés comme étant disponibles au cours de la semaine de référence.

Découvert bancaire – Montant de dépassement d'un chèque ou d'autres paiements des fonds déposés.

Dépôts à terme – Instrument de dépôt habituellement disponible auprès des sociétés de fiducie et des banques à charte, qui nécessite un investissement minimum à un taux prédéterminé d'intérêt pour une durée déterminée. Les taux d'intérêt varient selon le montant investi et l'échéance.

Dettes – Montant dû par un parti à un autre parti pour de l'argent, des biens ou des services reçus.

Dividende – Montant versé aux actionnaires d'une compagnie à partir des bénéfices réalisés par cette compagnie.

Emploi – Les personnes occupées sont celles qui :

- a) ont fait un travail quelconque² dans le cadre d'un emploi ou dans une entreprise; ou
- b) avaient un emploi, mais n'étaient pas au travail à cause d'une maladie ou d'une incapacité, pour obligations personnelles ou familiales, pour des vacances, à la suite d'un conflit de travail ou du fait de tout autre facteur (cela n'inclut pas les personnes mises à pied, celles qui étaient inactives entre deux emplois occasionnels et celles qui avaient un emploi devant commencer à une date ultérieure).

Emploi à temps plein – L'emploi à temps plein comprend les personnes qui travaillent habituellement 30 heures ou plus par semaine à leur emploi principal ou à leur unique emploi.

Emploi à temps partiel – L'emploi à temps partiel tient compte des personnes qui travaillent habituellement moins de 30 heures par semaine à leur emploi principal ou à leur unique emploi.

Famille – Groupe de deux personnes ou plus qui vivent dans le même logement et qui sont apparentées par le sang, par alliance ou par adoption ou qui vivent en union libre.

Fiducie – Une fiducie est un arrangement grâce auquel le droit de propriété est détenu par une partie, le « fiduciaire » (ou gestionnaire), au profit d'une autre partie, le « bénéficiaire ».

Fonds communs de placement – Un ensemble de valeurs mobilières qui sont achetées par une compagnie d'investissement et vendues comme groupe ou fonds particuliers. Les investisseurs achètent des unités de ces fonds.

Fonds de revenu de retraite (FRR) – Fond auquel des sommes d'un REER peuvent être transférées. Les versements d'un FRR peuvent varier, mais un montant minimum doit être retiré chaque année.

Hypothèque – Prêt pour lequel une résidence ou un autre bien immobilier sert de garantie.

Inactivité – Les personnes inactives sont celles qui, considérant les conditions qui existaient sur le marché du travail dans leur région, ne désiraient ou ne pouvaient offrir ou fournir leurs services au cours de la semaine de référence

Intérêt – Paiement versé à un taux déterminé pour l'utilisation d'argent emprunté.

² Le travail comprend tout travail contre rémunération ou bénéfice, c'est-à-dire le travail rémunéré dans le contexte d'un rapport employeur-employé ou le travail autonome. Cela comprend aussi le travail familial non rémunéré, c'est-à-dire un travail non rémunéré qui contribue directement à l'exploitation d'une ferme, d'une entreprise ou d'un cabinet de professionnels appartenant à un membre apparenté du même ménage et exploité par lui. Ces activités comprennent la tenue de livres, la vente de produits, le service aux tables, etc. Les tâches comme les tâches ménagères ou l'entretien de la maison ne sont pas considérées comme du travail familial non rémunéré.

Ligne de crédit – Entente formelle entre un emprunteur et un prêteur (habituellement une institution financière) qui permet à l'emprunteur d'obtenir autant ou aussi peu d'argent qu'il souhaite jusqu'à concurrence d'un maximum déterminé au préalable (ou limite de crédit). Aux fins de l'enquête, le montant à déclarer est le montant à rembourser à la ligne de crédit.

Ménage – Un ménage comprend une personne ou un groupe de personnes (apparentées ou non) qui occupent un logement et qui n'ont pas de lieu de résidence habituel ailleurs.

Obligations – Titre de créance délivré par un gouvernement ou une société. Les taux d'intérêt sont fixes pour la durée de l'obligation, mais celle-ci peut être vendue à une valeur inférieure ou supérieure à sa valeur nominale.

Personne la mieux renseignée (PMR) – La PMR est la personne du ménage qui connaît le plus la situation financière. On a demandé aux répondants de s'identifier en ce qui a trait à la PRM dans deux sections du questionnaire – Dépenses quotidiennes et Gestion financière.

Profession – L'Enquête canadienne sur les capacités financières fournit des données concernant la profession à laquelle appartiennent les personnes occupées et en chômage, ainsi que les personnes inactives qui ont occupé un emploi au cours des 12 derniers mois. Ces codes reposent sur la Classification nationale des professions - statistiques (CNP-S) de 2006.

Régime d'accession à la propriété (RAP) – Il s'agit d'un régime parrainé par le gouvernement qui permet aux personnes de retirer jusqu'à 20 000 \$ libres d'impôt de leur Régime enregistré d'épargne-retraite (REER) pour l'achat d'une maison.

Régime enregistré d'épargne-études (REEE) – Outil d'épargne conçu pour permettre à des personnes d'accumuler des revenus pour les études postsecondaires. Généralement, ce sont les parents qui tentent d'économiser pour les études postsecondaires de leurs enfants qui souscrivent aux régimes. Il peut être avantageux d'investir dans un REEE, parce que le gouvernement fédéral verse une contribution et que le revenu produit est libre d'impôt jusqu'à ce qu'il soit retiré pour les études postsecondaires de l'enfant.

Régime enregistré d'épargne-invalidité (REEI) – Régime qui permet d'investir des fonds à l'abri de l'impôt jusqu'à leur retrait. Il vise à aider les parents et d'autres personnes à économiser pour la sécurité financière à long terme d'un enfant handicapé.

Régime enregistré d'épargne-retraite (RER ou REER) – Régime d'accumulation de capital conçu pour encourager l'épargne en vue de la retraite. Les cotisations sont déductibles d'impôt, selon des limites déterminées. Les revenus d'investissement dans un RER sont libres d'impôt, mais les prestations sont imposables. Les montants accumulés dans ces régimes comprennent les montants investis à l'origine plus l'intérêt / les gains accumulés.

Régime de pensions du Canada (RPC) / Régie de rentes du Québec (RRQ) – Prestations de retraite versées à l'âge de 65 ans aux personnes qui ont travaillé au Canada. Comprend aussi les prestations de survivant, comme les pensions de veufs ou veuves, les prestations d'orphelin et les prestations d'invalidité pour les retraités handicapés.

Sécurité de la vieillesse (SV) – Prestation mensuelle mise à la disposition de la plupart des Canadiens de 65 ans et plus qui ont vécu au Canada pendant au moins dix ans.

Situation vis-à-vis l'activité – Désigne la situation du répondant sur le marché du travail : un membre de la population active âgé de 15 ans ou plus (à l'exclusion des pensionnaires d'établissements) est soit occupé, soit en chômage, soit inactif.

Société de fiducie – Institution financière qui fournit des services financiers et de fiducie à des personnes et des sociétés. Une part importante des activités des sociétés de fiducie consistent à agir comme fiduciaires pour d'autres sociétés relativement aux régimes de retraite, à l'émission d'obligations,

etc. Il s'agit d'intermédiaires financiers actifs, qui acceptent des dépôts et qui font des prêts de divers types.

Valeur nette (Equity) – Intérêt résiduel des actifs après déduction des passifs connexes. Par exemple, la valeur nette d'une maison est égale à la valeur de la maison moins la dette hypothécaire.

Valeur nette (Net Value) – Valeur qui résulte de la déduction des coûts de fonctionnement du prix.

5.0 Méthodologie de l'enquête

L'Enquête canadienne sur les capacités financières (ECCF) a été tenue comme enquête ponctuelle entre le 11 février et le 9 mai 2009, sous forme d'enquête par composition aléatoire (CA), une technique permettant de générer des numéros de téléphone au hasard par ordinateur. Les interviews ont été menées par téléphone.

5.1 Population visée

La population cible de l'ECCF était constituée de toutes les personnes de 18 ans et plus vivant au Canada, sauf pour les deux exceptions suivantes :

- 1) les résidents du Yukon, des Territoires du Nord-Ouest et du Nunavut, et
- 2) les pensionnaires à plein temps des établissements.

Étant donné que l'enquête a été menée à partir d'un échantillon de numéros de téléphone, les ménages (et par conséquent les personnes vivant dans ces ménages) qui n'ont pas de ligne téléphonique terrestre ont été exclus de la population de l'échantillon. Cela signifie que les personnes sans téléphone et les personnes possédant uniquement un téléphone cellulaire ont été exclues. Les personnes sans ligne téléphonique terrestre représentent environ 8 % de la population cible. Toutefois, les estimations de l'enquête ont été pondérées pour les inclure.

5.2 Stratification

Afin d'assurer la représentation dans l'échantillon des personnes vivant dans toutes les régions du Canada, chacune des dix provinces a été divisée en strates ou régions géographiques. Les régions métropolitaines de recensement (RMR) sont des régions définies à partir du recensement de la population et correspondent approximativement aux villes comptant 100 000 habitants ou plus. De nombreuses RMR ont été considérées comme des strates distinctes. C'est le cas pour St. John's, Halifax, Saint John, Montréal, ville de Québec, Toronto, Ottawa, Hamilton, Winnipeg, Regina, Saskatoon, Calgary, Edmonton et Vancouver. Les RMR qui restent en Ontario, au Québec et en Colombie-Britannique ont été combinées en deux strates distinctes. Généralement, à l'intérieur de chaque province, une strate autre qu'une RMR a été créée, sauf à l'Île-du-Prince-Édouard, où il n'y avait qu'une strate pour l'ensemble de la province. Cela a donné lieu à un plan de sondage comptant 27 strates au total.

5.3 Plan d'échantillonnage et répartition de l'échantillon

Le plan d'échantillonnage correspond à un échantillon aléatoire stratifié à deux phases de numéros de téléphones. À la première phase, les ménages sont sélectionnés au moyen de la CA. À la deuxième phase, une personne du ménage contacté est sélectionnée.

Étant donné que l'enquête vise principalement à produire des estimations fiables au niveau national, mais tente aussi d'établir des estimations de qualité raisonnable aux niveaux provincial et de la RMR, on a eu recours à une répartition de Kish. Du fait de ce compromis, l'échantillon des provinces plus importantes compte un plus grand nombre de répondants, mais le nombre n'est pas strictement proportionnel à la population de chacune.

La taille de l'échantillon initial des numéros de téléphone dépendait du taux de réponse prévu et du taux de succès attendu de la CA (proportion des numéros de téléphone échantillonnés qui sont intégrés à l'échantillon comme ménages). On a estimé qu'au total plus de 53 000 numéros de téléphone étaient nécessaires pour obtenir 20 000 répondants. Cela repose sur une hypothèse de taux de réponse et de succès de 66 %, qui variait substantiellement selon la province, avec une moyenne globale prévue d'environ 40 %.

5.4 Tirage de l'échantillon

L'échantillon de l'ECCF a été produit au moyen d'une technique perfectionnée d'échantillonnage par CA appelée élimination des banques non valides (ÉBNV). Dans chaque combinaison de provinces et de strates, une liste des banques valides (indicatif régional et cinq chiffres suivants) a été dressée à partir des fichiers administratifs des compagnies de téléphone. Pour les fins des enquêtes sociales, une banque valide se définit comme une banque renfermant au moins un numéro de téléphone résidentiel valide. Par conséquent, toutes les banques qui renferment uniquement des numéros de téléphone non assignés, non valides ou d'affaires sont exclues de la base de sondage.

Puis, un échantillon systématique des banques (avec remplacement) a été sélectionné dans chacune des strates. Pour chaque banque sélectionnée, un numéro à deux chiffres (de 00 à 99) a été généré aléatoirement. Ce numéro aléatoire a été ajouté à la banque pour former un numéro de téléphone complet. Cette méthode a permis de sélectionner des numéros de téléphone résidentiels inscrits et non inscrits, des numéros d'affaires et des numéros non valides (qui ne sont pas ou qui n'ont jamais été en service). Une procédure de filtrage visant à éliminer les numéros hors service et d'affaires connus a été effectuée avant l'envoi de l'échantillon à l'équipe chargée de l'interview téléphonique assistée par ordinateur (ITAO).

On a composé tous les numéros de téléphone figurant dans l'échantillon de l'ITAO pour déterminer s'ils étaient rattachés à un ménage. Le cas échéant, on a demandé à la personne ayant pris l'appel de fournir des renseignements sur les différents membres du ménage. Les âges des membres du ménage ont été utilisés pour déterminer qui dans le ménage serait sélectionné pour l'interview. Les personnes ainsi sélectionnées ont été interviewées dans la langue officielle de leur choix et les interviews par procuration n'ont pas été permises.

5.5 Taille de l'échantillon

Le tableau suivant montre le nombre de ménages dans l'échantillon de l'ECCF.

Province	Taille de l'échantillon
Terre-Neuve-et-Labrador	1 224
Île-du-Prince-Édouard	572
Nouvelle-Écosse	1 326
Nouveau-Brunswick	1 213
Québec	5 824
Ontario	8 185
Manitoba	1 427
Saskatchewan	1 976
Alberta	2 748
Colombie-Britannique	3 061
Canada	27 555

5.6 Structure du questionnaire

L'enquête a permis de recueillir une quantité importante de données pour chaque répondant sélectionné, ainsi que des renseignements concernant le ménage. On était particulièrement

intéressé à déterminer si le répondant sélectionné se considérait comme la personne la mieux renseignée du ménage concernant les dépenses quotidiennes et la gestion financière. On a eu recours à l'auto-identification pour ces deux sections de l'enquête. Le questionnaire complet se trouve à la chapitre 12.0 du présent document. Les sections du questionnaire figurent ci-après.

Introduction : Formulaire de contrôle et établissement de la liste des membres du ménage

Le formulaire de contrôle sert à guider l'intervieweur au début de l'interview et fournit un tableau pour établir la liste des membres du ménage. À partir de cette liste, un membre admissible est sélectionné de façon aléatoire pour répondre à l'enquête.

Démographie (DM)

Cette section sert à recueillir certaines données démographiques de base.

Emploi (LF)

La section sur l'emploi permet de déterminer le statut d'emploi actuel, si le répondant (et son conjoint / partenaire, le cas échéant) a travaillé au cours des 12 derniers mois, et le type de travail.

Dépenses quotidiennes (OE)

Cette section de l'enquête porte sur les dépenses au jour le jour et les paiements de factures courantes. Les renseignements recueillis dans cette section ont trait à la façon dont les personnes suivent leurs finances et gèrent leur argent.

Gestion financière (FM)

Cette section du questionnaire porte sur la planification financière à plus long terme. Elle comprend des questions concernant les dépenses importantes supérieures à 10 000 \$, la planification de la retraite et la planification des études postsecondaires des enfants.

Dépenses importantes (ME)

Cette section du questionnaire demande comment les répondants planifient pour des dépenses importantes telles que l'achat d'une maison, d'une voiture, d'un chalet ou pour la noce d'un enfant.

Financement des études postsecondaires (EF)

Cette section porte sur la planification financière dans l'éventualité qu'un enfant poursuive des études postsecondaires telles qu'au le collège (collège d'enseignement général et professionnel (CEGEP) au Québec), à l'université ou à une école de formation professionnelle ou technique.

Planification pour la retraite (RP)

Cette section comporte des questions sur la planification pour la retraite.

Actifs et Passifs (AD)

Cette section du questionnaire comporte des questions concernant les actifs et les passifs de la personne ou du ménage, selon le cas. Ces renseignements contribuent à établir le profil de la situation financière du ménage.

Revenu (IN)

Ces renseignements fournissent un contexte concernant le revenu de la personne et, au besoin, du ménage.

Choix financiers (FC)

Cette section de l'enquête porte sur la façon d'aborder les choix financiers et est pertinente pour les questions de planification et de responsabilité.

Évaluation personnelle subjective (SA)

Dans cette section de l'enquête, les répondants fournissent une auto-évaluation de leurs connaissances des questions financières.

Évaluation personnelle objective (OA)

La dernière section du questionnaire comprend un petit jeu questionnaire auquel les répondants doivent répondre.

6.0 Collecte des données

La collecte des données de l'Enquête canadienne sur les capacités financières (ECCF) a été menée entre février 2009 et le début de mai 2009.

6.1 Conception du questionnaire

Dans le cas de l'Enquête canadienne sur les capacités financières, il avait été proposé, dès sa conception, que l'information soit recueillie par une enquête téléphonique; une approche qui s'est avérée une réussite dans d'autres pays avec un sujet similaire. Une première ronde de mise à l'essai cognitif, y compris des interviews personnelles et des discussions de groupe, à travers le Canada au printemps 2007 a confirmé qu'il s'agissait bel et bien de la meilleure façon de procéder.

Avec l'ajout de Ministère des Finances Canada et la Banque du Canada comme partenaires actifs, le contenu a été modifié pour refléter les besoins en données de chacun de nos partenaires. Bien sur, ceci a mené à une deuxième ronde de mise à l'essai cognitif dans quelques villes sélectionnées au printemps 2008. L'application interview téléphonique assisté par ordinateur (ITAO) a été développée et mis à l'essai durant l'été et l'automne 2008.

6.2 Interview

Les intervieweurs de Statistique Canada sont des employés recrutés et formés pour mener des enquêtes auprès des ménages. Les intervieweurs qui ont participé à l'ECCF ont reçu une formation spéciale en prévision de l'enquête.

Les données ont été recueillies à l'aide des techniques de l'interview téléphonique assistée par ordinateur. Un module frontal contient un ensemble de codes de réponse standard, qui s'appliquent à toutes les issues possibles des appels, ainsi que les scénarios correspondants qui sont lus par les intervieweurs. Une approche normalisée a été utilisée pour présenter l'organisme, le nom et le but de l'enquête, les clients de l'enquête, l'utilisation qui sera faite des résultats et la durée de l'interview.

L'application ITAO garantissait l'entrée des seules réponses valides et le bon enchaînement des questions. Des contrôles intégrés à l'application garantissaient la cohérence des réponses, repéraient et corrigeaient les valeurs aberrantes et déterminaient à qui étaient posées les différentes questions. Ainsi, à la fin du processus de collecte, les données étaient déjà passablement « épurées ».

6.3 Supervision et contrôle de qualité

Tous les intervieweurs de Statistique Canada travaillent sous la supervision d'un groupe d'intervieweurs principaux qui ont pour responsabilité de veiller à ce que les intervieweurs connaissent bien les concepts et les procédures de l'enquête et d'assurer aussi un contrôle périodique des intervieweurs et un examen des documents remplis par ces derniers. Les intervieweurs principaux sont, quant à eux, supervisés par les gestionnaires de programme dans chacun des bureaux régionaux de Statistique Canada.

7.0 Traitement des données

Le principal produit de l'Enquête canadienne sur les capacités financières (ECCF) est un fichier de microdonnées « épuré ». Ce chapitre présente un bref résumé des étapes de traitement inhérentes à la production de ce fichier.

7.1 Saisie des données

Les intervieweurs saisissent directement les réponses aux questions de l'enquête au moment de l'interview (interview téléphonique assistée par ordinateur), à l'aide d'une version informatisée du questionnaire. L'emploi d'un questionnaire automatisé réduit les délais et coûts de traitement associés à la saisie des données, aux erreurs de transcription et à la transmission des données. On soumet les données des réponses à un cryptage pour en préserver le caractère confidentiel, puis on les transmet par modem au bureau régional concerné de Statistique Canada. Les données sont par la suite transmises à Ottawa, au moyen d'une ligne protégée, pour y être traitées davantage.

Une partie du contrôle se fait directement au moment de l'interview. Lorsque les renseignements entrés dépassent les fourchettes (trop faibles ou trop élevés) des valeurs attendues, ou qu'ils entrent en contradiction avec des renseignements entrés auparavant, l'intervieweur voit apparaître à l'écran de l'ordinateur des messages lui demandant de modifier les renseignements. Toutefois, pour certaines questions, les intervieweurs ont la possibilité de passer outre aux contrôles (vérification avec avertissement) et de sauter des questions si le répondant ne connaît pas la réponse ou refuse de répondre. Pour cette raison, on soumet les données des réponses à d'autres processus de vérification et d'imputation après réception au bureau central.

7.2 Vérification

Les fichiers de texte électroniques contenant les transmissions quotidiennes de cas terminés sont combinés en un seul fichier d'enquête « brut ». À la fin de la collecte, ce fichier doit contenir un enregistrement pour chaque personne figurant dans l'échantillon. Avant tout autre traitement, on effectue une vérification pour repérer et supprimer d'éventuels enregistrements en double et exclure les enregistrements classés « non-réponse » ou « hors champ ».

Par conséquent, la correction se fait en modifiant les données au niveau de la variable individuelle. La première étape de la vérification consiste à déterminer quels sont les éléments des résultats de l'enquête qui doivent être conservés dans le fichier maître de l'enquête. Par la suite, les caractères invalides sont supprimés et les éléments de données sont dûment formatés. Les champs de texte sont retirés des fichiers principaux et ajoutés à un fichier distinct pour le codage.

Le premier type d'erreur traitée a été les erreurs dans l'enchaînement des questions, des questions qui ne s'appliquaient pas au répondant (et auxquelles on n'aurait donc pas dû répondre) renfermant parfois des réponses. Dans ces cas, un contrôle informatique a éliminé automatiquement les données superflues en suivant l'ordre du questionnaire dicté par les réponses à des questions antérieures et, parfois, subséquentes. Dans le cas des sauts de question fondés sur des questions auxquelles on a répondu, toutes les questions sautées sont codées « Enchaînement valide » (6, 96, 996, etc.). Dans le cas des sauts de question fondés sur des réponses « Ne sait pas » ou « Refus », on attribue à toutes les questions sautées le code « Non déclaré » (9, 99, 999, etc.). Aux éléments vides restants, on attribue une valeur numérique (9, 99, 999, etc., selon la longueur de la variable). Ces codes sont réservés à des fins de traitement et signifient que l'élément était « Non déclaré ».

7.3 Codage des questions ouvertes

Quelques données élémentaires ont été consignées dans le questionnaire par les intervieweurs sous forme de questions ouvertes. Cela se produit généralement lorsqu'un répondant sélectionne la réponse « Autre - Précisez » à une question. Dans ces cas, les réponses sont passées en revue, afin de déterminer si elles devraient être attribuées à l'une des catégories déterminées au préalable, jointes à d'autres réponses similaires pour créer une nouvelle catégorie, ou demeurer telles quelles.

7.4 Imputation

L'imputation est le processus qui fournit des valeurs valides concernant les variables qui ont été retenues pour être modifiées, soit en raison de renseignements invalides, soit en raison de renseignements manquants. Les nouvelles valeurs sont établies de façon à préserver la structure sous-jacente des données et à garantir que les enregistrements qui en résultent passeront tous les contrôles requis. En d'autres mots, l'objectif n'est pas de reproduire les véritables valeurs des microdonnées mais plutôt d'établir des enregistrements intrinsèquement cohérents qui permettront de produire de bonnes estimations agrégées.

Nous pouvons faire la distinction entre trois types de non-réponse. La non-réponse complète est lorsque le répondant ne fournit pas le nombre minimal de réponses. Ces enregistrements sont supprimés et seront pris en compte lors du processus de pondération (voir le chapitre 11.0). La non-réponse ponctuelle est lorsque le répondant ne fournit pas une réponse à une question mais qu'il passe à la question suivante. Ce type de non-réponse est habituellement traité en utilisant le code « non déclaré » ou en ayant recours à l'imputation. Enfin, la non-réponse partielle est lorsque le répondant fournit le nombre minimal de réponses mais ne termine pas l'interview. Ces enregistrements peuvent être traités comme des cas de non-réponse complète ou ponctuelle.

Dans le cas de l'ECCF, l'imputation par enregistrement donneur a été utilisée pour fournir les éléments d'information manquants aux enregistrements avec non-réponse ponctuelle ou partielle pour les montants de revenus personnels et revenus ménages. De plus amples renseignements sur le processus d'imputation sont donnés au chapitre 8.0 (Qualité des données).

7.5 Création de variables dérivées

Au total, 36 éléments de données inclus dans le fichier de microdonnées ont été calculés en combinant des postes du questionnaire pour faciliter l'analyse des données. La plupart sont des variables continues ayant trait à l'âge et au nombre d'années de service. Ces variables ont été regroupées d'après des intervalles prédéterminés pour faciliter le processus d'analyse.

7.6 Pondération

Le principe qui sous-tend une estimation pour un échantillon probabiliste, tel que l'ECCF, veut que chacune des personnes incluses dans l'échantillon « représente », en plus d'elle-même, plusieurs autres personnes qui en sont exclues. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne incluse dans l'échantillon représente 50 membres de la population.

La phase de la pondération est une étape où l'on calcule ce nombre (ou poids) pour chaque enregistrement. Ce poids, qui figure dans le fichier de microdonnées, **doit** servir à calculer des estimations significatives à partir de l'enquête. Si, par exemple, le nombre de personnes au Canada qui n'ont pas un compte bancaire personnel doit être estimé, cette opération s'effectue en sélectionnant les enregistrements renvoyant aux personnes incluses à l'intérieur de

l'échantillon qui présentent cette caractéristique (OE_Q02A = 0) et en additionnant les poids inscrits dans ces enregistrements.

Le chapitre 11.0 renferme des détails au sujet de la méthode utilisée pour calculer ces poids.

7.7 *Suppression de renseignements confidentiels*

Il convient de souligner que les fichiers de microdonnées « à grande diffusion » (FMGD) peuvent différer des fichiers « maîtres » de l'enquête que conserve Statistique Canada. Ces différences sont habituellement le résultat de mesures prises pour protéger l'anonymat des répondants à une enquête. Les mesures les plus courantes sont la suppression de variables du fichier, le regroupement de valeurs en des catégories plus étendues et le codage de valeurs spécifiques à la catégorie « Non déclaré ». Les utilisateurs ayant besoin d'avoir accès à de l'information exclue des fichiers de microdonnées peuvent acheter des totalisations spéciales. Les estimations produites seront communiquées à l'utilisateur, sous réserve du respect des lignes directrices pour l'analyse et la diffusion dont le chapitre 9.0 de ce document fournit un aperçu.

8.0 Qualité des données

8.1 Taux de réponse

Le tableau suivant résume les taux de réponses obtenus par province pour l'Enquête canadienne sur les capacités financières (ECCF).

Province	Nombre total de ménages	Nombre total de répondants	Taux de réponse
Terre-Neuve-et-Labrador	1 224	670	54,7
Île-du-Prince-Édouard	572	324	56,7
Nouvelle-Écosse	1 326	779	58,8
Nouveau-Brunswick	1 213	684	56,4
Québec	5 824	3 336	57,3
Ontario	8 185	4 519	55,2
Manitoba	1 427	795	55,7
Saskatchewan	1 976	1 213	61,4
Alberta	2 748	1 690	61,5
Colombie-Britannique	3 061	1 509	49,3
Canada	27 555	15 519	56,3

Un **répondant** possède les caractéristiques suivantes :

- La liste de membres du ménage a été remplie sans aucun refus à propos de l'âge des personnes.
- La personne sélectionnée était âgée de 18 ans ou plus au moment de l'interview (confirmé avec la personne sélectionnée).
- La personne sélectionnée a répondu au moins à deux-tiers des items clés dans trois des premiers cinq modules du questionnaire (Démographie, Emploi, Dépenses quotidiennes, Gestion financière et Dépenses importantes) ainsi qu'au moins un item clé du module Gestion financière.

8.2 Erreurs relatives à l'enquête

Les estimations calculées à partir de cette enquête reposent sur un échantillon de ménages. Des estimations légèrement différentes auraient pu être obtenues si un recensement complet avait été effectué en reprenant le même questionnaire et en faisant appel aux mêmes intervieweurs, superviseurs, méthodes de traitement, etc. que ceux effectivement utilisés dans l'enquête. L'écart entre les estimations découlant de l'échantillon et celles que donnerait un dénombrement complet réalisé dans des conditions semblables est appelé erreur d'échantillonnage de l'estimation.

Des erreurs qui ne sont pas liées à l'échantillonnage peuvent se produire à presque toutes les étapes des opérations d'enquête. Les intervieweurs peuvent avoir mal compris les instructions, les enquêtés peuvent se tromper en répondant aux questions, les réponses peuvent être mal saisies sur le questionnaire et des erreurs peuvent survenir lors du traitement et de la totalisation des données. Ces erreurs sont toutes des exemples d'erreurs non dues à l'échantillonnage.

Sur un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations calculées à partir de l'enquête. Toutefois, les erreurs systématiques contribuent à biaiser les estimations de l'enquête. Énormément de temps et d'efforts ont été consacrés à réduire les erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été prises à chacune des étapes du cycle de collecte et de traitement des données afin de contrôler la qualité des données. Ces mesures comprennent les essais cognitifs pour assurer la compréhension des concepts, la formation poussée des intervieweurs concernant les procédures de l'enquête et de l'application de l'interview téléphonique assistée par ordinateur (ITAO), l'observation des intervieweurs en vue de cerner les problèmes liés à la conception du questionnaire ou à une mauvaise compréhension des instructions, et l'évaluation de l'application ITAO pour s'assurer que les contrôles des limites, les vérifications et le déroulement des questions étaient tous programmés correctement.

8.2.1 Collecte des données

La formation des intervieweurs pour l'ECCF consistait à lire le Manuel du surveillant, le Manuel des procédures, le Manuel de l'intervieweur, la pratique avec des cas simulés de l'ECCF sur ordinateur et une discussion avec l'intervieweur principal des questions de l'intervieweur avant le début de l'enquête. L'information fournie aux intervieweurs comprenait une description du contexte et des objectifs de l'enquête ainsi qu'un glossaire de terminologie et une série de questions et de réponses. La période de collecte était du 11 février au 9 mai 2009.

8.2.2 Traitement des données

Le traitement des données de l'ECCF a été fait par étape, y compris la vérification, le codage, le contrôle, l'imputation, l'estimation, la confidentialité, etc. À chaque étape, une photo des fichiers de sortie est prise, et il est facile de faire une vérification en comparant les fichiers de l'étape en cours avec ceux de l'étape antérieure. Cette façon de procéder a beaucoup amélioré l'étape du traitement des données.

8.2.3 Non-réponse

L'une des principales sources d'erreurs non dues à l'échantillonnage observées dans le cadre des enquêtes est l'effet de la non-réponse sur les résultats de l'enquête. L'étendue d'une non-réponse varie d'une non-réponse partielle (le fait de ne pas répondre simplement à une ou à plusieurs questions) à une non-réponse totale. S'il y a eu non-réponse totale, c'est parce que l'intervieweur a été incapable de communiquer avec le répondant, qu'aucun membre du ménage n'a pu fournir l'information demandée ou que le répondant a refusé de participer à l'enquête. Les non-réponses totales ont été traitées en ajustant le poids des personnes qui ont répondu au questionnaire d'enquête de façon à le contrebalancer pour ceux qui n'y ont pas répondu.

Dans la plupart des cas, il y a eu non-réponse partielle au questionnaire d'enquête lorsque le répondant n'a pas compris ou a mal interprété une question, a refusé d'y répondre ou ne pouvait se rappeler l'information demandée. Afin de produire des données complètes concernant la répartition du revenu personnel et du revenu ménage, des valeurs ont été imputées lorsqu'elles étaient manquantes.

Les imputations ont été faites à l'aide de donneurs qui avaient été choisis au moyen d'une fonction de caractérisation. Pour tous les enregistrements contenant de la non-réponse ponctuelle ou de la non-réponse partielle (également appelés enregistrements receveurs), certaines caractéristiques ont été comparées avec celles de tous les donneurs. Lorsque les caractéristiques étaient les mêmes entre un donneur et le receveur, une valeur était ajoutée au résultat de ce donneur. Le donneur ayant obtenu le

résultat le plus élevé était réputé être le donneur « le plus semblable », et il était choisi pour fournir les éléments d'information manquants des non-répondants. S'il y avait plus d'un donneur avec le résultat le plus élevé, on procédait à une sélection aléatoire. Le bassin de donneurs était constitué de telle façon que la valeur imputée assignée au receveur, parallèlement à d'autres éléments d'information non imputés du receveur, serait encore acceptée à la vérification.

On a imputé le revenu personnel et celui du ménage (ensemble si nécessaire, et toujours en se servant du même donneur). Le tableau qui suit montre le taux d'imputation pour chacune des variables lorsque approprié.

Au total, presque 10 000 répondants (63 %) constituaient des donneurs éligibles ayant fourni à la fois le revenu personnel et celui du ménage. On avait demandé à ceux qui n'avaient fourni d'estimation en dollars de leurs revenus une série de questions visant à dériver une catégorie de revenu. Presque 2 000 répondants (13 %) n'avaient pas fourni d'information sur leurs revenus. Les catégories de revenu et l'information manquante sur le revenu ont été imputées par les valeurs des donneurs grâce à une série d'étapes, en fonction de l'information disponible pour les autres variables utilisées pour former les groupes d'imputation. La dernière étape consistait à convertir tous les revenus, qu'ils soient déclarés ou imputés, selon des catégories de quartiles, de quintiles, et de déciles pour des raisons analytiques.

	Revenu personnel	Revenu du ménage
Imputé	4 775	5 223
Total	15 519	15 519
Taux (%)	30,8	33,7

Le processus d'imputation de l'ECCF a donné de bons résultats et il a contribué à compléter les réponses incomplètes avec l'expérience d'autres répondants ayant des caractéristiques semblables ou identiques. Il augmentera le nombre d'unités utilisées dans toute analyse effectuée par des chercheurs.

8.2.4 Mesure de l'erreur d'échantillonnage

Puisqu'il est inévitable que des estimations établies à partir d'une enquête-échantillon (ou par sondage) soient sujettes à une erreur d'échantillonnage, une saine pratique de la statistique exige que les chercheurs fournissent aux utilisateurs une certaine indication de l'importance de cette erreur d'échantillonnage. Cette section de la documentation renferme un aperçu des mesures de l'erreur d'échantillonnage dont Statistique Canada se sert couramment et dont le Bureau conseille vivement aux utilisateurs qui produisent des estimations à partir de ce fichier de microdonnées à employer également.

La base pour mesurer l'importance potentielle des erreurs d'échantillonnage est l'erreur-type des estimations calculées à partir des résultats d'une enquête.

En raison, cependant, de la diversité des estimations pouvant être produites à partir d'une enquête, l'erreur-type d'une estimation est habituellement exprimée en fonction de l'estimation à laquelle elle se rapporte. La mesure résultante, appelée coefficient de variation (CV) d'une estimation, s'obtient en divisant l'erreur-type de l'estimation par l'estimation elle-même et s'exprime en pourcentage de l'estimation.

Par exemple, supposons que, d'après les résultats de l'enquête, l'on estime que 13,5 % des Canadiens ont déclaré vérifier leur solde de compte personnel ou conjoint

quotidiennement, et l'on constate que l'erreur-type de cette estimation est de 0,0034. Le coefficient de variation de l'estimation est donc calculé comme suit:

$$\left(\frac{0,0034}{0,135} \right) \times 100 \% = 2,5 \%$$

De plus amples renseignements sur le calcul des coefficients de variation se trouvent au chapitre 10.0.

9.0 Lignes directrices pour la totalisation, l'analyse et la diffusion de données

Ce chapitre de la documentation renferme un aperçu des lignes directrices que doivent respecter les utilisateurs qui totalisent, analysent, publient ou autrement diffusent des données calculées à partir des fichiers de microdonnées de l'enquête. Ces lignes directrices devraient permettre aux utilisateurs de microdonnées de produire les mêmes chiffres que ceux produits par Statistique Canada, tout en étant en mesure d'obtenir des chiffres actuellement inédits de façon conforme à ces lignes directrices établies.

9.1 Lignes directrices pour l'arrondissement d'estimations

Afin que les estimations qui sont destinées à la publication ou à toute autre forme de diffusion qui sont calculées à partir de ces fichiers de microdonnées correspondent à celles produites par Statistique Canada, nous conseillons vivement aux utilisateurs de respecter les lignes directrices qui suivent en ce qui concerne l'arrondissement de telles estimations :

- a) Les estimations dans le corps principal d'un tableau statistique doivent être arrondies à la centaine près à l'aide de la technique d'arrondissement normale. Selon cette technique, si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, le dernier chiffre à conserver est augmenté de 1. Par exemple, selon la technique d'arrondissement normale à la centaine près, si les deux derniers chiffres se situent entre 00 et 49, ils sont remplacés par 00 et le chiffre précédent (le chiffre des centaines) reste inchangé. Si les derniers chiffres se situent entre 50 et 99, ils sont remplacés par 00 et le chiffre précédent est augmenté de 1.
- b) Les totaux partiels marginaux et les totaux marginaux des tableaux statistiques doivent être calculés à partir de leurs composantes non arrondies correspondantes, puis ensuite être arrondis à leur tour à la centaine près à l'aide de la technique d'arrondissement normale.
- c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir de composantes non arrondies (c'est-à-dire des numérateurs et/ou des dénominateurs), puis être arrondis à leur tour à une décimale à l'aide de la technique d'arrondissement normale. Dans le cas d'un arrondissement normal à un seul chiffre, si le dernier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, le dernier chiffre à conserver est augmenté de 1.
- d) Les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs composantes non arrondies correspondantes, puis être arrondies à leur tour à la centaine près (ou à la décimale près) à l'aide de la technique d'arrondissement normale.
- e) Dans les cas, où, en raison de limitations d'ordre technique ou de toutes autres limites, une technique d'arrondissement autre que la technique normale est utilisée produisant des estimations à être publiées ou autrement diffusées différentes des estimations correspondantes publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces différences dans le ou les documents à publier ou à diffuser.
- f) En aucun cas, les utilisateurs ne doivent publier ou autrement diffuser des estimations non arrondies. Des estimations non arrondies laissent entendre qu'elles sont plus précises qu'elles ne le sont en réalité.

9.2 Lignes directrices pour la pondération de l'échantillon en vue de la totalisation

Le plan d'échantillonnage utilisé pour l'Enquête canadienne sur les capacités financières (ECCF) n'était pas autopondéré. Lorsqu'ils produisent des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent appliquer les poids appropriés.

Si les poids appropriés ne sont pas utilisés, les estimations calculées à partir des fichiers de microdonnées ne peuvent être considérées comme représentatives de la population visée par l'enquête et ne correspondront pas à celles produites par Statistique Canada.

Les utilisateurs devraient également prendre note que certains progiciels pourraient peut-être ne pas permettre la production d'estimations correspondant exactement à celles qu'offre Statistique Canada, en raison du mode de traitement du poids par ces progiciels.

9.3 Définitions de types d'estimations : catégoriques et quantitatives

Avant de discuter de la façon dont on peut totaliser et analyser les données de l'ECCF, il est utile de décrire les deux principaux types d'estimations ponctuelles des caractéristiques de la population qui peuvent être produites à partir du fichier de microdonnées créé pour l'ECCF.

9.3.1 Estimations catégoriques

Les estimations catégoriques sont des estimations du nombre ou du pourcentage de membres de la population visée par l'enquête possédant certaines caractéristiques ou faisant partie d'une catégorie définie. Le nombre d'adultes canadiens qui ont un budget pour leur ménage ou la proportion en retard de paiements constituent des exemples de telles estimations. Une estimation du nombre de personnes possédant une certaine caractéristique peut aussi être désignée comme une estimation d'un agrégat.

Exemples de questions catégoriques:

Q : Établissez-vous un budget pour votre ménage?
R : Oui / Non

Q : À ce jour, quel pourcentage du prix d'achat total avez-vous mis de côté pour l'achat de votre future maison?
R : Moins de 5 % / 5 à 10 % / 11 à 20 % / 21 à 50 % / 51 à 75 % / 76 à 100 %

9.3.2 Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes et d'autres mesures d'une tendance centrale de quantités reposant sur certains ou sur tous les membres de la population visée par l'enquête. Elles comprennent aussi expressément des estimations de la forme \hat{X} / \hat{Y} où \hat{X} est une estimation de la quantité totale pour la population visée par l'enquête et \hat{Y} , est une estimation du nombre de personnes dans la population visée par l'enquête qui contribuent à cette quantité totale.

Un exemple d'estimation quantitative est le nombre moyen de comptes de banque personnels détenus par les femmes adultes au Canada. Le numérateur est une

estimation du nombre total de comptes de banque personnels et son dénominateur est le nombre de femmes adultes (18 ans et plus) au Canada.

Exemple de questions quantitatives :

Q : Combien avez-vous actuellement de comptes de chèques ou d'épargne personnels dans une banque, une caisse populaire ou une société de fiducie?

R : |_|_| comptes

Q : Pendant combien d'années de plus prévoyez-vous faire des paiements hypothécaires pour votre résidence principale?

R : |_|_| années

9.3.3 Totalisation d'estimations catégoriques

On peut obtenir des estimations du nombre de gens possédant une certaine caractéristique à partir du fichier de microdonnées en additionnant les poids finals de tous les enregistrements possédant la ou les caractéristiques qui nous intéressent. On obtient des proportions et des rapports de la forme \hat{X} / \hat{Y} en :

- additionnant les poids finals des enregistrements présentant la caractéristique qui nous intéresse pour le numérateur (\hat{X}),
- additionnant les poids finals des enregistrements présentant la caractéristique qui nous intéresse pour le dénominateur (\hat{Y}), puis en
- divisant l'estimation a) par celle de b) (\hat{X} / \hat{Y}).

9.3.4 Totalisation d'estimations quantitatives

On peut obtenir des estimations de quantités à partir du fichier de microdonnées en multipliant la valeur de la variable qui nous intéresse par le poids final de chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements qui nous intéressent. Par exemple, pour obtenir une estimation du nombre total de comptes de banque personnels détenus par des femmes adultes au Canada, multipliez la valeur déclarée à la question OE_Q02A (nombre de comptes de chèques ou d'épargne personnels dans une banque) par le poids final de l'enregistrement, puis additionnez cette valeur pour tous les enregistrements où la variable SEX = 2 (femme).

Pour obtenir une moyenne pondérée de la forme \hat{X} / \hat{Y} , le numérateur (\hat{X}) est calculé comme une estimation quantitative et le dénominateur (\hat{Y}) est calculé comme une estimation catégorique. Par exemple, pour estimer le nombre moyen de comptes de banque personnels détenus par les femmes :

- estimez le nombre total de comptes de banque personnels (\hat{X}), tel qu'il est décrit ci-dessus,
- estimez le nombre de femmes au Canada (\hat{Y}) incluses dans cette catégorie en additionnant les poids finals de tous les enregistrements où la variable SEX = 2, puis
- divisez l'estimation a) par l'estimation b) (\hat{X} / \hat{Y}).

9.4 Lignes directrices pour l'analyse statistique

L'ECCF repose sur un plan d'échantillonnage complexe comportant une stratification, de multiples étapes de sélection ainsi que des probabilités inégales de sélection des répondants. L'utilisation des données provenant d'enquêtes aussi complexes présente des problèmes pour les analystes, parce que le plan d'enquête et les probabilités de sélection influent sur les procédures d'estimation et de calcul de la variance qui devraient être utilisées. Il faut utiliser les poids de l'enquête pour que les estimations et les analyses des données de l'enquête soient exemptes de biais.

Bien que de nombreuses procédures d'analyse que l'on trouve à l'intérieur de progiciels statistiques permettent d'utiliser des poids, la signification ou la définition du poids inclus dans ces procédures peut différer de ce qui convient dans le contexte d'une enquête-échantillon, de telle sorte que dans bien des cas les estimations produites au moyen de ces progiciels sont correctes, mais que les variances calculées sont piètres. Les variances approximatives pour des estimations simples comme des totaux, des proportions et des rapports (pour des variables qualitatives) peuvent être calculées à partir des tables de variabilité d'échantillonnage approximative qui accompagnent les données.

Pour d'autres techniques d'analyse (de régression linéaire, de régression logistique et de l'analyse de variance, par exemple), il existe une méthode qui peut rendre les variances calculées par l'application des progiciels normalisés plus significatives, en intégrant les probabilités inégales de sélection. L'application de cette méthode entraîne une remise à l'échelle des poids de façon à ce que le poids moyen soit de 1.

Supposons, par exemple, qu'il faut effectuer l'analyse de tous les répondants de sexe masculin. Les étapes à suivre pour remettre à l'échelle les poids sont les suivantes :

- 1) sélectionner tous les répondants du fichier qui ont déclaré SEXE = homme;
- 2) calculer le poids MOYEN pour ces enregistrements en additionnant les poids originaux des personnes établis à partir du fichier de microdonnées pour ces enregistrements puis diviser cette somme par le nombre de répondants ayant déclaré SEXE = homme;
- 3) pour chacun de ces répondants, calculer un poids REMIS À L'ÉCHELLE égal au poids original de la personne divisé par le poids MOYEN;
- 4) effectuer l'analyse portant sur ces répondants en utilisant le poids REMIS À L'ÉCHELLE.

Parce qu'on ne tient toujours compte ni de la stratification ni des grappes du plan d'échantillonnage, les estimations des variances calculées avec cette méthode risquent cependant d'être des sous-estimations.

Il faut connaître les détails du plan d'enquête pour calculer des estimations des variances plus précises. De tels détails ne peuvent être fournis dans le fichier de microdonnées en raison de la confidentialité. Statistique Canada peut, contre remboursement des frais, calculer des variances qui tiennent compte du plan complet d'échantillonnage pour beaucoup de statistiques.

9.5 Lignes directrices pour la diffusion de coefficients de variation

Avant de diffuser et/ou de publier toutes estimations établies à partir de l'ECCF, les utilisateurs devraient premièrement déterminer le niveau de qualité de cette estimation. Les niveaux de qualité sont *acceptable*, *médiocre* et *inacceptable*. Les erreurs d'échantillonnage et non dues à l'échantillonnage, dont il a été question au chapitre 8.0, influencent la qualité des données. Aux

fins du présent document, cependant, on ne déterminera le niveau de qualité d'une estimation qu'à partir d'une erreur d'échantillonnage dont rend compte le coefficient de variation indiqué à l'intérieur du tableau qui figure ci-dessous. Les utilisateurs devraient néanmoins s'assurer de lire le chapitre 8.0 pour être plus pleinement informés des caractéristiques relatives à la qualité de ces données.

On devrait premièrement déterminer le nombre de répondants retenus pour le calcul de l'estimation. Si ce nombre est inférieur à 30, il faudrait considérer l'estimation pondérée comme étant de qualité inacceptable

Pour les estimations pondérées fondées sur les tailles d'échantillon de 30 ou plus, les utilisateurs devraient déterminer le coefficient de variation de l'estimation et suivre les lignes directrices relatives au niveau de qualité qui figurent ci-dessous. Celles-ci devraient être appliquées, pour la détermination du niveau de qualité d'une estimation, aux estimations pondérées arrondies.

On peut considérer qu'il est possible de divulguer toutes les estimations. Celles d'un niveau de qualité médiocre ou inacceptable doivent cependant être accompagnées d'une mise en garde pour avertir les utilisateurs subséquents.

Lignes directrices relatives au niveau de qualité de l'estimation

Niveau de qualité de l'estimation	Lignes directrices
1) Acceptable	<p>Les estimations proviennent d'une taille d'échantillon de 30 ou plus, et présentent de faibles coefficients de variation, de l'ordre de 0,0 à 16,5 %.</p> <p>Aucune mise en garde n'est requise.</p>
2) Médiocre	<p>Les estimations proviennent d'une taille d'échantillon de 30 ou plus, et présentent des coefficients de variation élevés, de l'ordre de 16,6 à 33,3 %.</p> <p>Ces estimations devraient être signalées par la lettre E (ou un quelconque identificateur similaire). Elles devraient être accompagnées d'une mise en garde avertissant les utilisateurs subséquents des niveaux élevés d'erreur associés aux estimations.</p>
3) Inacceptable	<p>Les estimations proviennent d'une taille d'échantillon inférieure à 30, ou présentent des coefficients de variation très élevés, supérieurs à 33,3 %.</p> <p>Statistique Canada recommande de ne pas diffuser d'estimations de qualité inacceptable. Si un utilisateur choisit cependant de le faire, ces estimations devraient alors être signalées à l'aide de la lettre F (ou d'un quelconque identificateur similaire) et devraient être accompagnées de la mise en garde suivante :</p> <p>« Nous informons l'utilisateur que ces estimations (désignées avec la lettre F) ne respectent pas les normes de qualité de Statistique Canada. Les conclusions qui reposeront sur ces données ne seront pas fiables et seront très probablement invalides. »</p>

9.6 Seuils pour la diffusion des estimations

Le tableau ci-dessous fournit une indication de la précision des estimations démographiques ainsi que les seuils de diffusion associés aux trois niveaux de qualité de l'estimation présentés à la section précédente. Ces seuils proviennent des tables de coefficients de variation (CV) dont il sera question au chapitre 10.0.

Par exemple, d'après le tableau, la qualité d'une estimation pondérée de 15 000 personnes possédant une caractéristique donnée à Terre-Neuve-et-Labrador est médiocre.

Veillez noter que ces seuils de diffusion correspondent aux estimations de chiffres de population seulement. Dans le cas d'estimations de rapports, les utilisateurs ne devraient pas utiliser la valeur du numérateur (ni le dénominateur) afin de trouver le niveau de qualité de l'estimation correspondant. La règle 4 à la section 10.1 ainsi que l'exemple 4 à la section 10.1.1 expliquent la bonne procédure à suivre dans le cas d'un rapport.

Province and region	CV acceptable 0,0 à 16,5 %	CV médiocre 16,6 à 33,3 %	CV inacceptable > 33,3 %
Terre-Neuve-et-Labrador	27 500 et plus	7 000 à < 27 500	moins de 7 000
Île-du-Prince-Édouard	13 000 et plus	3 500 à < 13 000	moins de 3 500
Nouvelle-Écosse	43 000 et plus	11 000 à < 43 000	moins de 11 000
Nouveau-Brunswick	46 000 et plus	12 000 à < 46 000	moins de 12 000
Québec	93 000 et plus	23 000 à < 93 000	moins de 23 000
Ontario	113 500 et plus	28 000 à < 113 500	moins de 28 000
Manitoba	52 500 et plus	13 500 à < 52 500	moins de 13 500
Saskatchewan	33 000 et plus	8 500 à < 33 000	moins de 8 500
Alberta	77 000 et plus	19 500 à < 77 000	moins de 19 500
Colombie-Britannique	129 500 et plus	32 500 à < 129 500	moins de 32 500
Canada	99 000 et plus	24 500 à < 99 000	moins de 24 500

10.0 Tables de variabilité d'échantillonnage approximative

Afin de fournir des coefficients de variation (CV) qui pourraient s'appliquer à une gamme étendue d'estimations catégoriques produites à partir de ce fichier de microdonnées et auxquels il serait facilement possible pour l'utilisateur d'avoir accès, un ensemble de tables de variabilité d'échantillonnage approximative a été produit. Ces tables de CV permettent à l'utilisateur d'obtenir un coefficient de variation approximatif fondé sur la taille de l'estimation calculée à partir des données de l'enquête.

Les coefficients de variation sont calculés à l'aide de la formule de la variance pour un échantillonnage aléatoire simple et en y incorporant un facteur qui reflète la nature du plan d'échantillonnage, qui est à plusieurs degrés et qui prévoit la formation de grappes. Ce facteur, appelé l'effet du plan, a été déterminé en calculant premièrement les effets du plan pour une gamme étendue de caractéristiques, puis en choisissant parmi ceux-ci une valeur modérée (habituellement le 75^e percentile) à utiliser à l'intérieur des tables de CV qui s'appliqueraient ensuite à l'ensemble entier des caractéristiques.

Le tableau ci-dessous indique la valeur modérée des effets du plan, ainsi que les tailles de l'échantillon et les chiffres de population selon la province qui ont été utilisés pour produire les tables de variabilité d'échantillonnage approximative de l'Enquête canadienne sur les capacités financières (ECCF).

Province et région	Effet du plan	Taille de l'échantillon	Population
Terre-Neuve-et-Labrador	1,32	670	410 773
Île-du-Prince-Édouard	1,21	324	109 723
Nouvelle-Écosse	1,29	779	750 481
Nouveau-Brunswick	1,53	684	602 590
Québec	1,39	3 336	6 167 275
Ontario	1,39	4 519	10 161 677
Manitoba	1,32	795	911 804
Saskatchewan	1,47	1 213	771 190
Alberta	1,33	1 690	2 740 104
Colombie-Britannique	1,54	1 509	3 586 822
Canada	1,60	15 519	26 212 439

Tous les coefficients de variation inclus dans les tables de variabilité d'échantillonnage approximative sont approximatifs et donc non officiels. Des estimations de la variance réelle pour des variables précises peuvent être obtenues auprès de Statistique Canada, contre remboursement des frais. Étant donné que le CV approximatif est une estimation prudente, l'utilisation de la variance réelle estimée pourrait faire passer l'estimation d'un niveau de qualité à un autre. Par exemple, une estimation *médiocre* pourrait devenir *acceptable* si elle était fondée sur le calcul du CV exact.

Rappelez-vous que : Si le nombre d'observations sur lesquelles une estimation est basée est inférieur à 30, l'estimation pondérée est très probablement inacceptable et Statistique Canada recommande de ne pas diffuser une telle estimation, quelle que soit la valeur du coefficient de variation.

10.1 Comment utiliser les tables de coefficients de variation pour des estimations catégoriques

Les règles qui suivent devraient permettre à l'utilisateur de déterminer les coefficients de variation approximatifs à partir des tables de variabilité d'échantillonnage approximative pour des estimations du nombre, de la proportion ou du pourcentage de membres de la population visée par l'enquête possédant une certaine caractéristique et pour des rapports et des différences entre de telles estimations.

Règle 1 : Estimations du nombre de personnes possédant une caractéristique donnée (agrégats)

Le coefficient de variation dépend uniquement de la taille de l'estimation elle-même. Dans la table de variabilité d'échantillonnage approximative pour la région géographique appropriée, repérez le nombre estimé dans la colonne la plus à gauche (intitulée « Numérateur du pourcentage ») et suivez les astérisques (le cas échéant) jusqu'au premier chiffre rencontré. Ce chiffre est le coefficient de variation approximatif.

Règle 2 : Estimations de proportions ou de pourcentages de personnes possédant une caractéristique donnée

Le coefficient de variation d'une proportion estimée ou d'un pourcentage estimé dépend à la fois de la taille de la proportion ou du pourcentage et de la taille du total sur lequel la proportion ou le pourcentage repose. Les proportions estimées ou les pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque la proportion ou que le pourcentage repose sur un sous-groupe de la population. Par exemple, la proportion de femmes canadiennes qui ont un budget pour le ménage est plus fiable que le nombre estimé de femmes canadiennes qui ont un budget pour le ménage. (Remarquez que dans les tables la valeur des coefficients de variation diminue lorsqu'on les lit de gauche à droite.)

Lorsque la proportion ou que le pourcentage repose sur la population totale de la région géographique visée par la table, le CV de la proportion ou du pourcentage est le même que le CV du numérateur de la proportion ou du pourcentage. Dans ce cas, la règle 1 peut être appliquée.

Lorsque la proportion ou que le pourcentage repose sur un sous-ensemble de la population totale (p. ex., un groupe d'âge ou un sexe particulier), particulier, on devrait faire référence à la proportion ou au pourcentage (dans le haut de la table) et au numérateur de la proportion ou du pourcentage (dans la colonne de gauche de la table). L'intersection de la rangée et de la colonne appropriées donne le coefficient de variation.

Règle 3 : Estimations de différences entre des agrégats et des pourcentages

L'erreur-type d'une différence entre deux estimations est approximativement égale à la racine carrée de la somme des carrés de chaque erreur-type considérée séparément. C'est-à-dire que l'erreur-type d'une différence ($\hat{d} = \hat{X}_1 - \hat{X}_2$) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1 \alpha_1)^2 + (\hat{X}_2 \alpha_2)^2}$$

où \hat{X}_1 est l'estimation 1, \hat{X}_2 est l'estimation 2 et α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement. Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}}/\hat{d}$. Cette

formule est exacte pour la différence entre des caractéristiques distinctes et non corrélées, mais n'est autrement qu'approximative.

Règle 4 : Estimations de rapports

Dans le cas où le numérateur est un sous-ensemble du dénominateur, le rapport devrait être converti en un pourcentage et la règle 2 appliquée. Cela s'appliquerait, par exemple, au cas où le dénominateur est le nombre de femmes ayant un budget pour leur ménage et le numérateur, le nombre de femmes ayant un budget pour leur ménage qui déclarent toujours respecter ce budget.

Dans le cas où le numérateur n'est pas un sous-ensemble du dénominateur, comme dans l'exemple du rapport du nombre d'hommes ayant un budget pour leur ménage comparativement au nombre de femmes ayant un budget pour leur ménage, l'erreur-type du rapport des estimations est approximativement égale à la racine carrée de la somme des carrés de chaque coefficient de variation considéré séparément multipliée par \hat{R} . C'est-à-dire que l'erreur-type d'un rapport ($\hat{R} = \hat{X}_1 / \hat{X}_2$) est :

$$\sigma_{\hat{R}} = \hat{R} \sqrt{\alpha_1^2 + \alpha_2^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et de \hat{X}_2 respectivement. Le coefficient de variation de \hat{R} est donné par $\sigma_{\hat{R}} / \hat{R}$. La formule tendra à surestimer l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélés positivement et à la sous-estimer si \hat{X}_1 et \hat{X}_2 sont corrélés négativement.

Règle 5 : Estimations de différences entre des rapports

Dans ce cas, les règles 3 et 4 sont combinées. On détermine premièrement les CV pour les deux rapports à l'aide de la règle 4, puis on trouve le CV de leur différence au moyen de la règle 3.

10.1.1 Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques

Les exemples ci-dessous utilisent des données du fichier de l'ECCF et sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter.

Exemple 1 : Estimations du nombre de personnes possédant une caractéristique donnée (agrégats)

Supposons qu'un utilisateur estime que 13 233 741 Canadiens adultes déclarent avoir un budget pour leur ménage. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Reportez-vous à la table des coefficients de variation pour le CANADA.
- 2) L'agrégat estimé (13 233 741) ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, c'est-à-dire 12 500 000.
- 3) On trouve le coefficient de variation pour un agrégat estimé en se reportant à la première entrée autre que des astérisques sur cette rangée, c'est-à-dire 1,0 %.

- 4) Le coefficient de variation approximatif de l'estimation est donc 1,0 %. Le résultat selon lequel il y avait 13 233 741 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1) Canadiens adultes ont un budget pour leur ménage, peut être publié sans réserves.

Exemple 2 : Estimations de proportions ou de pourcentages de personnes possédant une caractéristique donnée

Supposons que l'utilisateur estime que $4\,881\,629 / 13\,233\,741 = 36,9\%$ la proportion d'adultes canadiens ayant un budget pour leur ménage déclarent toujours respecter ce budget. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Reportez-vous à la table des coefficients de variation pour le CANADA.
- 2) Parce que l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c'est-à-dire d'adultes ayant un budget pour leur ménage), il faut utiliser à la fois le pourcentage (36,9 %) et la portion numérateur du pourcentage (4 881 629) pour déterminer le coefficient de variation.
- 3) Le numérateur, 4 881 629, ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, soit 5 000 000. De même, l'estimation du pourcentage ne figure dans l'en-tête d'aucune colonne; il faut donc utiliser la proportion qui s'en rapproche le plus, soit 35,0 %.
- 4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 1,9 %, est le coefficient de variation à employer.
- 5) Le coefficient de variation approximatif de l'estimation est donc 1,9 %. Le résultat selon lequel 36,9 % des adultes ayant un budget pour leur ménage déclarent toujours respecter ce budget, peut être publié sans réserves.

Enquête canadienne sur les capacités financières, 2009

Tables de la variabilité d'échantillonnage approximative - Canada - Tous les âges

NUMERATEUR DU POURCENTAGE (en milliers)	POURCENTAGE ESTIME														
	0,1%	1,0%	2,0%	5,0%	10,0%	15,0%	20,0%	25,0%	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%	
1	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	90,0	52,0
2	#####	#####	#####	#####	#####	#####	#####	#####	#####	97,2	93,7	90,0	82,2	63,7	36,7
3	94,8	94,4	93,9	92,5	90,0	87,5	84,9	82,2	79,4	76,5	73,5	67,1	52,0	30,0	
4	82,1	81,8	81,3	80,1	78,0	75,8	73,5	71,2	68,7	66,2	63,7	58,1	45,0	26,0	
5	73,5	73,1	72,8	71,6	69,7	67,8	65,7	63,7	61,5	59,3	56,9	52,0	40,3	23,2	
6	67,1	66,8	66,4	65,4	63,7	61,9	60,0	58,1	56,1	54,1	52,0	47,4	36,7	21,2	
7	62,1	61,8	61,5	60,5	58,9	57,3	55,6	53,8	52,0	50,1	48,1	43,9	34,0	19,6	
...
...
...
2 000	*****	t*****	t*****	*****	3,5	3,4	3,3	3,2	3,1	3,0	2,8	2,6	2,0	1,2	
3 000	*****	t*****	t*****	*****	t*****	2,8	2,7	2,6	2,5	2,4	2,3	2,1	1,6	0,9	
4 000	*****	t*****	t*****	*****	t*****	t*****	2,3	2,3	2,2	2,1	2,0	1,8	1,4	0,8	
5 000	*****	t*****	t*****	*****	t*****	t*****	2,1	2,0	1,9	1,9	1,8	1,6	1,3	0,7	
6 000	*****	t*****	t*****	*****	t*****	t*****	*****	1,8	1,8	1,7	1,6	1,5	1,2	0,7	
7 000	*****	t*****	t*****	*****	t*****	t*****	*****	*****	1,6	1,6	1,5	1,4	1,1	0,6	
8 000	*****	t*****	t*****	*****	t*****	t*****	*****	*****	*****	1,5	1,4	1,3	1,0	0,6	
9 000	*****	t*****	t*****	*****	t*****	t*****	*****	*****	*****	1,4	1,3	1,2	0,9	0,5	
10 000	*****	t*****	t*****	*****	t*****	t*****	*****	*****	*****	*****	1,3	1,2	0,9	0,5	
12 500	*****	t*****	t*****	*****	t*****	t*****	*****	*****	*****	*****	*****	1,0	0,8	0,5	
15 000	*****	t*****	t*****	*****	t*****	t*****	*****	*****	*****	*****	*****	*****	0,7	0,4	
20 000	*****	t*****	t*****	*****	t*****	t*****	*****	*****	*****	*****	*****	*****	*****	0,4	

NOTE: Pour bien utiliser ces tables, veuillez consulter la documentation sur les microdonnées.

Exemple 3 : Estimations de différences entre des agrégats ou des pourcentages

Supposons qu'un utilisateur estime que $2\,478\,643 / 7\,055\,176 = 35,1\%$ la proportion des femmes qui ont un budget pour leur ménage déclarent toujours respecter ce budget, par rapport à $2\,402\,986 / 6\,178\,565 = 38,9\%$ la proportion des hommes qui ont un budget pour leur ménage déclarent toujours respecter ce budget. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

- 1) L'utilisation de la table des coefficients de variation du CANADA de la même façon que celle décrite dans l'exemple 2, donne un CV de l'estimation pour les femmes de 3,0 %, et un CV de l'estimation pour les hommes de 2,8 %.

- 1) En utilisant la règle 3, l'erreur-type d'une différence ($\hat{d} = \hat{X}_1 - \hat{X}_2$) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1 \alpha_1)^2 + (\hat{X}_2 \alpha_2)^2}$$

où \hat{X}_1 est l'estimation 1 (femmes), \hat{X}_2 , est l'estimation 2 (hommes), et α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et de \hat{X}_2 respectivement.

C'est-à-dire que l'erreur-type de la différence $\hat{d} = 0,351 - 0,389 = -0,038$ est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0,351)(0,030)]^2 + [(0,389)(0,028)]^2} \\ &= \sqrt{(0,000111) + (0,000119)} \\ &= 0,015\end{aligned}$$

- 2) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,015 / (-0,038) = -0,395$
- 3) Le coefficient de variation approximatif de la différence entre les estimations est donc 39,5 %. La différence entre les estimations est considérée inacceptable et Statistique Canada recommande de ne pas publier cette estimation. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre F (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les utilisateurs subséquents en garde contre les hauts taux d'erreur associés à l'estimation.

Exemple 4 : Estimations de rapports

Supposons que l'utilisateur estime que 2 478 643 le nombre de femmes qui ont un budget pour leur ménage déclarent toujours respecter ce budget, par rapport à 2 402 986 hommes en ont un et le respectent. L'utilisateur est intéressé à comparer l'estimation des femmes à celles des hommes, sous la forme d'un rapport. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation de rapport, où le numérateur de l'estimation (\hat{X}_1) est le nombre de femmes ayant un budget pour leur ménage qui déclarent toujours respecter ce budget. Le dénominateur de l'estimation (\hat{X}_2) est le nombre d'hommes qui ont un budget et qui déclarent toujours le respecter.
- 2) Reportez-vous à la table des coefficients de variation pour le CANADA.
- 3) Le numérateur de cette estimation de rapport est 2 478 643. Le chiffre qui s'en rapproche le plus est 2 000 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 3,5 %.
- 4) Le dénominateur de cette estimation de rapport est 2 402 986. Le chiffre qui s'en rapproche le plus est 2 000 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 3,5 %.
- 5) Le coefficient de variation approximatif de l'estimation du rapport est donc donné par la règle 4, qui est :

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^2 + \alpha_2^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement. C'est-à-dire que :

$$\begin{aligned} \alpha_{\hat{R}} &= \sqrt{(0,035)^2 + (0,035)^2} \\ &= \sqrt{0,001225 + 0,001225} \\ &= 0,049 \end{aligned}$$

- 6) Le rapport obtenu entre les femmes et les hommes qui ont un budget pour leur ménage et qui déclarent toujours le respecter est 2 478 643 / 2 402 986, c'est-à-dire 1,03 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1). Le coefficient de variation de cette estimation est de 4,9 %, ce qui fait qu'on peut la diffuser sans réserve.

Enquête canadienne sur les capacités financières, 2009

Tables de la variabilité d'échantillonnage approximative - Canada - Tous les âges

NUMÉRATEUR DU POURCENTAGE (en milliers)	ESTIMATED PERCENTAGE														
	0,1%	1,0%	2,0%	5,0%	10,0%	15,0%	20,0%	25,0%	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%	
1	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	90,0	52,0	
2	#####	#####	#####	#####	#####	#####	#####	#####	97,2	93,7	90,0	82,2	63,7	36,7	
3	94,8	94,4	93,9	92,5	90,0	87,5	84,9	82,2	79,4	76,5	73,5	67,1	52,0	30,0	
4	82,1	81,8	81,3	80,1	78,0	75,8	73,5	71,2	68,7	66,2	63,7	58,1	45,0	26,0	
5	73,5	73,1	72,8	71,6	69,7	67,8	65,7	63,7	61,5	59,3	56,9	52,0	40,3	23,2	
6	67,1	66,8	66,4	65,4	63,7	61,9	60,0	58,1	56,1	54,1	52,0	47,4	36,7	21,2	
7	62,1	61,8	61,5	60,5	58,9	57,3	55,6	53,8	52,0	50,1	48,1	43,9	34,0	19,6	
...	
...	
...	
750	*****	t*****	t*****	5,8	5,7	5,5	5,4	5,2	5,0	4,8	4,6	4,2	3,3	1,9	
1 000	*****	t*****	t*****	5,1	4,9	4,8	4,6	4,5	4,3	4,2	4,0	3,7	2,8	1,6	
1 500	*****	t*****	t*****	*****	4,0	3,9	3,8	3,7	3,6	3,4	3,3	3,0	2,3	1,3	
2 000	*****	t*****	t*****	*****	3,5	3,4	3,3	3,2	3,1	3,0	2,8	2,6	2,0	1,2	
3 000	*****	t*****	t*****	*****	t*****	2,8	2,7	2,6	2,5	2,4	2,3	2,1	1,6	0,9	
4 000	*****	t*****	t*****	*****	t*****	t*****	2,3	2,3	2,2	2,1	2,0	1,8	1,4	0,8	
5 000	*****	t*****	t*****	*****	t*****	t*****	2,1	2,0	1,9	1,9	1,8	1,6	1,3	0,7	
6 000	*****	t*****	t*****	*****	t*****	t*****	*****	1,8	1,8	1,7	1,6	1,5	1,2	0,7	
7 000	*****	t*****	t*****	*****	t*****	t*****	*****	t*****	1,6	1,6	1,5	1,4	1,1	0,6	
8 000	*****	t*****	t*****	*****	t*****	t*****	*****	t*****	t*****	1,5	1,4	1,3	1,0	0,6	
9 000	*****	t*****	t*****	*****	t*****	t*****	*****	t*****	t*****	1,4	1,3	1,2	0,9	0,5	
10 000	*****	t*****	t*****	*****	t*****	t*****	*****	t*****	t*****	*****	1,3	1,2	0,9	0,5	
12 500	*****	t*****	t*****	*****	t*****	t*****	*****	t*****	t*****	*****	t*****	1,0	0,8	0,5	
15 000	*****	t*****	t*****	*****	t*****	t*****	*****	t*****	t*****	*****	t*****	t*****	0,7	0,4	
20 000	*****	t*****	t*****	*****	t*****	t*****	*****	t*****	t*****	*****	t*****	t*****	t*****	0,4	

NOTE: Pour bien utiliser ces tables, veuillez consulter la documentation sur les microdonnées.

Exemple 5 : Estimations de différences entre des rapports

Supposons qu'un utilisateur estime que le rapport entre les femmes qui ont un budget pour leur ménage et qui déclarent toujours le respecter et les hommes qui ont un budget pour leur ménage et qui déclarent toujours le respecter est de 0,83 pour la Nouvelle-Écosse, tandis qu'il est de 1,18 pour l'Alberta. L'utilisateur est intéressé à comparer les deux rapports pour voir s'il y a une différence statistique entre ceux-ci. Comment l'utilisateur détermine-t-il le coefficient de variation de cette différence?

- 1) Tout d'abord calculez le coefficient de variation approximatif pour le rapport de la Nouvelle-Écosse (\hat{R}_1) et le rapport de l'Alberta (\hat{R}_2) tel qu'il est décrit dans l'exemple 4. Le CV approximatif pour le rapport de la Nouvelle-Écosse est de 19,4 % et 15,0 % pour celui l'Alberta, de.
- 2) En utilisant la règle 3, l'erreur-type d'une différence ($\hat{d} = \hat{R}_1 - \hat{R}_2$) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{R}_1 \alpha_1)^2 + (\hat{R}_2 \alpha_2)^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{R}_1 et \hat{R}_2 respectivement. C'est-à-dire que l'erreur-type de la différence $\hat{d} = 0,83 - 1,18 = -0,35$ est :

$$\begin{aligned} \sigma_{\hat{d}} &= \sqrt{[(0,83)(0,194)]^2 + [(1,18)(0,150)]^2} \\ &= \sqrt{(0,0259) + (0,0313)} \\ &= 0,239 \end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,239 / (-0,35) = -0,683$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 68,3 %. La différence entre les estimations est considérée inacceptable et Statistique Canada recommande de ne pas la publier. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre F (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les utilisateurs subséquents en garde contre les hauts taux d'erreur associés à l'estimation.

Enquête canadienne sur les capacités financières, 2009

Tables de la variabilité d'échantillonnage approximative - Nouvelle-Écosse - Tous les âges

NUMÉRATEUR DU POURCENTAGE (en milliers)	POURCENTAGE ESTIMÉ													
	0,1%	1,0%	2,0%	5,0%	10,0%	15,0%	20,0%	25,0%	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%
1	*****	110,9	110,3	108,6	105,7	102,7	99,7	96,5	93,2	89,8	86,3	78,8	61,0	35,2
2	*****	78,4	78,0	76,8	74,7	72,6	70,5	68,2	65,9	63,5	61,0	55,7	43,2	24,9
...
...
...
50	*****	*****	*****	*****	14,9	14,5	14,1	13,6	13,2	12,7	12,2	11,1	8,6	5,0
55	*****	*****	*****	*****	14,3	13,9	13,4	13,0	12,6	12,1	11,6	10,6	8,2	4,8
60	*****	*****	*****	*****	13,6	13,3	12,9	12,5	12,0	11,6	11,1	10,2	7,9	4,5
65	*****	*****	*****	*****	13,1	12,7	12,4	12,0	11,6	11,1	10,7	9,8	7,6	4,4
70	*****	*****	*****	*****	12,6	12,3	11,9	11,5	11,1	10,7	10,3	9,4	7,3	4,2
75	*****	*****	*****	*****	12,2	11,9	11,5	11,1	10,8	10,4	10,0	9,1	7,0	4,1
80	*****	*****	*****	*****	*****	11,5	11,1	10,8	10,4	10,0	9,6	8,8	6,8	3,9
85	*****	*****	*****	*****	*****	11,1	10,8	10,5	10,1	9,7	9,4	8,5	6,6	3,8
90	*****	*****	*****	*****	*****	10,8	10,5	10,2	9,8	9,5	9,1	8,3	6,4	3,7
95	*****	*****	*****	*****	*****	10,5	10,2	9,9	9,6	9,2	8,9	8,1	6,3	3,6
100	*****	*****	*****	*****	*****	10,3	10,0	9,6	9,3	9,0	8,6	7,9	6,1	3,5
125	*****	*****	*****	*****	*****	*****	8,9	8,6	8,3	8,0	7,7	7,0	5,5	3,2
150	*****	*****	*****	*****	*****	*****	8,1	7,9	7,6	7,3	7,0	6,4	5,0	2,9
200	*****	*****	*****	*****	*****	*****	*****	*****	6,6	6,4	6,1	5,6	4,3	2,5
250	*****	*****	*****	*****	*****	*****	*****	*****	*****	5,7	5,5	5,0	3,9	2,2
300	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	5,0	4,5	3,5	2,0
350	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	4,2	3,3	1,9
400	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	3,1	1,8
450	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,9	1,7
500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,7	1,6

NOTE: Pour bien utiliser ces tables, veuillez consulter la documentation sur les microdonnées.

Enquête canadienne sur les capacités financières, 2009

Tables de la variabilité d'échantillonnage approximative - Alberta - Tous les âges

NUMÉRATEUR DU POURCENTAGE (en milliers)	POURCENTAGE ESTIMÉ													
	0,1%	1,0%	2,0%	5,0%	10,0%	15,0%	20,0%	25,0%	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%
1	146,7	146,1	145,3	143,1	139,3	135,3	131,3	127,1	122,8	118,4	113,7	103,8	80,4	46,4
2	103,8	103,3	102,8	101,2	98,5	95,7	92,8	89,9	86,8	83,7	80,4	73,4	56,9	32,8
3	*****	84,3	83,9	82,6	80,4	78,1	75,8	73,4	70,9	68,3	65,7	59,9	46,4	26,8
...
...
...
125	*****	*****	*****	12,8	12,5	12,1	11,7	11,4	11,0	10,6	10,2	9,3	7,2	4,2
150	*****	*****	*****	*****	11,4	11,1	10,7	10,4	10,0	9,7	9,3	8,5	6,6	3,8
200	*****	*****	*****	*****	9,8	9,6	9,3	9,0	8,7	8,4	8,0	7,3	5,7	3,3
250	*****	*****	*****	*****	8,8	8,6	8,3	8,0	7,8	7,5	7,2	6,6	5,1	2,9
300	*****	*****	*****	*****	*****	7,8	7,6	7,3	7,1	6,8	6,6	6,0	4,6	2,7
350	*****	*****	*****	*****	*****	7,2	7,0	6,8	6,6	6,3	6,1	5,5	4,3	2,5
400	*****	*****	*****	*****	*****	6,8	6,6	6,4	6,1	5,9	5,7	5,2	4,0	2,3
450	*****	*****	*****	*****	*****	*****	6,2	6,0	5,8	5,6	5,4	4,9	3,8	2,2
500	*****	*****	*****	*****	*****	*****	5,9	5,7	5,5	5,3	5,1	4,6	3,6	2,1
750	*****	*****	*****	*****	*****	*****	*****	*****	4,5	4,3	4,2	3,8	2,9	1,7
1 000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	3,6	3,3	2,5	1,5
1 500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,1	1,2
2 000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1,0

NOTE: Pour bien utiliser ces tables, veuillez consulter la documentation sur les microdonnées.

10.2 Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance

Bien que les coefficients de variation soient beaucoup utilisés, l'intervalle de confiance d'une estimation est une mesure plus intuitivement significative de l'erreur d'échantillonnage. Un intervalle de confiance constitue une déclaration du niveau de confiance selon laquelle la valeur vraie pour la population se situe à l'intérieur d'une gamme précisée de valeurs. Par exemple, un intervalle de confiance de 95 % peut être décrit comme suit :

Si l'échantillonnage de la population est répété indéfiniment, chaque échantillon menant à un nouvel intervalle de confiance pour une estimation, l'intervalle englobera alors dans 95 % des échantillons la valeur vraie de la population.

En utilisant l'erreur-type d'une estimation, des intervalles de confiance pour des estimations peuvent être obtenues en partant de l'hypothèse qu'aux termes d'un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique donnée de la population se répartiront normalement autour de la valeur vraie de la population. Selon cette hypothèse, 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. Ces différents degrés de confiance sont désignés sous le nom de niveaux de confiance.

Des intervalles de confiance pour une estimation \hat{X} sont généralement exprimés sous forme de deux chiffres, un inférieur et un supérieur à l'estimation, comme étant $(\hat{X} - k, \hat{X} + k)$, où k est déterminé suivant le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

Des intervalles de confiance pour une estimation peuvent être calculés directement à partir des tables de variabilité d'échantillonnage approximative, en déterminant d'abord à partir de la table appropriée le coefficient de variation de l'estimation \hat{X} , puis en utilisant la formule suivante pour le convertir à un intervalle de confiance ($IC_{\hat{x}}$) :

$$IC_{\hat{x}} = (\hat{X} - t\hat{X}\alpha_{\hat{x}}, \hat{X} + t\hat{X}\alpha_{\hat{x}})$$

où $\alpha_{\hat{x}}$ est le coefficient de variation déterminé de \hat{X} , et

- $t = 1$ si l'on désire un intervalle de confiance de 68 %;
- $t = 1,6$ si l'on désire un intervalle de confiance de 90 %;
- $t = 2$ si l'on désire un intervalle de confiance de 95 %;
- $t = 2,6$ si l'on désire un intervalle de confiance de 99 %.

Nota : Les lignes directrices pour la diffusion des estimations s'appliquent également aux intervalles de confiance. S'il est impossible, par exemple, de diffuser une estimation, on ne peut alors pas non plus communiquer un intervalle de confiance.

10.2.1 Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance

Un intervalle de confiance de 95 % pour la proportion estimée d'adultes canadiens ayant un budget pour leur ménage qui déclarent toujours respecter ce budget (d'après l'exemple 2 à la section 10.1.1) serait calculé comme suit :

$$\hat{X} = 36,9 \% \text{ (ou exprimé sous forme de proportion } 0,369)$$

$$t = 2$$

$$\alpha_{\hat{x}} = 1,9 \% \text{ (0,019 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.}$$

$$CI_{\hat{x}} = \{0,369 - (2) (0,369) (0,019), 0,369 + (2) (0,369) (0,019)\}$$

$$CI_{\hat{x}} = \{0,369 - 0,014, 0,369 + 0,014\}$$

$$CI_{\hat{x}} = \{0,355, 0,383\}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 35,5 % et 38,3 % des adultes canadiens ayant un budget pour leur ménage déclarent toujours respecter ce budget.

10.3 Comment utiliser les tables de coefficients de variation pour effectuer un test t

Des erreurs-types peuvent aussi être utilisées pour effectuer des tests d'hypothèses, une procédure destinée à distinguer des paramètres d'une population à l'aide d'estimations d'un échantillon. Ces estimations peuvent être des chiffres, des moyennes, des pourcentages, des rapports, etc. Les tests peuvent être effectués à divers niveaux de signification, où un niveau de signification est la probabilité de conclure que les caractéristiques sont différentes quand, en fait, elles sont identiques.

Supposons que \hat{X}_1 et \hat{X}_2 sont des estimations d'un échantillon pour deux caractéristiques qui nous intéressent. Supposons également que l'erreur-type de la différence $\hat{X}_1 - \hat{X}_2$ est $\sigma_{\hat{d}}$.

Si $t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ se situe entre -2 et 2, aucune conclusion à propos de la différence entre les

caractéristiques n'est alors justifiée au niveau de signification de 5 %. Si, cependant, ce rapport est inférieur à -2 ou supérieur à +2, la différence observée est significative au niveau de 0,05. C'est-à-dire que la différence entre les estimations est significative.

10.3.1 Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t

Supposons que l'utilisateur désire tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y pas de différence entre les proportions de femmes qui ont un budget

pour leur ménage qui déclarent toujours respecter ce budget et la proportion d'hommes qui ont un budget pour leur ménage qui déclarent toujours respecter ce budget. D'après l'exemple 3 à la section 10.1.1, il s'est avéré que l'erreur-type de la différence entre ces deux estimations était 0,015. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_d} = \frac{0,351 - 0,389}{0,015} = \frac{-0,038}{0,015} = -2,53$$

Puisque $t = -2,53$ est inférieur à -2 , il faut en conclure qu'il existe une différence significative entre les deux estimations au niveau de signification de 0,05.

10.4 Coefficients de variation pour des estimations quantitatives

Il faudrait produire des tables spéciales afin de déterminer l'erreur d'échantillonnage d'estimations quantitatives, ce qui n'a pas été fait, parce que la plupart des variables pour l'ECCF sont principalement de nature catégorique.

En général cependant, le coefficient de variation d'un total quantitatif sera supérieur au coefficient de variation de l'estimation de la catégorie correspondante (c'est-à-dire l'estimation du nombre de personnes retenues dans l'estimation quantitative). S'il est impossible de diffuser l'estimation de la catégorie correspondante, on ne pourra pas non plus communiquer l'estimation quantitative. Par exemple, le coefficient de variation du nombre total de comptes de banque personnels détenus par des femmes serait supérieur au coefficient de variation de la proportion correspondante de femmes ayant un compte de banque personnel. Si, par conséquent, le coefficient de variation de la proportion est inacceptable (rendant la proportion non diffusable), il en sera de même du coefficient de variation de l'estimation quantitative correspondante (rendant cette estimation quantitative non diffusable).

Des coefficients de variation de telles estimations peuvent être calculés, au besoin, pour une estimation précise à l'aide d'une technique appelée pseudo-répétition, ce qui veut dire diviser les enregistrements inclus dans les fichiers de microdonnées en sous-groupes (ou répétitions) et déterminer la variation à l'intérieur de l'estimation de répétition en répétition. Les utilisateurs qui désirent calculer des coefficients de variation pour des estimations quantitatives peuvent communiquer avec Statistique Canada afin d'obtenir des conseils sur l'allocation d'enregistrements à des répétitions appropriées et sur les formules à employer à l'intérieur de ces calculs.

10.5 Tables de coefficients de variation

Consulter le fichier ECCF2009_CVTabF.pdf pour les tables de coefficients de variation.

10.6 Méthode bootstrap moyenne pour estimer la variance

Pour juger de la qualité d'une estimation et en calculer le CV, on doit d'abord établir l'écart-type. On a aussi besoin de l'écart-type d'estimations pour les intervalles de confiance. Dans l'ECCF, 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 variance qu'on puisse appliquer. Il faut donc procéder par approximation. Si on se sert de la méthode bootstrap moyenne, c'est qu'il faut tenir compte de l'échantillonnage et de l'étalonnage au moment d'estimer la variance. C'est ce que fait la méthode bootstrap moyenne et, avec le programme Bootvar dont il sera question à la sous-section suivante, l'utilisateur dispose déjà d'une méthode d'une application assez facile.

Dans l'ECCF, on emploie la méthode bootstrap moyenne que décrit W. Yung (Yung, W. (1997b) Estimation de la variance des fichiers de microdonnées à grande diffusion, *Symposium 1997, Nouvelles orientations pour les enquêtes et les recensements*, Statistique Canada).

On prélève indépendamment sur chaque strate un échantillon aléatoire simple de $(n - 1)$ des unités n de l'échantillon. À noter que, comme la sélection s'opère avec remise, une unité peut être choisie plusieurs fois. On reprend l'opération R fois pour ainsi obtenir R échantillons bootstrap. On calcule une valeur moyenne initiale de pondération bootstrap à partir des R échantillons pour chaque unité échantillonnée de la strate. Toute l'opération (prélèvement d'échantillons aléatoires simples et repondération de chaque strate) a lieu à B reprises – B étant une valeur élevée –, ce qui donne B valeurs initiales de pondération bootstrap. Dans l'ECCF, R est de 20 et B de 250 en temps normal pour 250 valeurs de pondération.

On repondère alors comme dans la pondération courante par correction de non-réponse, étalonnage, etc. On obtient 250 valeurs de pondération bootstrap moyenne finale pour chaque unité de l'échantillon. On rapporte la variation des 250 estimations possibles correspondant aux 250 valeurs bootstrap moyennes à la variance de l'estimateur par pondération courante; la variance peut être estimée par ce moyen. Pour plusieurs raisons, l'utilisateur pourrait vouloir calculer le CV des estimations par la méthode bootstrap moyenne. En voici quelques-unes :

- Premièrement, s'il désire des estimations à un niveau géographique inférieur à celui de la province (au niveau des régions urbaines et rurales, par exemple), les Tables de la variabilité d'échantillonnage approximative ne peuvent suffire. Par le programme d'estimation de variance bootstrap, on peut recourir aux techniques d'estimation de domaine pour dégager le CV de ces estimations.
- Deuxièmement, si l'utilisateur veut une analyse plus fine par régression linéaire ou logistique pour l'estimation des coefficients, les Tables de la variabilité d'échantillonnage approximative ne lui donneront pas les CV correspondants en toute précision. Bien qu'un certain nombre de progiciels statistiques permettent d'intégrer la pondération d'échantillonnage à l'analyse, les variances produites ne tiennent souvent pas tout à fait compte du plan d'échantillonnage ni ne traduisent l'étalonnage de pondération contrairement à ce qui se passe dans le programme d'estimation de variance bootstrap.
- Troisièmement, dans le cas des estimations de variables quantitatives, il faut consulter des tableaux séparés pour établir l'erreur d'échantillonnage.

10.7 Progiciels statistiques pour estimer la variance

Statistique Canada a élaboré un programme qui peut livrer des estimations de variance bootstrap moyenne. C'est le programme Bootvar.

Celui-ci est disponible en format SAS et SPSS. Il est formé de macro instructions d'estimation de variance pour les totaux, les rapports et les différences entre rapports, ainsi que de régression linéaire et logistique.

Le Bootvar peut être téléchargé à partir du site Internet des Centres de données de recherche (CDR) de Statistique Canada. Il faut accepter la licence d'adhésion automatique Bootvar avant de pouvoir lire les fichiers.

SAS: http://www.statcan.gc.ca/rdc-cdr/bootvar_sas-fra.htm

SPSS: http://www.statcan.gc.ca/rdc-cdr/bootvar_spss-fra.htm

11.0 Ponderation

Pour le fichier de microdonnées, des poids statistiques ont été placés pour chaque enregistrement pour représenter le nombre de personnes que chaque enregistrement échantillonné représente. Un poids a été calculé pour chaque personne.

La pondération pour l'Enquête canadienne sur les capacités financières (ECCF) comprend plusieurs étapes :

- calcul du poids de base,
- le facteur de compensation pour la non-réponse,
- l'élimination des enregistrements hors du champ de l'enquête,
- un ajustement pour la sélection d'une personne dans le ménage, et finalement
- un ajustement pour rendre les estimations de la population cohérentes avec les totaux connus, soit province-âge-sexe des projections démographiques du recensement pour les personnes de 18 ans et plus.

11.1 Procédures de pondération

1. Calcul du poids pour le numéro de téléphone

Chaque numéro de téléphone dans l'échantillon a reçu un poids de base, W_1 , égal à l'inverse de sa probabilité de sélection.

$$W_1 = \left(\frac{\text{Nombre total de numéros de téléphone pouvant être échantillonnés pour province}}{\text{Nombre de numéros de téléphone échantillonnés pour province}} \right)$$

Il y avait 68 462 numéros de téléphone dans l'échantillon avec des poids attribués.

2. Ajustement pour les numéros de téléphone non résolus

Il y a eu 5 191 numéros de téléphone qui n'ont pas été résolus, laissant 63 271 numéros de téléphone résolus. Il n'a pas été déterminé si ces numéros de téléphone non résolus appartenaient à un ménage, une entreprise ou étaient hors du champ de l'enquête. Chaque numéro de téléphone avait un indicateur signalant si l'on s'attendait que le numéro serait une résidence, une entreprise ou de genre inconnu et un autre indicateur signalant s'il avait été éliminé avant la collecte comme étant hors service ou une entreprise. L'ajustement pour les numéros de téléphone non résolus a été fait par province, par genre de ligne prévu, et selon si le numéro a été envoyé aux intervieweurs ou pas.

Pour chaque province-genre de ligne prévu-envoyé,

$$W_2 = W_1 * \left(\frac{\sum W_1 \text{ pour numéros de téléphone résolus} + \sum W_1 \text{ pour numéros de téléphone non résolus}}{\sum W_1 \text{ pour numéros de téléphone résolus}} \right)$$

3. Éliminer les numéros de téléphone hors du champ de l'enquête

Les numéros de téléphone correspondant à une entreprise, hors service ou hors du champ de l'enquête, tels les chalets, ont été éliminés après l'ajustement pour la non-réponse du téléphone. Veuillez prendre note que si les données pour le ménage ou pour une personne étaient présentes alors le numéro de téléphone était considéré comme étant celui d'un ménage. Il y

avait 38 040 numéros de téléphone qui étaient hors du champ de l'enquête et 25 231 numéros qui appartenaient à un ménage.

4. Ajustement pour la non-réponse pour le nombre de lignes téléphoniques dans le ménage

Le nombre de lignes téléphoniques dans le ménage a été calculé. Si le nombre de lignes téléphoniques différentes dans le ménage ne pouvait être calculé mais que des données existaient pour le ménage ou pour une personne, la valeur de 1 a été imputée pour conserver de bonnes données. Après l'imputation, il restait 7 384 numéros de téléphone pour lesquels le nombre de lignes manquait toujours. Donc, il y avait 17 847 ménages dont l'information a été calculée ou imputée. L'ajustement a été fait au niveau province.

$$W_3 = W_2 * \left(\frac{\sum W_2 \text{ pour ménages avec nombre de lignes} + \sum W_2 \text{ pour ménages avec nombre de lignes manquant}}{\sum W_2 \text{ pour ménages avec nombre de lignes}} \right)$$

5. Calcul du poids des ménages avec l'ajustement pour les lignes téléphoniques multiples

Les poids des ménages avec plus d'une ligne téléphonique (avec différents numéros de téléphone) ont été ajustés à la baisse pour prendre en compte le fait que ces ménages avaient une plus grande probabilité de sélection. Le poids pour chaque ménage a été divisé par le nombre de lignes téléphoniques résidentielles distinctes (maximum 4) qui desservaient le ménage. L'ajustement a été fait au niveau province.

$$W_4 = \left(\frac{W_3}{\text{Nombre de lignes téléphoniques dans le champ d'enquête dans le ménage}} \right)$$

6. Ajustement pour les ménages non-répondants

Les répondants du ménage ont répondu aux questions dans la liste des membres. Si ces questions n'ont pas été complétées, refusées ou remplies partiellement, le ménage a été considéré comme non-répondant. Il y a eu 71 non-répondants. Donc 17 776 ménages pondérés dans le champ d'enquête ont été utilisés et ajustés par province.

$$W_5 = W_4 * \left(\frac{\sum W_4 \text{ pour ménages répondants} + \sum W_4 \text{ pour ménages non répondants}}{\sum W_4 \text{ pour ménages répondants}} \right)$$

7. Assigner les poids de ménages aux personnes sélectionnées

Tous les individus sélectionnés des ménages répondants dans l'enquête ayant une liste complète des membres (c.-à-d. aucun âge manquant) reçoivent les poids appartenant aux leurs ménages.

$$W_6 = W_5$$

8. Calcul du sous-poids de la personne sélectionnée

Les poids de tous les individus sélectionnés sont augmentés à partir de l'information dans la liste des membres dans le but de représenter le nombre de personnes dans le ménage qui sont admissibles à être sélectionné (l'âge de 18 ou plus).

$$W_7 = W_6 * (\text{nombre de personnes dans le ménage})$$

9. Ajustement pour les individus non-répondants

Le fichier de données comprend les enregistrements des répondants individuels qui ont donné suffisamment d'information aux questions demandées. Il y a eu 2 257 non-répondants.

Donc, nous avons utilisé 15 519 poids des individus dans le champ d'enquête et ajusté selon la province, le groupe d'âge tiré de la liste des membres (18 à 24, 25 à 44, 45 à 64, 65 ans et plus) et le sexe.

$$W_8 = W_7 * \left(\frac{\sum W_7 \text{ pour personnes répondante } s + \sum W_7 \text{ pour personnes non répondante } s}{\sum W_7 \text{ pour personnes répondante } s} \right)$$

10. Ajustement pour totaux externes

On a calculé un ajustement aux poids des personnes pour s'assurer que les estimations de la population étaient cohérentes avec des totaux externes de population pour les personnes de 18 ans et plus. Ceci est connu comme la post-stratification. Les totaux externes suivants, comme projetés pour février 2009, ont été utilisés :

- 1) Totaux de population selon la province, le sexe et les groupes d'âge suivants : 18 à 24, 25 à 29, 30 à 34, 35 à 39, 40 à 44, 45 à 49, 50 à 54, 55 à 59, 60 à 64, 65 à 69 et 70 ans et plus.
- 2) Totaux de population ayant 18 ans ou plus vivant dans les régions métropolitaines de recensement.

La méthode appelée GREG, soit la régression généralisée, a été utilisée pour modifier les poids afin de s'assurer que les estimations de l'enquête concordaient avec les totaux externes simultanément pour les deux dimensions.

Après cette étape, les poids obtenus pour les personnes sont considérés comme finals et ont été mis au fichier.

12.0 Questionnaires

Le questionnaire de l'Enquête canadienne sur les capacités financières (ECCF) a servi à recueillir des données pour l'enquête en 2009. Le fichier ECCF2009_QuestF.pdf renferme le questionnaire français

13.0 Cliché d'enregistrement à valeurs univariées

Consulter le fichier ECCF2009_LvCds.pdf pour le cliché d'enregistrement à valeurs univariées