



Guide de l'utilisateur des microdonnées

**ENQUÊTE DE SURVEILLANCE DE L'USAGE DU
TABAC AU CANADA**

CYCLE 1

FÉVRIER - JUIN 2002



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	13
5.1	Champ de l'enquête	13
5.2	Stratification	13
5.3	Plan de sondage et répartition de l'échantillon	13
5.4	Tirage de l'échantillon	14
6.0	Collecte des données	17
6.1	Conception du questionnaire	17
6.2	Collecte et vérification des données	17
7.0	Traitement des données	19
7.1	Saisie des données	19
7.2	Vérification	19
7.3	Création de variables calculées	19
7.4	Pondération	19
7.5	Suppression des renseignements confidentiels	20
8.0	Qualité des données	21
8.1	Taux de réponse des ménages - février à juin 2002	22
8.2	Taux de réponse des personnes - février à juin 2002	23
8.3	Erreurs dans l'enquête	25
8.4	Non-réponse totale	25
8.5	Non-réponse partielle	25
8.6	Couverture	25
8.7	Mesure de l'erreur d'échantillonnage	26
9.0	Lignes directrices pour la totalisation, l'analyse et la diffusion de données	27
9.1	Lignes directrices sur l'arrondissement d'estimations	27
9.2	Lignes directrices sur la pondération d'un échantillon pour une totalisation	28
9.2.1	Définitions de types d'estimations : catégoriques et quantitatives	28
9.2.2	Totalisation d'estimations catégoriques	29
9.2.3	Totalisation d'estimations quantitatives	29
9.3	Lignes directrices pour l'analyse statistique de données	30
9.4	Lignes directrices sur la diffusion de coefficients de variation	31

10.0	Tables de variabilité d'échantillonnage approximative	33
10.1	Comment utiliser les tables de CV pour les estimations catégoriques	35
10.1.1	Exemples d'utilisation des tables de CV pour des estimations catégoriques	37
10.2	Comment utiliser les tables de CV pour obtenir des limites de confiance	39
10.2.1	Exemple d'utilisation de tables de CV pour obtenir des limites de confiance	40
10.3	Comment utiliser les tables de CV pour effectuer un test t	41
10.3.1	Exemple d'utilisation des tables de CV pour effectuer un test t	41
10.4	Coefficients de variation d'estimations quantitatives	41
10.5	Seuils pour la diffusion des estimations pour l'Enquête de surveillance de l'usage du tabac au Canada - Fichier ménages	42
10.6	Seuils pour la diffusion des estimations pour l'Enquête de surveillance de l'usage du tabac au Canada - Fichier personnes	43
10.7	Tables de CV - Fichier ménages	44
10.8	Tables de CV - Fichier personnes	44
11.0	Pondération	45
11.1	Procédures de pondération pour les Fichiers ménages et personnes	45
11.2	Procédures de pondération pour le Fichier ménages	46
11.3	Procédures de pondération pour le Fichier personnes	47
12.0	Questionnaire	49
13.0	Disposition d'enregistrement avec fréquences univariés	51
13.1	Disposition d'enregistrement avec fréquences univariés – Fichier ménages	51
13.2	Disposition d'enregistrement avec fréquences univariés – Fichier personnes	51

1.0 Introduction

L'Enquête de surveillance de l'usage du tabac au Canada (ESUTC) a été menée par Statistique Canada entre février et juin 2002, avec la collaboration et l'appui de Santé Canada. Ce manuel a été rédigé en vue de faciliter la manipulation du fichier de microdonnées portant sur les résultats de l'enquête.

Veuillez adresser toutes les demandes concernant l'ensemble des données ou leurs utilisations à :

Statistique Canada

Services à la clientèle
Division des enquêtes spéciales
Téléphone : (613) 951-3321 ou appelez sans frais : 1 800 461-9050
Télécopieur : (613) 951-4527
Courriel : des@statcan.ca

Elizabeth Majewski
Division des enquêtes spéciales
2^e étage, Immeuble Principal, pré Tunney
Ottawa (Ontario) K1A 0T6
Téléphone : (613) 951-4584
Télécopieur : (613) 951-0562
Courriel : elizabeth.majewski@statcan.ca

Santé Canada

Murray Kaiserman
Bureau de la recherche, de la surveillance et de l'évaluation
Programme de la lutte au tabagisme
Direction générale de la santé environnementale et de la sécurité des consommateurs
Immeuble MacDonald, ind. l'adr. 3507C
123, rue Slater, pièce A723
Ottawa (Ontario) K1A 0K9
Téléphone : (613) 954-5851
Télécopieur : (613) 954-2292
Courriel : Murray_Kaiserman@hc-sc.gc.ca

Anne Zaborski
Bureau de la recherche, de la surveillance et de l'évaluation
Programme de la lutte au tabagisme
Direction générale de la santé environnementale et de la sécurité des consommateurs
Immeuble MacDonald, ind. l'adr. 3507C
123, rue Slater, pièce A712
Ottawa (Ontario) K1A 0K9
Téléphone : (613) 954-0152
Télécopieur : (613) 954-2292
Courriel : Anne_Zaborski@hc-sc.gc.ca

IL EST IMPORTANT POUR LES UTILISATEURS DE SE FAMILIARISER AVEC LE CONTENU DE CE DOCUMENT AVANT DE PUBLIER OU DE DIFFUSER TOUTES ESTIMATIONS PRODUITES À PARTIR DU FICHIER DE MICRODONNÉES DE L'ENQUÊTE DE SURVEILLANCE DE L'USAGE DU TABAC AU CANADA.

2.0 Contexte

Depuis les années 60, Statistique Canada mène des enquêtes spéciales sur le tabagisme pour Santé Canada sous forme de suppléments à l'Enquête sur la population active du Canada et d'enquêtes téléphoniques à composition aléatoire.

En février 1994, une modification de la législation qui autorisait la diminution des taxes sur les cigarettes a été adoptée. Comme il n'existait aucune donnée d'enquête datant immédiatement avant l'adoption de cette modification, mesurer exactement l'impact de ce changement était difficile pour Santé Canada et pour les analystes intéressés.

Vu le désir de Santé Canada à suivre de près les conséquences des changements législatifs ainsi que l'effet des politiques antitabac sur les comportements des fumeurs, elle a mis sur pied l'Enquête de surveillance de l'usage du tabac au Canada (ESUTC). Cette enquête a pour but de fournir à Santé Canada ainsi qu'à ses partenaires, des données continues et fiables sur l'usage du tabac et sur des sujets connexes.

Depuis 1999, deux fichiers de l'ESUTC ont été publiés annuellement. Le premier contient des données recueillies durant la période de février à juin et l'autre, de juillet à décembre. De plus, un sommaire de l'année est distribué. Le fichier actuel comprend les données pour la période de février à juin 2002.

3.0 Objectifs

Le principal objectif de l'enquête est de fournir en permanence une source de données sur la prévalence de l'usage du tabac permettant de suivre l'évolution de la prévalence. Cet objectif est distinct de celui de l'Enquête nationale sur la santé de la population (ENSP), qui recueille des données sur l'usage du tabac à partir d'un échantillon longitudinal afin de déterminer quelles sont les personnes qui changent leurs habitudes de fumer, les facteurs possibles qui contribuent à ce changement ainsi que les facteurs de risque possibles pour ceux qui commencent à fumer et qui continuent à le faire plus ou moins longtemps. Comme l'ENSP recueille des données tous les deux ans et que celles-ci sont diffusées environ un an après la fin du cycle de collecte, elle ne comble pas les besoins de Santé Canada qui exige une couverture temporelle continue et une diffusion rapide. De plus, les données de l'ENSP ne sont pas suffisamment détaillées dans le cas de la population la plus à risque, soit les 15 à 24 ans.

L'Enquête de surveillance de l'usage du tabac au Canada aidera Santé Canada à observer la prévalence de l'usage du tabac par province et par groupe d'âge -sexe (15 à 19 ans, 20 à 24 ans, 25 à 34 ans, 35 à 44 ans, 45 ans ou plus) deux fois par an. Les données comprises dans ces fichiers représentent les données recueillies pour les six mois de 2002 (de février à juin). Les données continueront à être recueillies de façon constante, selon la disponibilité des fonds.

4.0 Concepts et définitions

Étant donné que l'Enquête de surveillance de l'usage du tabac au Canada a été réalisée par téléphone, on a fait appel à une terminologie facile à comprendre pour éviter les longues explications. L'analyse et l'interprétation des données exigent néanmoins le recours à des concepts et définitions standards. Les questions de l'enquête ont été conçues à la lumière de ces définitions.

Situation actuelle

- | | |
|-------------------------|--|
| 1) Fumeur quotidien : | Personne qui fume actuellement des cigarettes tous les jours. |
| 2) Fumeur occasionnel : | Personne qui fume des cigarettes à l'heure actuelle, mais non tous les jours. |
| 3) Non-fumeur : | Personne qui ne fume pas la cigarette à l'heure actuelle. |
| 4) Fumeur : | Personne qui fume actuellement des cigarettes quotidiennement ou à l'occasion. |

Antécédents

- | | |
|-----------------------------------|---|
| 1) Ancien fumeur : | Personne qui a fumé au moins 100 cigarettes au cours de sa vie, mais qui ne fume pas à l'heure actuelle. |
| 2) Fumeur expérimental : | Personne qui a fumé au moins une cigarette, mais moins de 100 cigarettes, et qui ne fume pas la cigarette à l'heure actuelle. |
| 3) Abstinente à perpétuité : | Personne qui n'a jamais fumé de cigarette de sa vie. |
| 4) Personne ayant déjà fumé : | Personne qui est un fumeur ou un ancien fumeur. |
| 5) Personne n'ayant jamais fumé : | Fumeur expérimental ou abstinente à perpétuité. |

Prévalence de l'usage de la cigarette

Proportion de la population qui fume des cigarettes à l'heure actuelle.

Âge exact en comparaison à l'âge de la liste

L'information sur l'âge des répondants est obtenue à partir de deux sources : de la part d'un répondant qui fournit l'âge de tous les membres du ménage (liste de l'âge) et, à la fin de l'interview, de la part de la personne répondante elle-même à qui on demande la date de naissance. L'âge exact (T_Age) provient de la date de naissance et l'âge de la liste est utilisé uniquement lorsque la date de naissance n'est pas disponible. Habituellement, il y a des divergences entre l'âge de la liste et l'âge exact. Alors que l'âge de la liste détermine l'enchaînement des questions et le traitement des données, les variables de l'âge, dans le fichier final, désignent l'âge exact.

5.0 Méthodologie de l'enquête

Le premier cycle de l'Enquête de surveillance de l'usage du tabac au Canada a été réalisé entre le 1^{er} février et le 30 juin 2002 par enquête téléphonique à composition aléatoire (CA), une technique qui consiste à générer des numéros de téléphone au hasard par ordinateur. L'interview a été menée au téléphone.

5.1 Champ de l'enquête

Toute personne âgée de 15 ans ou plus et vivant au Canada faisait partie de la population cible de l'Enquête de surveillance de l'usage du tabac au Canada, exception faite des personnes suivantes :

- 1) les résidents du Yukon, des Territoires du Nord-Ouest et du Nunavut, et
- 2) les pensionnaires à temps plein d'un établissement institutionnel.

Comme l'enquête se faisait à partir d'un échantillon de numéros de téléphone, les ménages (et donc tous ceux qui en font partie) ne possédant pas de téléphone ont été exclus de l'échantillon. Les personnes n'ayant pas de téléphone représentent moins de 3 % de la population cible. Toutefois, les estimations de l'enquête ont été pondérées afin d'inclure les personnes n'ayant pas de téléphone.

5.2 Stratification

Pour s'assurer que des personnes de partout au Canada soient représentées dans l'échantillon, on a divisé chacune des dix provinces en strates ou aires géographiques. En général, à l'intérieur de chaque province, une strate de régions métropolitaines de recensement (RMR) et une autre de régions autres que des RMR étaient délimitées. L'Île-du-Prince-Édouard ne possédait qu'une seule strate. En Ontario et au Québec, Toronto et Montréal possédaient chacune une troisième strate. Les RMR, régions définies aux fins du recensement, correspondent approximativement à des villes d'au moins 100 000 habitants.

5.3 Plan de sondage et répartition de l'échantillon

Le plan de sondage est un échantillon spécial aléatoire stratifié à deux phases de numéros de téléphone. On utilise un plan de sondage à deux phases afin d'accroître la représentation des personnes appartenant aux groupes d'âge de 15 à 19 ans ou de 20 à 24 ans dans l'échantillon. Durant la première phase, on sélectionne les ménages par CA et durant la seconde, une ou deux personnes (ou aucune) sont sélectionnées d'après la composition du ménage.

Comme l'objectif principal de l'enquête est de produire des estimations fiables dans les dix provinces, on vise **un nombre égal de répondants dans chaque province**. On cherche à obtenir les réponses de 5 000 personnes âgées entre 15 et 24 ans et de 5 000 personnes âgées de 25 ans ou plus à l'échelle du Canada, c'est-à-dire de 500 personnes de chaque groupe d'âge par province. La taille initiale de l'échantillon de numéros de téléphone dépendait du taux de réponse prévu et du taux de succès prévu de la CA (proportion des numéros de téléphone échantillonnés appartenant à des ménages). Pour obtenir les tailles d'échantillon voulues, il a fallu apporter deux rajustements à la méthodologie standard de la CA. Premièrement, les probabilités de sélection au sein du ménage étaient inégales. Deuxièmement, les ménages comptant uniquement des personnes de 25 ans ou plus ont été sous-échantillonnés. On a estimé qu'au total, presque 130 000 numéros de téléphone par année seraient nécessaires pour obtenir les 20 000 répondants par année requis. Cela

suppose un taux de réponse de 75 % et qu'environ 20 % des ménages contiennent des personnes âgées entre 15 et 24 ans. Le taux de succès varie considérablement selon la province, avec une moyenne générale prévue d'environ 40 %.

Pour le 1er cycle de 2002, l'échantillon de l'Ontario a été doublé afin qu'une analyse plus détaillée au niveau infraprovincial soit possible. Les objectifs visés ainsi que les numéros de téléphone requis ont été ajustés en conséquence.

5.4 Tirage de l'échantillon

L'échantillon de l'Enquête de surveillance de l'usage du tabac au Canada a été créé à l'aide d'une méthode améliorée d'échantillonnage par CA, appelée la méthode d'élimination des banques non valides (EBNV). Pour chaque combinaison province-strate, une liste de banques valides (indicatif régional suivi de cinq chiffres) a été dressée à partir des fichiers administratifs des compagnies de téléphone. Une banque valide se définit, pour les fins des enquêtes sociales, comme une banque qui contient au moins un numéro de téléphone résidentiel valide. Ainsi, toutes les banques ayant seulement des numéros de téléphone non assignés, non valides, ou d'affaires ont été exclues de la base de sondage.

Ensuite, un échantillon systématique des banques (avec banques de remplacement) a été sélectionné à l'intérieur de chaque strate. Pour chaque banque sélectionnée, un numéro à deux chiffres (de 00 à 99) a été produit au hasard. Ce numéro aléatoire a été ajouté à la banque pour former un numéro de téléphone complet. Grâce à cette méthode, il était possible d'incorporer dans l'échantillon des numéros de téléphone résidentiels publiés et non publiés, de même que des numéros de téléphone d'affaires ou non valides (c.-à-d. qui ne sont pas ou qui n'ont jamais été en service).

Chaque numéro de téléphone de l'échantillon a été composé pour vérifier s'il correspondait bien à un ménage. Si c'était le cas, on demandait à l'interlocuteur de fournir des informations sur les membres individuels du ménage. On utilisait l'âge des membres pour déterminer qui dans le ménage serait sélectionné pour l'interview sur l'usage du tabac. Les interviews par procuration n'étaient pas acceptées.

Pour s'assurer de joindre un nombre suffisant de personnes appartenant aux jeunes groupes d'âge, la sélection aléatoire a été structurée pour qu'au moins une personne âgée entre 15 et 19 ans ou entre 20 et 24 ans soit sélectionnée au sein du ménage, s'il en existait une. En effet, environ 76 % de tous les ménages au Canada comprennent seulement des personnes de plus de 25 ans, 20 % regroupent des personnes de plus de 25 ans qui vivent avec des personnes de 15 à 19 ans ou de 20 à 24 ans, et seulement 4 % des ménages ne contiennent personne âgée de plus de 25 ans. Si tous les âges étaient sélectionnés et retenus avec la même probabilité, le groupe des 25 ans et plus serait surreprésenté compte tenu des objectifs de l'enquête. De ce fait, pour économiser le coût d'interviews additionnels, une partie des personnes sélectionnées dans le groupe d'âge des 25 ans et plus ont été rejetées et n'ont pas été interviewées sur l'usage du tabac. Lorsque plus d'un des groupes d'âge de 15 à 19 ans, de 20 à 24 ans, ou de 25 ans et plus étaient représentés dans le ménage, deux personnes étaient sélectionnées. Les deux personnes sélectionnées dans un même ménage appartenaient toujours à des groupes d'âge différents. Cela garantissait qu'aucun impact négatif sur la précision des estimations par groupe d'âge ne serait causé par une corrélation à l'intérieur du ménage. Il y avait un léger impact sur la précision de l'estimation totale pour tous les âges, mais la taille de l'échantillon était suffisante pour que les conséquences restent minimales.

La sélection des personnes s'est faite selon la logique détaillée suivante :

- 1) Si toutes les personnes du ménage ont entre 15 et 19 ans, une personne est sélectionnée au hasard.

- 2) Si toutes les personnes du ménage ont entre 20 et 24 ans, une personne est sélectionnée au hasard.
- 3) Si toutes les personnes du ménage ont 25 ans ou plus, une personne est sélectionnée au hasard. Toutefois, cette personne n'est retenue que dans une proportion des cas.
- 4) Si des membres du ménage ont entre 15 et 19 ans et que les autres ont entre 20 et 24 ans, deux personnes sont sélectionnées au hasard, une de chaque groupe d'âge.
- 5) Si des membres du ménage ont entre 15 et 19 ans et que les autres ont 25 ans ou plus, deux personnes sont sélectionnées au hasard, une de chaque groupe d'âge. Toutefois, la personne sélectionnée du groupe 25 ans et plus n'est retenue que dans une proportion des cas.
- 6) Si des membres du ménage ont entre 20 et 24 ans et que les autres ont 25 ans ou plus, deux personnes sont sélectionnées au hasard, une de chaque groupe d'âge. Toutefois, la personne sélectionnée du groupe 25 ans et plus n'est retenue que dans une proportion des cas.
- 7) Si les trois groupes d'âge sont représentés dans le ménage, deux des groupes d'âge sont sélectionnés au hasard et la règle 4), 5), ou 6) s'applique.

6.0 Collecte des données

6.1 Conception du questionnaire

La conception du questionnaire de cette enquête se base fortement sur l'Enquête sur le tabagisme au Canada de 1994. Quelques questions ont été ajoutées par souci de cohérence avec les enquêtes internationales qui utilisent le concept de l'habitude de consommation de tabac « durant les 30 derniers jours ». Le questionnaire du 1^{er} cycle de 2002 contient plusieurs questions qui n'étaient pas incluses dans le cycle 2 de l'ESUTC de 2001.

Des spécifications définissant les limites valides et garantissant la cohérence d'une question à l'autre ont été intégrées dans l'application de l'interview téléphonique assistée par ordinateur (ITAO) dans la mesure du possible. Des contrôles de cohérences additionnels ont été réalisés durant la phase de traitement des données.

6.2 Collecte et vérification des données

Les interviews ont été menées chaque mois de février à juin 2002.

Les données ont été recueillies à l'aide des techniques de l'interview téléphonique assistée par ordinateur (ITAO). Le système ITAO contient plusieurs modules génériques qui s'adaptent rapidement à la plupart des types d'enquête. Un module frontal contient un ensemble de codes de réponse standards, qui s'appliquent à toutes les issues possible 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. Nous avons expliqué aux répondants comment ils avaient été sélectionnés pour l'enquête, que leur participation à l'enquête était volontaire et que les renseignements fournis resteraient strictement confidentiels. Les intervieweurs avaient accès à des écrans d'aide grâce auxquels ils pouvaient répondre aux questions les plus fréquemment posées par les répondants.

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 ».

Les intervieweurs ont reçu une formation sur l'utilisation de l'application ITAO et sur le contenu de l'enquête. En plus de la formation en classe, les intervieweurs complétaient une série d'interviews simulées afin de se familiariser avec l'enquête, ses concepts et ses définitions. Tout est mis en œuvre pour conserver le même groupe d'intervieweurs chaque mois. Cela réduit le temps de formation et produit des données de meilleure qualité et plus cohérentes.

Les cas ont été distribués dans quatre bureaux régionaux de Statistique Canada. La charge de travail et les intervieweurs de chaque bureau étaient supervisés par un chargé de projet. L'ordonnanceur automatique utilisé dans le système ITAO garantissait que les cas étaient assignés au hasard aux intervieweurs et que les appels se faisaient à différents moments de la journée pendant des jours différents de la semaine pour maximiser la probabilité de contact. Un maximum de 17 appels par cas ont été tentés. Une fois le maximum atteint, un intervieweur principal révisait le cas et déterminait si d'autres appels seraient tentés.

7.0 *Traitement des données*

Le principal produit de l'Enquête de surveillance de l'usage du tabac au Canada sont deux fichiers de microdonnées « épurés », le premier sur les ménages, le second sur les personnes. Ce chapitre présente un bref aperçu des étapes du traitement associées à la réalisation de ces fichiers.

7.1 *Saisie des données*

Les données ayant été recueillies à l'aide de l'ITAO, un système de collecte de données séparé n'était pas nécessaire puisque les informations étaient entrées directement dans les systèmes des bureaux régionaux par les intervieweurs durant l'interview.

7.2 *Vérification*

La première étape du traitement de l'enquête a été de combiner les fichiers mensuels en un seul fichier. Toutes les valeurs « hors-limites » des fichiers de données ont été remplacées par des blancs. Ce processus a été conçu pour faciliter les vérifications ultérieures.

Le premier type d'erreurs traitées étaient les erreurs dans l'enchaînement des questions, qui se traduisaient par des réponses à des questions qui ne s'adressaient pas au répondant (et donc auxquelles il n'aurait pas dû répondre). Dans ce cas, une vérification par ordinateur a automatiquement éliminé les données superflues d'après l'enchaînement dicté par les réponses aux questions précédentes, ou occasionnellement aux suivantes.

Les erreurs qui englobaient un manque d'information à l'intérieur de questions auxquelles on aurait dû répondre ont été le second type d'erreurs traitées. Dans ce cas, un cas de non-réponse ou « non déclaré » était attribué au poste.

7.3 *Création de variables calculées*

Un certain nombre de données élémentaires du fichier de microdonnées ont été calculées en regroupant des postes du questionnaire pour faciliter l'analyse des données. Le nombre moyen de cigarettes fumées quotidiennement, la suffisance du revenu, etc., sont des exemples de variables calculées.

7.4 *Pondération*

Le principe à la base de l'estimation dans le cas d'un échantillon probabiliste veut que chaque personne en représente plusieurs autres qui ne font pas partie de l'échantillon. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne représente 50 membres de la population.

La phase de la pondération est une étape où l'on calcule ce nombre (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 l'on doit estimer, par exemple, le nombre de personnes au Canada qui fument tous les jours, on le fait en sélectionnant les enregistrements renvoyant aux personnes incluses à l'intérieur de l'échantillon qui possèdent cette caractéristique (Q010 = '1') et en additionnant les poids inscrits de ces enregistrements. La pondération des ménages et des personnes se fait séparément tous les six mois.

Le chapitre 11 renferme des détails au sujet de la méthode utilisée pour calculer ces poids.

7.5 Suppression des renseignements confidentiels

Il convient de mentionner que le fichier de microdonnées « à grande diffusion » décrit ci-dessus diffère à certains égards importants du fichier principal de l'enquête conservé par Statistique Canada. Ces différences sont le résultat de mesures prises pour protéger l'anonymat des répondants à l'enquête. Les utilisateurs qui désirent accéder à des renseignements exclus du fichier de microdonnées peuvent acheter des totalisations personnalisées. Les estimations produites seront communiquées à l'utilisateur, à la condition qu'elles soient conformes aux lignes directrices pour l'analyse et la diffusion décrites au chapitre 9 du présent document.

Identificateurs géographiques

Fichier ménages et Fichier personnes :

Les fichiers principaux des données de l'enquête comprennent des identificateurs géographiques explicites pour la province et la strate (RMR, régions autres que des RMR, Toronto ou Montréal). Les fichiers de microdonnées à grande diffusion de l'enquête comprennent les mêmes identificateurs géographiques pour la province et la strate, sans aucune suppression.

Fichier personnes :

Pour la première fois, le fichier principal de données du 1^{er} cycle de 2002 comprenait les trois premiers chiffres du code postal du répondant.

Composition du ménage selon l'âge

Fichier ménages et Fichier personnes :

On peut obtenir la composition du ménage selon l'âge, c'est-à-dire le nombre de membres du ménage (limité à deux) dans les groupes d'âge suivants : 0 à 14 ans, 15 à 24 ans, 25 à 44 ans et 45 ans et plus.

8.0 Qualité des données

Pour l'Enquête de surveillance de l'usage du tabac au Canada, les taux de réponse calculés tiennent compte :

Le taux de numéros de téléphone résolus et le taux de succès s'appliquent à la fois au Fichier ménages et au Fichier personnes.

- Taux de numéros de téléphone résolus, un numéro de téléphone résolu étant un numéro résidentiel, d'affaires ou non admissible confirmés.
- Taux de succès, un succès étant un numéro résidentiel confirmé appartenant à un ménage contenant des **données valides sur le ménage** ou contenant des **données valides sur la personne**.

Ce taux de réponse des ménages ne s'applique qu'au Fichier ménages.

- Taux de réponse des ménages, une réponse étant un ménage pour lequel l'âge de toutes les personnes sur la liste a été fourni et des **données valides sur le ménage**.

Le taux de réponse des ménages, le taux d'achèvement des listes, le taux de réponse des personnes et le taux de réponse combiné des ménages ne s'appliquent que pour le Fichier personnes.

- Taux de réponse des ménages, une réponse étant un numéro de téléphone résidentiel confirmé ou des **données valides sur la personne** existaient avec une liste remplie.
- Taux d'achèvement des listes, une réponse étant un ménage pour lequel l'âge de toutes les personnes sur la liste a été fourni.
- Taux de réponse des personnes, une réponse étant un enregistrement contenant l'âge de toutes les personnes sur la liste et des **données valides sur la personne**.
- Taux de réponse combiné des ménages égal le taux de réponse des ménages multiplié par le taux de réponse des personnes.

Taux de numéros de téléphone résolus et taux de succès par province

Provinces	Total des n ^{os} de téléphone produits	Total des n ^{os} résolus	Taux de n ^{os} de téléphone résolus	Nombre total de ménages	Taux de succès
Terre-Neuve-et-Labrador	6 833	6 624	96,9 %	2 280	34,4 %
Île-du-Prince-Édouard	5 575	5 231	93,8 %	2 215	42,3 %
Nouvelle-Écosse	6 983	6 635	95,0 %	2 724	41,1 %
Nouveau-Brunswick	9 404	9 094	96,7 %	2 988	32,9 %
Québec	5 237	5 153	98,4 %	2 521	48,9 %
Ontario	12 024	11 539	96,0 %	4 732	41,0 %
Manitoba	6 157	6 157	100,0 %	2 527	41,0 %
Saskatchewan	6 320	6 320	100,0 %	2 463	39,0 %
Alberta	5 080	5 080	100,0 %	2 289	45,1 %
Colombie-Britannique	6 260	6 009	96,0 %	2 879	47,9 %
Canada	69 873	67 842	97,1 %	27 618	40,7 %

8.1 Taux de réponse des ménages - février à juin 2002

Le **répondant d'un ménage** doit remplir la liste sans aucun refus à propos de l'âge des personnes, et des données valides sur le ménage doivent exister. Il y a eu 2 528 (9,2 %) ménages non répondants.

Taux de réponse des ménages par province

Provinces	Nombre total de ménages	Ménages répondants	Taux de réponse des ménages
Terre-Neuve-et-Labrador	2 280	2 128	93,3 %
Île-du-Prince-Édouard	2 215	2 040	92,1 %
Nouvelle-Écosse	2 724	2 495	91,6 %
Nouveau-Brunswick	2 988	2 708	90,6 %
Québec	2 521	2 234	88,6 %
Ontario	4 732	4 370	92,3 %
Manitoba	2 527	2 242	88,7 %
Saskatchewan	2 463	2 176	88,3 %
Alberta	2 289	1 978	86,4 %
Colombie-Britannique	2 879	2 719	94,4 %
Canada	27 618	25 090	90,8 %

Taux de réponse des ménages par mois d'enquête

Mois de l'enquête	Nombre total de ménages	Ménages répondants	Taux de réponse des ménages
Février	5 495	4 937	89,8 %
Mars	5 617	5 023	89,4 %
Avril	5 492	5 037	91,7 %
Mai	5 560	5 066	91,1 %
Juin	5 454	5 027	92,2 %
Total	27 618	25 090	90,8 %

8.2 Taux de réponse des personnes - février à juin 2002

Une **personne répondante** possède les caractéristiques suivantes :

- Le numéro de téléphone de la personne sélectionnée appartient à un ménage.
- La liste a été remplie sans aucun refus à propos de l'âge des personnes.
- La personne sélectionnée était âgée de 15 ans ou plus au moment de l'enquête. L'âge inscrit dans la liste a été vérifié à l'aide de la date de naissance donnée par la personne sélectionnée.
- La personne sélectionnée a répondu au moins aux questions clés concernant les habitudes de fumer.

Dans 14 036 ménages, des données sur le ménage ont été recueillies mais personne n'a été sélectionné pour continuer l'Enquête sur l'usage du tabac. (Pour plus d'information, voir (Tirage de l'échantillon) à la section 5.4.) Dans les ménages restants, on a sélectionné une personne dans 9 512 ménages et deux personnes dans 1 611 ménages. Il y a eu seulement 57 (0,4 %) personnes qui ont refusé de remplir le questionnaire et 1 245 (9,8 %) personnes qui n'ont pas répondu pour d'autres raisons.

Taux de réponse des ménages, taux d'achèvement des listes et taux de réponse combiné des ménages par province

Provinces	Total des ménages	Total des ménages avec liste remplie	Taux de réponse des ménages	Ménages avec données de liste valides	Taux d'achèvement des listes	Taux de réponse combiné des ménages
Terre-Neuve-et-Labrador	2 280	2 183	95,7 %	2 132	97,7 %	93,5 %
Ile-du-Prince-Édouard	2 215	2 096	94,6 %	2 044	97,5 %	92,3 %
Nouvelle-Écosse	2 724	2 564	94,1 %	2 502	97,6 %	91,9 %
Nouveau-Brunswick	2 988	2 802	93,8 %	2 718	97,0 %	91,0 %
Québec	2 521	2 298	91,2 %	2 238	97,4 %	88,8 %
Ontario	4 732	4 491	94,9 %	4 376	97,4 %	92,5 %
Manitoba	2 527	2 307	91,3 %	2 251	97,6 %	89,1 %
Saskatchewan	2 463	2 235	90,7 %	2 188	97,9 %	88,8 %
Alberta	2 289	2 024	88,4 %	1 990	98,3 %	86,9 %
Colombie-Britannique	2 879	2 775	96,4 %	2 720	98,0 %	94,5 %
Canada	27 618	25 775	93,3 %	25 159	97,6 %	91,1 %

Taux de réponse des personnes par province

Provinces	Total des personnes sélectionnées	Total des personnes répondantes	Taux de réponse des personnes
Terre-Neuve-et-Labrador	1 158	1 023	88,3 %
Île-du-Prince-Édouard	1 124	1 007	89,6 %
Nouvelle-Écosse	1 166	1 028	88,2 %
Nouveau-Brunswick	1 435	1 243	86,6 %
Québec	1 120	999	89,2 %
Ontario	2 190	2 023	92,4 %
Manitoba	1 105	983	89,0 %
Saskatchewan	1 100	992	90,2 %
Alberta	1 082	951	87,9 %
Colombie-Britannique	1 254	1 183	94,3 %
Canada	12 734	11 432	89,8 %

Taux de réponse des personnes par mois d'enquête

Mois de l'enquête	Total des personnes sélectionnées	Total des personnes répondantes	Taux de réponse des personnes
Février	2 487	2 244	90,2 %
Mars	2 487	2 211	88,9 %
Avril	2 590	2 339	90,3 %
Mai	2 598	2 354	90,6 %
Juin	2 572	2 284	88,8 %
Total	12 734	11 432	89,8 %

Nombre cible de répondants et taux de réponse des personnes par groupe d'âge

Groupes d'âge	Nombre cible de répondants	Total des personnes sélectionnées	Total des personnes répondantes	Taux de réponse des personnes
15-19	2 500	3 387	3 054	90,2 %
20-24	2 500	2 905	2 540	87,4 %
25 +	5 000	6 442	5 838	90,6 %
Total	10 000	12 734	11 432	89,8 %

8.3 Erreurs dans l'enquête

Les estimations dérivées de cette enquête sont fondées sur un échantillon de ménages. On aurait pu obtenir des estimations légèrement différentes si l'on avait procédé à un recensement complet en faisant appel au même questionnaire et aux mêmes intervieweurs, superviseurs, méthodes de traitement, etc., que ceux effectivement utilisés. L'écart entre les estimations découlant de l'échantillon et celles tirées d'un dénombrement complet effectué dans des conditions semblables est appelée erreur d'échantillonnage de l'estimation.

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

Sur un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations tirées de l'enquête. Toutefois, les erreurs qui se produisent systématiquement contribueront à biaiser les estimations de l'enquête. On a consacré beaucoup de temps et d'efforts à réduire le nombre d'erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été prises à chaque étape du cycle de collecte et de traitement des données afin de contrôler la qualité des données. On a notamment donné aux intervieweurs une formation poussée pour ce qui est des procédures d'enquête et de l'application ITAO, observé les intervieweurs pour repérer les problèmes liés à la conception du questionnaire ou à la mauvaise interprétation des instructions, et testé l'application ITAO pour s'assurer que les contrôles des limites, les vérifications et le déroulement des questions étaient tous programmés correctement.

8.4 Non-réponse totale

Dans bien des enquêtes, la non-réponse totale peut être une source importante d'erreurs non dues à l'échantillonnage, selon la mesure dans laquelle les répondants et les non-répondants diffèrent quant aux caractéristiques présentées. Il y a non-réponse totale quand l'intervieweur n'a pu communiquer avec le répondant, ou quand le répondant a refusé de participer à l'enquête. On a traité la non-réponse totale en ajustant le poids des ménages ou des personnes qui ont répondu à l'enquête afin de compenser celles qui n'y ont pas répondu.

8.5 Non-réponse partielle

Il y a non-réponse partielle à l'enquête quand le répondant n'a pas compris ou a mal interprété une question, quand il a refusé de répondre à une question ou quand il ne pouvait se rappeler le renseignement requis. Des codes dans le fichier de microdonnées indiquent les non-réponses partielles, c'est-à-dire refus, ne sait pas.

8.6 Couverture

Tel qu'il est mentionné à la section 5.1 (Champ de l'enquête), moins de 3 % des ménages au Canada n'ont pas le téléphone. Les personnes qui vivent dans ces ménages ont peut-être des caractéristiques uniques qui ne seront pas reflétées dans les estimations de l'enquête. Les utilisateurs devraient faire preuve de prudence lorsqu'ils analysent des sous-groupes de la population dont les caractéristiques peuvent être corrélées au fait de ne pas avoir le téléphone.

8.7 **Mesure de l'erreur d'échantillonnage**

Comme les estimations d'une enquête par sondage comportent inévitablement une erreur d'échantillonnage, de bonnes méthodes statistiques exigent que les chercheurs fournissent aux utilisateurs une certaine indication de la grandeur de cette erreur d'échantillonnage. Cette section de la documentation décrit les mesures de l'erreur d'échantillonnage utilisées habituellement par Statistique Canada et dont le Bureau conseille vivement l'emploi aux utilisateurs qui font des estimations à partir de ce fichier de microdonnées.

La mesure de l'importance éventuelle des erreurs d'échantillonnage est fondée sur l'erreur-type des estimations découlant des résultats de l'enquête.

Cependant, en raison de la diversité des estimations que l'on peut tirer 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 qu'en se basant sur les résultats de l'enquête, on estime que 21,4 % des Canadiens fument actuellement la cigarette, et qu'on constate que cette estimation a une erreur-type de 0,0068. Le CV de l'estimation serait calculé ainsi :

$$\left(\frac{0,0068}{0,214} \right) \times 100\% = 3,2\%$$

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

Ce chapitre décrit les lignes directrices que doivent suivre les utilisateurs qui totalisent, analysent, publient ou autrement diffusent des données tirées des fichiers de microdonnées de l'enquête. Ces lignes directrices devraient permettre aux utilisateurs d'arriver aux mêmes chiffres que Statistique Canada, tout en étant en mesure d'obtenir, conformément à ces lignes directrices, des chiffres qui n'ont pas été publiés.

9.1 Lignes directrices sur l'arrondissement d'estimations

Afin que les estimations destinées à la publication ou à toute autre forme de diffusion qui sont tirées de ces fichiers de microdonnées correspondent à celles de Statistique Canada, nous conseillons vivement aux utilisateurs de se conformer aux lignes directrices ci-après en ce qui concerne l'arrondissement de ces estimations :

- a) Les estimations dans le corps d'un tableau statistique doivent être arrondies à la centaine près au moyen de la technique d'arrondissement classique. Dans 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, on augmente de 1 la valeur du dernier chiffre à conserver. Par exemple, lorsqu'on cherche à arrondir à la centaine près de la façon classique, si les deux derniers chiffres se situent entre 00 et 49, il faut les remplacer par 00 et laisser le chiffre précédent (le chiffre des centaines) tel quel. Si les deux derniers chiffres se situent entre 50 et 99, on les remplace par 00 et on augmente de 1 le chiffre précédent.
- b) Les totaux partiels marginaux et les totaux marginaux des tableaux statistiques doivent être calculés à partir de leurs composantes correspondantes non arrondies, puis arrondis à leur tour à la centaine près à l'aide de la technique d'arrondissement classique.
- c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir d'éléments non arrondis (c'est-à-dire les numérateurs et (ou) les dénominateurs), puis arrondis à une décimale au moyen de la technique d'arrondissement classique. Lorsqu'on cherche à arrondir à un seul chiffre par cette technique, 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, on augmente de 1 le dernier chiffre à conserver.
- d) Les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la centaine près (ou à la décimale près) en employant la technique d'arrondissement classique.
- e) Si, en raison de limitations d'ordre technique ou de toute autre nature, on utilise une autre technique d'arrondissement que l'arrondissement classique, de sorte que les estimations à publier ou à diffuser sous une forme quelconque diffèrent des estimations correspondantes publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces divergences dans le ou les documents à publier ou à diffuser.
- f) Des estimations non arrondies ne doivent en aucun cas être publiées ou diffusées par les utilisateurs. Des estimations qui ne sont pas arrondies laissent supposer qu'elles sont beaucoup plus précises qu'elles ne le sont en réalité.

9.2 Lignes directrices sur la pondération d'un échantillon pour une totalisation

Le plan d'échantillonnage utilisé pour l'Enquête de surveillance de l'usage du tabac au Canada (ESUTC) n'était pas autopondéré. Pour produire des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent employer le poids approprié.

Sinon, les estimations calculées à partir des fichiers de microdonnées ne peuvent être considérées comme représentatives de la population observée et ne correspondront pas à celles de Statistique Canada.

Les utilisateurs doivent également noter qu'en raison de la manière dont ils traitent le champ du poids, certains progiciels ne permettent peut-être pas d'obtenir des estimations correspondant exactement à celles de Statistique Canada.

9.2.1 Définitions de types d'estimations : catégoriques et quantitatives

Avant de discuter de la façon dont on peut totaliser et analyser les données de l'ESUTC, il est utile de décrire les deux principaux genres d'estimations ponctuelles des caractéristiques de la population qui peuvent être produites à partir des fichiers de microdonnées pour l'Enquête de surveillance de l'usage du tabac au Canada.

Estimations catégoriques

Les estimations catégoriques sont des estimations du nombre ou du pourcentage de personnes dans la population visée par l'enquête qui possèdent certaines caractéristiques ou qui font partie d'une catégorie définie. Le nombre de personnes qui fument actuellement des cigarettes ou la proportion de fumeurs quotidiens qui ont tenté de cesser de fumer sont des exemples de telles estimations. Une estimation du nombre de personnes possédant une certaine caractéristique peut aussi être appelée estimation d'un agrégat.

Exemples de questions catégoriques :

Q : Avez-vous fumé des cigarettes au cours des 30 DERNIERS JOURS?

R : Oui/Non

Q : Qu'est-ce qui vous a incité à cesser de fumer?

R : Problèmes de santé / Maladie liée à la cigarette ou décès d'un ami / Grossesse / Recommandation d'un médecin / Inquiétude au sujet de sa propre santé future

Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes ou d'autres mesures de tendance centrale de quantités basées sur certains ou sur tous les membres de la population visée par l'enquête. Elles comprennent aussi explicitement 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 cigarettes fumées, les samedis, par personne. Si l'on écrit cette moyenne sous la forme $\hat{X} / \hat{Y}$ le numérateur ($\hat{X}$) est une estimation du nombre total de cigarettes fumées les samedis alors que le dénominateur ($\hat{Y}$) est une estimation du nombre de personnes ayant déclaré avoir fumé les samedis.

Exemples de questions quantitatives :

Q : Au cours des 7 derniers jours, à compter d'hier, combien de cigarettes avez-vous fumées samedi?

R : cigarettes

Q : À quel âge avez-vous fumé votre première cigarette?

R : ans

9.2.2 **Totalisation d'estimations catégoriques**

On peut obtenir des estimations du nombre de personnes possédant une caractéristique particulière 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 les proportions et les rapports de la forme $\hat{X} / \hat{Y}$ en :

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

9.2.3 **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 cigarettes fumées le samedi, il faut multiplier la valeur déclarée à Q090SAT (nombre de cigarettes fumées le samedi) par le poids final de l'enregistrement, puis additionner cette valeur pour tous les enregistrements portant la mention Q090SAT < 96 (tous les répondants qui ont donné une réponse à ce champ).

Afin d'obtenir une moyenne pondérée exprimée sous la forme $\hat{X} / \hat{Y}$, le numérateur ($\hat{X}$) est calculé comme une estimation quantitative et le dénominateur ($\hat{Y}$), comme une estimation catégorique. Par exemple, pour estimer le nombre moyen de cigarettes fumées le samedi,

- a) estimez le nombre total de cigarettes fumées le samedi, comme on l'a vu ci-dessus;
- b) estimez le nombre de personnes dans cette catégorie par l'addition des poids finals de tous les enregistrements ayant la mention Q090SAT < 96; puis
- c) divisez l'estimation obtenue en a) par celle calculée en b).

9.3 Lignes directrices pour l'analyse statistique de données

L'Enquête de surveillance de l'usage du tabac au Canada est fondée sur un plan complexe, comportant une stratification et 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 difficultés aux analystes parce que le plan d'enquête et les probabilités de sélection influent sur les méthodes d'estimation et de calcul de la variance qui doivent être utilisées. Pour que les estimations et les analyses de données d'enquête soient sans biais, les poids de l'enquête doivent être utilisés.

Bien que de nombreuses méthodes d'analyse qui font partie des progiciels statistiques permettent d'utiliser des poids, le poids diffère sous l'angle de la définition ou du sens de celui qui convient à une enquête par sondage, de sorte que si les estimations faites au moyen de ces progiciels sont exactes dans bien des cas, les variances calculées n'ont pratiquement aucune signification. Les variances pour des estimations simples comme des totaux, des proportions et des rapports (pour des variables qualitatives) sont fournis dans les tables de variabilité d'échantillonnage approximative qui les accompagnent.

Pour d'autres techniques d'analyse (par exemple, la régression linéaire, la régression logistique, l'analyse de variance), il existe un moyen de rendre les variances calculées par l'application des progiciels normalisés plus significatives en incorporant les probabilités inégales de sélection. La méthode transforme les poids des enregistrements de manière à ce que le poids moyen soit un (1).

Par exemple, si nous voulions faire une analyse pour les répondants masculins, les étapes pour transformer les poids seraient les suivants :

- sélectionner tous les répondants du fichier qui ont répondu SEX=homme
- calculer le poids MOYEN pour ces enregistrements en totalisant les poids originaux du fichier de microdonnées pour ces enregistrements et le divisant par le nombre de répondants ayant répondu SEX=homme
- pour chacun de ces répondants, calculer un poids RECALCULÉ égal au poids original de ce répondant divisé par le poids MOYEN
- effectuer l'analyse de ces répondants en utilisant le poids RECALCULÉ.

Cependant, la stratification et le groupement du plan de sondage ne sont pas considérés dans cette méthode et l'estimation de la variance calculée de cette façon est probablement sous-évaluée.

Le calcul d'estimations de variance vraiment significatives requiert une connaissance approfondie du plan de l'enquête. Ces détails ne peuvent être fournis dans ce fichier de microdonnées à cause de la confidentialité. Les variances qui prennent en compte le plan complet de l'enquête peuvent être calculées pour plusieurs données par Statistique Canada sur une base de recouvrement des coûts.

9.4 Lignes directrices sur la diffusion de coefficients de variation

Avant de diffuser et (ou) de publier des estimations tirées de ce fichier de microdonnées, les utilisateurs doivent commencer par établir le niveau de qualité de l'estimation. Les niveaux de qualité sont *acceptable*, *médiocre* et *inacceptable*. La qualité des données est affectée par des erreurs d'échantillonnage et des erreurs non dues à l'échantillonnage tel qu'il a été discuté au chapitre 8. Cependant, aux fins de ce document, le niveau de qualité d'une estimation est déterminé seulement en fonction de l'erreur d'échantillonnage reflétée par son CV tel qu'il est indiqué au tableau suivant. Les utilisateurs devraient néanmoins s'assurer de lire le chapitre 8 pour mieux comprendre les caractéristiques relatives à la qualité de ces données.

Il faut commencer par établir le nombre de répondants qui ont été retenus pour le calcul de l'estimation. Si ce nombre est inférieur à 30, l'estimation pondérée doit être considérée de qualité inacceptable.

Pour les estimations pondérées basées sur des échantillons composés de 30 enquêtés ou plus, les utilisateurs doivent calculer le CV de l'estimation et suivre les lignes directrices relatives au niveau de qualité ci-dessous. Elles sont appliquées aux estimations pondérées arrondies.

Toutes les estimations peuvent faire l'objet de diffusion. Toutefois, celles de qualité médiocre ou inacceptable devraient être accompagnées d'un avertissement mettant en garde les prochains utilisateurs.

Lignes directrices relatives au niveau de qualité de l'estimation

Niveau de qualité de l'estimation	Lignes directrices
1) Acceptable	Caractéristiques des estimations : un échantillon de 30 ou plus, et faibles coefficients de variation compris entre 0,0 % - 16,5 % Aucune mise en garde nécessaire.
2) Médiocre	Caractéristiques des estimations : un échantillon de 30 ou plus, et des coefficients de variation élevés compris entre 16,6 % - 33,3 %. On doit désigner ces estimations en utilisant la lettre M (ou un autre identificateur semblable). Elles doivent être accompagnées d'un avertissement mettant les prochains utilisateurs en garde contre les hauts taux d'erreur associés aux estimations.
3) Inacceptable	Caractéristiques des estimations : un échantillon inférieur à 30, ou des coefficients de variation très élevés, supérieurs à 33,3 %. Statistique Canada recommande de ne pas diffuser d'estimations dont la qualité est inacceptable. Toutefois, si l'utilisateur choisit de le faire, il doit alors désigner les estimations en utilisant la lettre I (ou un autre identificateur semblable) et les diffuser avec l'avertissement suivant : « Nous informons l'utilisateur que ces estimations (désignées avec la lettre I)... ne respectent pas les normes de qualité de Statistique Canada pour ce programme statistique. Les conclusions qui reposeront sur ces données ne seront ni fiables ni valables très probablement. »

10.0 Tables de variabilité d'échantillonnage approximative

Afin de permettre aux utilisateurs d'avoir facilement accès à des coefficients de variation (CV) qui s'appliqueraient à une gamme étendue d'estimations catégoriques réalisées à partir de ce fichier de microdonnées, on a préparé un ensemble de « tables à consulter » appelées Tables de variabilité d'échantillonnage approximative. Ces tables permettent aux utilisateurs d'obtenir un coefficient de variation approximatif basé sur la taille de l'estimation calculée à partir des données de l'enquête.

Les coefficients de variation dans ces tables sont obtenus au moyen de la formule de la variance utilisée pour l'échantillonnage aléatoire simple en y incorporant un facteur reflétant la structure en grappes à plusieurs degrés du plan d'échantillonnage. Pour obtenir ce facteur, appelé effet du plan, on a commencé par calculer les effets du plan pour une gamme étendue de caractéristiques, puis on a choisi parmi les nombres ainsi obtenus une valeur modérée à employer dans les tables à consulter qui serait ensuite appliquée à l'ensemble des caractéristiques.

Le tableau suivant indique la valeur modérée des effets du plan, des tailles de l'échantillon et des chiffres de population par province, qui ont été utilisés pour produire les tables de variabilité d'échantillonnage approximative pour le Fichier ménages.

Fichier ménages

Provinces	Effet du plan	Taille de l'échantillon	Population
Terre-Neuve-et-Labrador	1,22	2 128	196 244
Île-du-Prince-Édouard	1,14	2 040	53 112
Nouvelle-Écosse	1,29	2 495	367 496
Nouveau-Brunswick	1,14	2 708	290 652
Québec	1,42	2 234	3 082 954
Ontario	1,07	4 370	4 482 452
Manitoba	1,12	2 242	430 462
Saskatchewan	1,07	2 176	384 358
Alberta	1,21	1 978	1 144 709
Colombie-Britannique	1,07	2 719	1 604 241
Canada	2,54	25 090	12 036 680

Le tableau suivant indique la valeur modérée des effets du plan, des tailles de l'échantillon et des chiffres de population par province et groupe d'âge, qui ont été utilisés pour produire les tables de variabilité d'échantillonnage approximative pour le Fichier personnes.

Fichier personnes

Province s	Groupe s d'âge	Effet du plan	Taille de l'échantillon	Population
Terre-Neuve-et-Labrador	Tous	2,13	1 023	440 346
	15-19	1,42	292	38 919
	20-24	1,98	210	37 625
	25+	1,47	521	363 802
Île-du-Prince-Édouard	Tous	1,90	1 007	111 855
	15-19	1,51	267	10 469
	20-24	1,61	221	9 597
	25+	1,32	519	91 789
Nouvelle-Écosse	Tous	1,88	1 028	767 689
	15-19	1,53	285	64 256
	20-24	1,55	193	61 198
	25+	1,35	550	642 235
Nouveau-Brunswick	Tous	2,06	1 243	615 485
	15-19	1,37	362	50 458
	20-24	1,47	293	49 958
	25+	1,31	588	515 070
Québec	Tous	2,39	999	6 063 602
	15-19	2,38	247	458 749
	20-24	2,62	228	510 309
	25+	1,69	524	5 094 544
Ontario	Tous	1,95	2 023	9 648 662
	15-19	1,41	520	795 300
	20-24	1,57	453	782 310
	25+	1,34	1 050	8 071 051
Manitoba	Tous	1,75	983	902 487
	15-19	1,22	239	82 020
	20-24	1,45	227	78 227
	25+	1,31	517	742 241
Saskatchewan	Tous	1,69	992	788 018
	15-19	1,53	266	77 505
	20-24	1,89	200	72 476
	25+	1,28	526	638 038
Alberta	Tous	1,88	951	2 448 041
	15-19	1,48	251	228 740
	20-24	1,48	211	234 176
	25+	1,39	489	1 985 125
Colombie-Britannique	Tous	2,19	1 183	3 377 466
	15-19	1,45	340	274 196
	20-24	1,67	272	277 804
	25+	1,49	571	2 825 467
Canada	Tous	3,91	11 432	25 163 652
	15-19	3,30	3 069	2 080 612
	20-24	3,40	2 508	2 113 679
	25+	2,74	5 855	20 969 361

Il convient de mentionner que tous les coefficients de variation sont approximatifs dans les tables de variabilité d'échantillonnage approximative et ne doivent donc pas être considérés comme des valeurs officielles. On peut se procurer, sur une base de recouvrement des coûts, des estimations de la variance réelle de variables particulières auprès de Statistique Canada. Étant donné que le CV approximatif est une estimation prudente, l'utilisation de la variance réelle estimée permettrait aux utilisateurs de passer 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.

N'oubliez pas : Si le nombre d'observations sur lesquelles une estimation est basée est inférieur à 30, l'estimation pondérée est considérée de qualité *inacceptable* et devrait être accompagnée d'une mise en garde quelle que soit la valeur de son coefficient de variation. Ceci est dû au fait que les formules utilisées pour obtenir une estimation de la variance ne donnent pas de bons résultats pour de petits échantillons.

10.1 Comment utiliser les tables de CV pour les estimations catégoriques

Les règles suivantes doivent permettre à l'utilisateur de calculer à partir des tables de variabilité d'échantillonnage approximative les coefficients de variation approximatifs d'estimations relatives au nombre, à la proportion ou au pourcentage de personnes dans la population enquêtée possédant une caractéristique donnée ainsi que des rapports et des écarts entre ces estimations.

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

Le CV dépend uniquement de la taille de l'estimation elle-même. Sur la table de variabilité d'échantillonnage approximative correspondant à la région géographique appropriée, repérez le nombre estimé dans la colonne à l'extrême gauche de la table (colonne intitulée « Numérateur du pourcentage ») et suivez les astérisques (le cas échéant) de gauche à droite jusqu'au premier nombre. Ce nombre constitue le CV approximatif.

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

Le CV d'une proportion ou d'un pourcentage estimé dépend à la fois de la grandeur de cette proportion ou de ce pourcentage et de la grandeur du total sur lequel est fondé cette proportion ou ce pourcentage. Les proportions ou les pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque cette proportion ou ce pourcentage est fondé sur un sous-ensemble de la population. Par exemple, la proportion « d'anciens fumeurs qui ont cessé pour des problèmes de santé » est une mesure plus fiable que le nombre estimatif total d'anciens fumeurs qui ont cessé pour des problèmes de santé. (Il faut noter que dans les tables, la valeur des coefficients de variation décroît de gauche à droite sur une même ligne.)

Lorsque la proportion ou le pourcentage est fondé sur la population totale de la région géographique à laquelle la table s'applique, le CV de la proportion ou du pourcentage est égal à celui du numérateur de la proportion ou du pourcentage. Dans ce cas, on peut appliquer la règle 1.

Lorsque la proportion ou le pourcentage est fondé sur un sous-ensemble de la population totale (par exemple, les personnes qui appartiennent à un groupe d'âge ou un sexe particulier), il faut se reporter à la proportion ou au pourcentage (en haut de la table) et au numérateur de la proportion ou du pourcentage (dans la colonne de gauche de la table). Le CV se trouve à l'intersection de la ligne et de la colonne appropriées.

Règle 3 : Estimations des différences entre des agrégats ou des pourcentages

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

$$s_{\hat{d}} = \sqrt{(\hat{X}_1 a_1)^2 + (\hat{X}_2 a_2)^2}$$

où $\hat{X}_1$ représente l'estimation 1, $\hat{X}_2$ l'estimation 2 et a_1 et a_2 sont les coefficients de variation de $\hat{X}_1$ et $\hat{X}_2$ respectivement. Le CV de $\hat{d}$ est donné par $s_{\hat{d}} / \hat{d}$. Cette formule donne un résultat exact en ce qui a trait à la différence entre des caractéristiques distinctes et non corrélées, mais elle ne donne que des résultats approximatifs dans les autres cas.

Règle 4 : Estimations de rapports

Si le numérateur est un sous-ensemble du dénominateur, il faut convertir le rapport en pourcentage et appliquer la règle 2. Cette situation s'appliquerait notamment au cas où le dénominateur est le nombre « de fumeurs » et le numérateur, le nombre de « fumeurs quotidiens ».

Si le numérateur n'est pas un sous-ensemble du dénominateur, (par exemple, le rapport du nombre de « fumeurs quotidiens » au nombre de « non-fumeurs ») l'écart-type du rapport entre les estimations est approximativement égal à la racine carrée de la somme des carrés de chaque coefficient de variation pris séparément multipliée par $\hat{R}$. L'erreur-type d'un rapport ($\hat{R} = \hat{X}_1 / \hat{X}_2$) est donc :

$$s_{\hat{R}} = \hat{R} \sqrt{a_1^2 + a_2^2}$$

où a_1 et a_2 sont les coefficients de variation de $\hat{X}_1$ et $\hat{X}_2$ respectivement. Le CV de $\hat{R}$ est donné par $s_{\hat{R}} / \hat{R}$. La formule tend à surestimer l'erreur si $\hat{X}_1$ et $\hat{X}_2$ sont corrélés positivement et à sous-estimer l'erreur si $\hat{X}_1$ et $\hat{X}_2$ sont corrélés négativement.

Règle 5 : Estimations des différences entre des rapports

Dans ce cas, on combine les règles 3 et 4. On commence par calculer les coefficients de variation des deux rapports au moyen de la règle 4, puis le CV de leur différence au moyen de la règle 3.

10.1.1 Exemples d'utilisation des tables de CV pour des estimations catégoriques

Les exemples ci-dessous utilisent des données du fichier du cycle 1 de 2002 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 durant la période de référence, à 5 385 164 le nombre de personnes qui sont des fumeurs actuels (DVSST1 = '1') au Canada. Comment l'utilisateur détermine-t-il le CV de cette estimation?

- 1) Reportez-vous à la table de CV pour le CANADA du Fichier personnes.
- 2) L'agrégat estimé (5 385 164) ne se trouve pas dans la colonne de gauche (la colonne « Numérateur du pourcentage ») il faut donc utiliser le nombre qui s'en rapproche le plus, c'est-à-dire 5 000 000.
- 3) Le CV pour un agrégat estimé est la première inscription autre que des astérisques sur cette ligne, c'est-à-dire 3,7 %.
- 4) Le CV approximatif de l'estimation est donc 3,7 %. Il n'y a donc pas lieu de faire une mise en garde lorsque l'on publie le résultat, soit qu'il y a eu 5 385 164 fumeurs durant la période de référence.

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

Supposons qu'un utilisateur estime à $2\,890\,124 / 12\,393\,013 = 23,3\%$ le pourcentage d'hommes qui fumaient au Canada durant la période de référence. Comment l'utilisateur détermine-t-il le CV de cette estimation?

- 1) Reportez-vous à la table de CV pour le CANADA du Fichier personnes. Le tableau CANADA est utilisé parce qu'il est le plus petit tableau qui contient le domaine de l'estimation, tous les hommes au Canada.
- 2) Parce que l'estimation est un pourcentage basé sur un sous-ensemble de la population totale (c.-à-d. les hommes), il faut utiliser à la fois le pourcentage (23,3 %) et la partie numérateur du pourcentage (2 890 124) pour déterminer le CV.
- 3) Le numérateur, 2 890 124 ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »), il faut donc utiliser le nombre qui s'en rapproche le plus, soit 3 000 000. De même, l'estimation du pourcentage ne figure pas parmi les en-têtes de colonnes, il faut donc utiliser le nombre qui s'en rapproche le plus, soit 25,0 %.
- 4) Le nombre qui se trouve à l'intersection de la ligne et de la colonne utilisées, soit 4,6 %, est le CV à employer.
- 5) Le CV approximatif de l'estimation est donc 4,6 %. Il n'y a donc pas lieu de faire une mise en garde lorsque l'on publie le résultat, soit que 23,3 % des hommes fument actuellement.

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

Supposons qu'un utilisateur estime à $2\,495\,041 / 12\,770\,639 = 19,5 \%$ le pourcentage de femmes qui fument actuellement et à $2\,890\,124 / 12\,393\,013 = 23,3 \%$ le pourcentage d'hommes qui fument actuellement au Canada. Comment l'utilisateur détermine-t-il le CV de la différence entre ces deux estimations?

- 1) À l'aide de la table de CV pour le CANADA du Fichier personnes utilisée de la même façon que dans l'exemple 2, vous établissez à 5,9 % le CV de l'estimation pour les femmes et à 4,6 % le CV de l'estimation pour les hommes.
- 2) Selon la règle 3, l'erreur-type pour une différence ($\hat{d} = \hat{X}_1 - \hat{X}_2$) est :

$$s_{\hat{d}} = \sqrt{(\hat{X}_1 a_1)^2 + (\hat{X}_2 a_2)^2}$$

où $\hat{X}_1$ est l'estimation 1 (hommes), $\hat{X}_2$ est l'estimation 2 (femmes) et a_1 et a_2 sont les coefficients de variation de $\hat{X}_1$ et $\hat{X}_2$ respectivement.

C'est-à-dire que l'erreur-type de la différence

$\hat{d} = (0,233 - 0,195) = 0,038$ est :

$$\begin{aligned} s_{\hat{d}} &= \sqrt{[(0,233)(0,046)]^2 + [(0,195)(0,059)]^2} \\ &= \sqrt{(0,00011) + (0,00013)} \\ &= 0,015 \end{aligned}$$

- 3) Le CV de $\hat{d}$ est donné par $s_{\hat{d}} / \hat{d} = 0,015 / 0,038 = 0,395$.
- 4) Le CV approximatif de la différence entre les estimations est donc 39,5 %. Cette estimation est considérée de qualité 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 I (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les prochains utilisateurs en garde contre les hauts taux d'erreur associés aux estimations.

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime à 242 786 le nombre de femmes qui fument actuellement âgées de 15 à 19 ans et à 212 902 le nombre d'hommes qui fument actuellement âgés de 15 à 19 ans. L'utilisateur est intéressé à comparer l'estimation des femmes à celle des hommes sous la forme d'un rapport. Comment l'utilisateur détermine-t-il le CV 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 qui fument actuellement âgées de 15 à 19 ans. Le dénominateur de l'estimation ($= \hat{X}_2$) est le nombre d'hommes qui fument actuellement âgés de 15 à 19 ans.
- 2) Reportez-vous à la table de CV pour le CANADA - 15-19 du Fichier personnes.
- 3) Le numérateur de cette estimation de rapport est 242 786. Le chiffre qui se rapproche le plus de ce nombre est 250 000. On détermine le CV pour cette estimation en trouvant la première inscription, autre que des astérisques sur cette ligne, soit 8,7 %.
- 4) Le dénominateur de cette estimation de rapport est 212 902. Le chiffre qui se rapproche le plus de ce nombre est 200 000. On détermine le CV de cette estimation en trouvant la première inscription, autre que des astérisques sur cette ligne, soit 10,0 %.
- 5) Le CV approximatif de l'estimation du rapport est donc donné par la règle 4, qui est :

$$a_{\hat{R}} = \sqrt{a_1^2 + a_2^2}$$

où a_1 et a_2 sont les coefficients de variation de $\hat{X}_1$ et $\hat{X}_2$ respectivement.

C'est-à-dire :

$$\begin{aligned} a_{\hat{R}} &= \sqrt{(0,087)^2 + (0,100)^2} \\ &= 0,133 \end{aligned}$$

Le rapport femmes:hommes fumeurs actuels âgés de 15 à 19 ans est égal à 242 786 / 212 902, soit 1,14. Le coefficient de variation de cette estimation est égal à 13,3 %, donc l'estimation peut être publiée sans mise en garde.

10.2 Comment utiliser les tables de CV pour obtenir des limites de confiance

Bien que les coefficients de variation soient largement utilisés, l'intervalle de confiance d'une estimation donne une mesure intuitive plus significative de l'erreur d'échantillonnage. Un intervalle de confiance est une façon d'énoncer la probabilité que la valeur vraie de la population se situe dans une plage de valeurs données. Par exemple, un intervalle de confiance à 95 % peut être décrit comme suit :

Si l'échantillonnage de la population se répète à l'infini, chacun des échantillons donnant un nouvel intervalle de confiance pour une estimation alors, dans 95 % des cas, l'intervalle contiendra la valeur vraie de la population.

Une fois déterminée l'erreur-type d'une estimation, on peut calculer des intervalles de confiance pour les estimations en partant de l'hypothèse qu'en procédant à un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique de la population sont réparties selon une distribution normale 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. On appelle ces différents degrés de confiance des niveaux de confiance.

Les intervalles de confiance pour une estimation $\hat{X}$ sont généralement exprimés sous forme de deux nombres, l'un étant inférieur à l'estimation et l'autre supérieur à celle-ci, sous la forme $(\hat{X} - k, \hat{X} + k)$ où k varie selon le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

On peut calculer directement les intervalles de confiance pour une estimation à partir des tables de variabilité d'échantillonnage approximative, en commençant par trouver, dans la table appropriée, le coefficient de variation de l'estimation $\hat{X}$, puis en utilisant la formule ci-dessous pour obtenir l'intervalle de confiance (CI pour « confidence interval » en anglais) correspondant :

$$CI_x = (\hat{X} - t\hat{X}a_{\hat{X}}, \hat{X} + t\hat{X}a_{\hat{X}})$$

où $a_{\hat{X}}$ est le coefficient de variation trouvé pour $\hat{X}$ et

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

Nota : Les lignes directrices qui s'appliquent à la diffusion des estimations s'appliquent aussi aux intervalles de confiance. Par exemple, si l'estimation est de qualité « inacceptable » alors l'intervalle de confiance est aussi de qualité « inacceptable ».

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

Voici comment on calculerait un intervalle de confiance à 95 % pour la proportion estimée de fumeurs actuels qui sont des hommes (d'après l'exemple 2 à la section 10.1.1) :

$$\hat{X} = 23,3 \% \text{ (ou exprimé sous forme de proportion } 0,233)$$

$$t = 2$$

$a_{\hat{X}} = 4,6 \%$ (0,046 sous forme de proportion) est le CV de cette estimation tel que déterminé à partir des tables.

$$CI_x = \{0,233 - (2) (0,233) (0,046), 0,233 + (2) (0,233) (0,046)\}$$

$$CI_x = \{0,233 - 0,021, 0,233 + 0,021\}$$

$$CI_x = \{0,212, 0,254\}$$

On peut dire, avec une probabilité de 95 %, qu'entre 21,2 % et 25,4 % des hommes fument actuellement.

10.3 **Comment utiliser les tables de CV pour effectuer un test t**

On peut aussi utiliser les erreurs-types pour effectuer des tests d'hypothèses, une technique qui permet de faire la distinction entre les paramètres d'une population à l'aide d'estimations basées sur un échantillon. Ces estimations peuvent être des nombres, 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 basées sur un échantillon pour deux caractéristiques qui nous intéressent. Supposons également que l'erreur-type de la différence $\hat{X}_1 - \hat{X}_2$ est $s_{\hat{d}}$.

$$\text{Si } t = \frac{\hat{X}_1 - \hat{X}_2}{s_{\hat{d}}}$$

est compris entre -2 et 2, alors on ne peut tirer de conclusion à propos de la différence entre les caractéristiques au niveau de signification de 5 %. Toutefois, si 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 CV pour effectuer un test t**

Supposons que nous désirons tester, au niveau de signification de 5 %, l'hypothèse qu'il y a une différence entre la proportion d'hommes qui fument actuellement et la proportion de femmes qui fument actuellement. Dans l'exemple 3 à la section 10.1.1, nous avons trouvé que l'erreur-type de la différence entre ces deux estimations était égal à 0,015. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{s_{\hat{d}}} = \frac{0,233 - 0,195}{0,015} = \frac{0,038}{0,015} = 2,53.$$

Puisque $t = 2,53$ est plus grand que 2, on doit conclure qu'il existe une différence significative au niveau de 0,05 entre les deux estimations.

10.4 **Coefficients de variation d'estimations quantitatives**

Pour calculer l'erreur d'échantillonnage d'estimations quantitatives, il faudrait élaborer des tables spéciales, ce qui n'a pas été fait puisque la majorité des variables de l'Enquête de surveillance sur l'usage du tabac au Canada sont surtout de types catégoriques.

Néanmoins, le CV d'un total quantitatif est généralement supérieur à celui de l'estimation catégorique correspondante (c'est-à-dire l'estimation du nombre de personnes contribuant à l'estimation quantitative). Si l'estimation catégorique correspondante ne peut être diffusée, il

en sera de même pour l'estimation quantitative. Par exemple, le CV du nombre total de cigarettes fumées le samedi serait plus élevé que le CV de la proportion de fumeurs actuels. Par conséquent, si le CV de la proportion ne peut être diffusé, ce sera aussi le cas pour le CV de l'estimation quantitative correspondante.

On peut calculer, au besoin, les coefficients de variation d'estimations de ce genre pour une estimation particulière au moyen d'une technique appelée pseudo-réplication, qui consiste à diviser les enregistrements des fichiers de microdonnées en sous-groupes (ou sous-échantillons) et à calculer la variabilité de l'estimation d'un sous-échantillon à l'autre. Les utilisateurs désireux de calculer le CV d'estimations quantitatives peuvent demander conseil à Statistique Canada en ce qui concerne l'allocation d'enregistrements en sous-échantillons appropriés et les formules à utiliser pour ces calculs.

10.5 Seuils pour la diffusion des estimations pour l'Enquête de surveillance de l'usage du tabac au Canada - Fichier ménages

La taille minimum des estimations pour les ménages au niveau provincial est indiquée dans le tableau ci-dessous. Les estimations plus petites que la taille indiquée dans la colonne « Inacceptable » doivent être publiées avec une mise en garde tel qu'il est décrit dans le tableau des qualités des estimations.

Tableau des seuils pour la diffusion des estimations – Fichier ménages

Provinces	Acceptable CV 0,0 % - 16,5 %	Médiocre CV 16,6 % - 33,3 %	Inacceptable CV > 33,3 %
Terre-Neuve-et-Labrador	4 000 & plus	1 000 à < 4 000	moins de 1 000
Île-du-Prince-Édouard	1 000 & plus	500 à < 1 000	moins de 500
Nouvelle-Écosse	7 000 & plus	1 500 à < 7 000	moins de 1 500
Nouveau-Brunswick	4 500 & plus	1 000 à < 4 500	moins de 1 000
Québec	70 500 & plus	17 500 à < 70 500	moins de 17 500
Ontario	40 000 & plus	10 000 à < 40 000	moins de 10 000
Manitoba	8 000 & plus	2 000 à < 8 000	moins de 2 000
Saskatchewan	7 000 & plus	1 500 à < 7 000	moins de 1 500
Alberta	25 000 & plus	6 500 à < 25 000	moins de 6 500
Colombie-Britannique	23 000 & plus	5 500 à < 23 000	moins de 5 500
Canada	44 500 & plus	11 000 à < 44 500	moins de 11 000

10.6 Seuils pour la diffusion des estimations pour l'Enquête de surveillance de l'usage du tabac au Canada - Fichier personnes

Le tableau suivant spécifie la taille minimum des estimations par province et groupe d'âge. Les estimations plus petites que la taille minimum présentées dans la colonne « Inacceptable » doivent être accompagnées d'une mise en garde appropriée.

Tableau des seuils pour la diffusion des estimations – Fichier personnes

Provinces	Groupes d'âge	Acceptable CV 0,0 % - 16,5 %	Médiocre CV 16,6 % - 33,3 %	Inacceptable CV > 33,3 %
Terre-Neuve-et-Labrador	Tous	31 500 & plus	8 000 à < 31 500	moins de 8 000
	15-19	6 000 & plus	1 500 à < 6 000	moins de 1 500
	20-24	9 500 & plus	3 000 à < 9 500	moins de 3 000
	25+	34 000 & plus	9 000 à < 34 000	moins de 9 000
Île-du-Prince-Édouard	Tous	7 000 & plus	2 000 à < 7 000	moins de 2 000
	15-19	2 000 & plus	500 à < 2 000	moins de 500
	20-24	2 000 & plus	500 à < 2 000	moins de 500
	25+	8 000 & plus	2 000 à < 8 000	moins de 2 000
Nouvelle-Ecosse	Tous	48 500 & plus	12 500 à < 48 500	moins de 12 500
	15-19	10 500 & plus	3 000 à < 10 500	moins de 3 000
	20-24	14 000 & plus	4 000 à < 14 000	moins de 4 000
	25+	53 000 & plus	14 000 à < 53 000	moins de 14 000
Nouveau-Brunswick	Tous	35 500 & plus	9 000 à < 35 500	moins de 9 000
	15-19	6 000 & plus	1 500 à < 6 000	moins de 1 500
	20-24	8 000 & plus	2 000 à < 8 000	moins de 2 000
	25+	39 000 & plus	10 000 à < 39 000	moins de 10 000
Québec	Tous	490 000 & plus	128 000 à < 490 000	moins de 128 000
	15-19	120 000 & plus	36 500 à < 120 000	moins de 36 500
	20-24	151 500 & plus	48 000 à < 151 500	moins de 48 000
	25+	539 500 & plus	144 000 à < 539 500	moins de 144 000
Ontario	Tous	330 000 & plus	83 000 à < 330 000	moins de 83 000
	15-19	72 000 & plus	19 000 à < 72 000	moins de 19 000
	20-24	88 500 & plus	23 500 à < 88 500	moins de 23 500
	25+	361 500 & plus	92 000 à < 361 500	moins de 92 000
Manitoba	Tous	55 500 & plus	14 500 à < 55 500	moins de 14 500
	15-19	13 000 & plus	3 500 à < 13 000	moins de 3 500
	20-24	15 000 & plus	4 500 à < 15 000	moins de 4 500
	25+	63 000 & plus	16 500 à < 63 000	moins de 16 500
Saskatchewan	Tous	46 500 & plus	12 000 à < 46 500	moins de 12 000
	15-19	13 500 & plus	4 000 à < 13 500	moins de 4 000
	20-24	18 500 & plus	5 500 à < 18 500	moins de 5 500
	25+	52 500 & plus	13 500 à < 52 500	moins de 13 500
Alberta	Tous	165 500 & plus	43 000 à < 165 500	moins de 43 000
	15-19	40 500 & plus	11 500 à < 40 500	moins de 11 500
	20-24	48 000 & plus	14 000 à < 48 000	moins de 14 000
	25+	187 500 & plus	49 500 à < 187 500	moins de 49 500
Colombie-Britannique	Tous	215 000 & plus	55 500 à < 215 000	moins de 55 500
	15-19	37 000 & plus	10 000 à < 37 000	moins de 10 000
	20-24	51 000 & plus	14 500 à < 51 000	moins de 14 500
	25+	247 000 & plus	65 000 à < 247 000	moins de 65 000
Canada	Tous	312 000 & plus	77 500 à < 312 000	moins de 77 500
	15-19	79 000 & plus	20 000 à < 79 000	moins de 20 000
	20-24	100 500 & plus	25 500 à < 100 500	moins de 25 500
	25+	354 500 & plus	88 000 à < 354 500	moins de 88 000

10.7 *Tables de CV - Fichier ménages*

Voir le fichier Ctums2002_C1_MN_CVTabSF.pdf pour les tables de CV pour le Fichier ménages du cycle 1 de 2002.

10.8 *Tables de CV - Fichier personnes*

Voir le fichier Ctums2002_C1_PR_CVTabSF.pdf pour les tables de CV pour le Fichier personnes du cycle 1 de 2002.

11.0 Pondération

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 ménage et sur un fichier différent un poids pour chaque personne.

La pondération pour le premier cycle de l'Enquête de surveillance sur l'usage du tabac au Canada comprend plusieurs étapes :

- calcul du poids de base,
- le facteur de compensation pour la non-réponse,
- un ajustement pour la sélection d'une ou deux personnes dans le ménage,
- l'élimination des enregistrements hors plan, et finalement
- un ajustement pour rendre les estimations de la population consistantes avec les totaux connus soient province-âge-sexe de la population estimée du recensement pour les personnes de 15 ans et plus.

11.1 Procédures de pondération pour les Fichiers ménages et personnes

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.

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

Il y avait 69 873 numéros de téléphone dans l'échantillon avec des poids.

2. Ajustement pour les numéros de téléphone non-résolus

Il y a eu 2 031 numéros de téléphone qui n'ont pas été résolus, laissant 67 842 numéros résolus. Nous n'avons pas pu déterminer si ces numéros non-résolus appartenaient à un ménage, une entreprise ou étaient hors du champ de l'enquête. Chaque numéro de téléphone avait un indicateur si on s'attendait que le numéro serait une résidence, une entreprise ou de genre inconnu. L'ajustement pour les numéros non-résolus a été fait par province-strate et le genre de numéro.

Pour chaque genre de ligne-province-strate,

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

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

Les numéros de téléphone correspondants 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 le numéro de téléphone était considéré comme étant celui d'un

ménage. Il y avait 40 224 numéros de téléphone qui étaient hors du champ de l'enquête et 27 618 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 était inconnu et que des données existaient pour le **ménage ou pour une personne**, la valeur de 1 a été imputé pour conserver des données. Après l'imputation, il restait 1 843 numéros de téléphone pour lesquels le nombre de lignes manquait toujours. Donc, il y avait 25 775 ménages dont l'information a été calculé ou imputé. L'ajustement a été fait au niveau province-strate.

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

5. Calcul du poids des ménages avec de multiples lignes téléphoniques

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) du ménage.

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

11.2 Procédures de pondération pour le Fichier ménages

6. Ajustement pour les ménages non-répondants

Les répondants du ménage ont répondu aux questions sur leurs habitudes de fumer. Si ces questions n'ont pas été complétées, refusées ou remplies partiellement, le ménage a été considéré comme non-répondant. Il y a eu 685 non-répondants. Donc 25 090 ménages pondérés dans le champ d'enquête ont été utilisés et ajustés par province-strate.

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

7. Ajustement pour totaux externes connus de ménages par strate

On a calculé un ajustement aux poids des ménages pour chaque enregistrement par province, strate et mois pour s'assurer que les estimations du nombre de ménages étaient consistantes avec des totaux externes connus de ménages. Le facteur d'ajustement par province-strate-mois (P-S-M) est défini comme :

$$W_6 = W_5 * \left(\frac{\text{Totaux externes connus de ménages pour } P-S-M}{\sum W_5 \text{ pour les ménages répondants dans l'échantillon par } P-S-M} \right)$$

Le poids des ménages, W_6 , obtenu après cette étape, est final et apparaît sur le fichier de microdonnées du ménage.

11.3 Procédures de pondération pour le Fichier personnes

6. Ajustement pour les ménages non-répondants

Sur le Fichier personnes, les répondants du ménage devaient répondre à la liste des membres sans aucun refus pour l'âge. Il y a eu 616 non-répondants. Donc 25 159 poids pour les ménages dans le champ d'enquête ont été utilisés et ajustés selon province-strate.

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

7. Calcul du poids de groupe

Tous les ménages répondants dans l'enquête ayant une liste complète des membres (aucun âge manquant) reçoivent un poids de groupe. À partir de la liste des membres, trois indicateurs sont placés pour indiquer la présence d'une personne dans les groupes d'âge 15-19, 20-24 et 25 et plus. Si un ou deux groupes d'âge sont présents, un individu a été sélectionné dans chaque groupe d'âge présent (c.-à-d. la probabilité de sélection du groupe est 1). Donc, le poids n'est pas augmenté. Toutefois, si les trois groupes d'âge sont présents, deux personnes ont été sélectionnées et la probabilité de sélectionner le groupe d'âge est de 2 sur 3 groupes d'âge. Le poids est donc augmenté par son inverse.

Si 1 ou 2 groupes d'âge présent(s), $W_6 = W_5$.

Si tous les 3 groupes d'âge présents, $W_6 = W_5 * 3/2$.

8. Éliminer les ménages pour lesquels personne n'a été sélectionné

Il y avait 14 036 ménages pour lesquels personne n'a été sélectionné pour compléter la portion sur l'usage du tabac ou la personne sélectionnée n'a pas été retenue à cause du sous-échantillonnage des individus. Ces ménages n'avaient donc aucune information au niveau de la personne. Approximativement 70 % des personnes de 25 ans et plus n'ont pas été retenues. Il y avait 11 123 ménages pour lesquels nous avons sélectionné un répondant. Dans 9 512 ménages nous avons sélectionné une personne et 1 611 où nous avons sélectionné deux personnes.

9. Calcul du poids des ménages pour les personnes sélectionnées

Ces $9\,512 + 2(1\,611) = 12\,734$ personnes sélectionnées sont associées avec les ménages répondants et ont conservées le poids correspondant, W_6 .

10. Calcul du sous-poids de la personne sélectionnée

Tous les individus dans la population de l'enquête ont reçu un poids. Le poids est augmenté par le nombre de personnes dans le groupe d'âge et l'inverse du facteur de sous-échantillonnage.

$$W_7 = W_6 * \left(\frac{\text{Nombre de personnes dans le groupe d'âge choisi}}{\text{Facteur de sous-échantillonnage}} \right)$$

Le facteur de sous-échantillonnage est de 1 pour les groupes d'âge 15-19 et 20-24. Le facteur de sous-échantillonnage pour le groupe d'âge 25 ans et plus avait été déjà assigné et variait de 23,0 % à 29,7 %, selon la province.

11. Ajustement pour les individus non-répondants

Le Fichier personnes comprend les enregistrements des répondants individuels qui ont donné l'information aux questions sur les habitudes de l'usage du tabac et qui ont fourni une date de naissance correspondant à l'âge indiqué dans la liste des membres. Il y a eu 1 302 non-répondants.

Donc, nous avons utilisé 11 432 poids des individus dans le champ d'enquête et ajusté selon la province, le groupe d'âge selon la liste des membres (15-19, 20-24, 25-44, 45-64, 65 et plus) et le sexe.

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

12. Ajustement pour totaux externes

On a calculé un ajustement aux poids des personnes pour chaque enregistrement par province, strate et mois pour s'assurer que les estimations de la population étaient consistantes avec des totaux externes de population pour les personnes de 15 ans et plus. Ceci est connu comme la post-stratification. Les totaux externes suivants ont été utilisés :

- 1) Totaux mensuels de population par province-strate, et
- 2) Pour le cycle 1 et le cycle 2 :
totaux de population par province, sexe et les groupes d'âge suivants :
15-19, 20-24, 25-34, 35-44, 45-54, 55-64, et 65 et plus. La moyenne de ces totaux a été faite pour la période d'enquête.

Pour le sommaire annuel :

totaux de population par province, sexe et les groupes d'âge suivants :
15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69 et 70 et plus. La moyenne de ces totaux a été faite pour la période d'enquête.

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 apparaissent au fichier de microdonnées de la personne.

12.0 Questionnaire

Voir le fichier Ctums2002_C1_QuestF.pdf pour le questionnaire français du cycle 1 de 2002.

13.0 Cliché d'enregistrement à valeurs univariées

13.1 Cliché d'enregistrement à valeurs univariées – Fichier ménages

Voir le fichier Ctums2002_C1_MN_LvCds.pdf pour le cliché d'articles avec les variables uniques pour le Fichier ménages du cycle 1 de 2002.

13.2 Cliché d'enregistrement à valeurs univariées – Fichier personnes

Voir le fichier Ctums2002_C1_PR_LvCds.pdf pour le cliché d'articles avec les variables uniques pour le Fichier personnes du cycle 1 de 2002.