

Guide de l'utilisateur des microdonnées

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

ANNUEL

FÉVRIER - DÉCEMBRE 2008



Statistique
Canada

Statistics
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	Population visée	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 dérivées	19
7.4	Pondération	19
7.5	Suppression des renseignements confidentiels	20
8.0	Qualité des données	23
8.1	Taux de réponse des ménages – février à décembre 2008	25
8.2	Taux de réponse des personnes – février à décembre 2008	27
8.3	Erreurs relatives à l'enquête	29
8.4	Non-réponse totale	30
8.5	Non-réponse partielle	30
8.6	Couverture	30
8.7	Mesure de l'erreur d'échantillonnage	30
9.0	Lignes directrices pour la totalisation, l'analyse et la diffusion de données	33
9.1	Lignes directrices pour l'arrondissement d'estimations	33
9.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation	34
9.3	Définitions de types d'estimations : catégoriques et quantitatives	34
9.3.1	Estimations catégoriques	34
9.3.2	Estimations quantitatives	34
9.3.3	Totalisation d'estimations catégoriques	35
9.3.4	Totalisation d'estimations quantitatives	35
9.4	Lignes directrices pour l'analyse statistique	36
9.5	Lignes directrices pour la diffusion de coefficients de variation	36
9.6	Seuils pour la diffusion des estimations pour l'Enquête de surveillance de l'usage du tabac au Canada - Fichier des ménages	38
9.7	Seuils pour la diffusion des estimations pour l'Enquête de surveillance de l'usage du tabac au Canada - Fichier des personnes	39

10.0	Tables de variabilité d'échantillonnage approximative	41
10.1	Comment utiliser les tables de coefficients de variation pour des estimations catégoriques.....	43
10.1.1	Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques	44
10.2	Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance	49
10.2.1	Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance	50
10.3	Comment utiliser les tables de coefficients de variation pour effectuer un test t.....	50
10.3.1	Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t	51
10.4	Coefficients de variation pour des estimations quantitatives.....	51
10.5	Tables des coefficients de variation – Fichier des ménages	52
10.6	Tables des coefficients de variation – Fichier des personnes	52
10.7	Méthode bootstrap moyenne pour estimer la variance.....	52
10.8	Progiciels statistiques pour estimer la variance.....	53
10.8.1	Autres progiciels.....	53
11.0	Pondération	57
11.1	Procédures de pondération pour les fichiers des ménages et des personnes.....	57
11.2	Procédures de pondération pour le fichier des ménages	58
11.3	Procédures de pondération pour le fichier des personnes	59
12.0	Questionnaire	61
13.0	Clichés d'enregistrement à valeurs univariées	63
13.1	Cliché d'enregistrement à valeurs univariées – Fichier des ménages	63
13.2	Cliché d'enregistrement à valeurs univariées – Fichier des personnes	63

1.0 Introduction

L'Enquête de surveillance de l'usage du tabac au Canada (ESUTC) a été menée par Statistique Canada entre février et décembre 2008, avec la collaboration et l'appui de Santé Canada. Le présent manuel a été produit pour faciliter la manipulation du fichier de microdonnées sur les résultats de l'enquête.

Toutes questions au sujet de l'ensemble de données ou de son utilisation devraient être adressées 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.gc.ca

Greg Phillips
Division des enquêtes spéciales
Immeuble Principal, 2^e étage
150, promenade Tunney's Pasture
Ottawa (Ontario) K1A 0T6
Téléphone : 613-951-3353
Télécopieur : 613-951-0562
Courriel : Greg.Phillips@statcan.gc.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

Judy Snider
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 A725
Ottawa (Ontario) K1A 0K9
Téléphone : 613-957-0697
Télécopieur : 613-954-2292
Courriel : Judy_Snider@hc-sc.gc.ca

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 décembre 2008.

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. L'ESUTC est la seule enquête de Statistique Canada qui répond au besoin de Santé Canada d'avoir une couverture continue dans le temps, des données disponibles rapidement ou suffisamment de détails sur les populations les plus à risque, à savoir les personnes de 15 à 24 ans. En comparaison, les mesures de la prévalence de l'usage du tabac de l'Enquête sur la santé dans les collectivités canadiennes sont occasionnelles et limitées.

L'Enquête de surveillance de l'usage du tabac au Canada permet à Santé Canada d'examiner la prévalence de l'usage du tabac selon la province, le sexe et le groupe d'âge (15 à 19 ans, 20 à 24 ans, 25 à 34 ans, 35 à 44 ans, 45 ans et plus) deux fois par an ainsi qu'annuellement. 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 de l'usage de la cigarette

- 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 actuel : Personne qui fume actuellement des cigarettes quotidiennement ou à l'occasion.

Antécédents de tabagisme

- 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 actuellement un fumeur ou un ancien fumeur.
- 5) Personne n'ayant jamais fumé : Personne qui a été un fumeur expérimental ou qui est un abstinente à perpétuité.

Prévalence de l'usage du tabac

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

Âge

L'information sur l'âge du répondant est obtenue à partir de deux sources : d'un répondant du ménage qui fournit l'âge de tous les membres du ménage (âge de la liste) et plus tard, au commencement de l'interview avec la personne sélectionnée, directement auprès de la personne répondante elle-même à qui on demande son âge. La variable DVAGE est l'âge fourni par le répondant sélectionné ou l'âge de la liste lorsque l'âge n'est pas disponible (p. ex. un refus).

5.0 Méthodologie de l'enquête

L'Enquête de surveillance de l'usage du tabac au Canada (ESUTC) a été réalisée entre le 1 février et le 31 décembre 2008. Cette enquête était téléphonique et 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 Population visée

La population cible pour l'ESUTC était composée de toutes les personnes âgées de 15 ans et plus vivant au Canada, à l'exception des personnes suivantes :

- 1) les résidents du Yukon, des Territoires du Nord-Ouest et du Nunavut,
- 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 ligne téléphonique terrestre ont été exclus de l'échantillon. Les personnes n'ayant pas de téléphone et ceux qui possèdent uniquement un téléphone cellulaire représentent à peu près 8 % de la population cible. Toutefois, les estimations de l'enquête ont été pondérées afin d'inclure les personnes n'ayant pas de ligne téléphonique terrestre.

5.2 Stratification

Afin de s'assurer que des personnes de partout au Canada étaient représentées dans l'échantillon, on a divisé chacune des 10 provinces en strates ou aires géographiques. En général, à l'intérieur de chaque province, une strate d'une région métropolitaine de recensement (RMR) et une strate d'une région autre qu'une RMR étaient délimitées. Dans l'Île-du-Prince-Édouard, il n'y avait qu'une strate. En Ontario et au Québec, il y avait une troisième strate pour Toronto et une troisième strate pour Montréal. Les RMR sont des régions définies aux fins du recensement et correspondent approximativement à des villes de 100 000 habitants ou plus.

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 et 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 10 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 et plus à l'échelle du Canada, c'est-à-dire de 500 personnes de chaque groupe d'âge par province par cycle. 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 âgées de 25 ans et plus ont été sous-échantillonnés. À l'origine, on a estimé qu'au total, presque 150 000 numéros de téléphone par année seraient nécessaires pour obtenir les 20 000 répondants par année requis. Cela suppose un taux de réponse de 72 % et qu'environ 23 % 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 %.

5.4 Tirage de l'échantillon

L'échantillon de l'ESUTC 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 des cinq prochains 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). Une activité de sélection visant à supprimer les numéros d'entreprise hors service et inconnus a été effectuée avant d'envoyer l'échantillon à l'unité d'interview téléphonique assistée par ordinateur (ITAO).

Chaque numéro de téléphone de l'échantillon de l'ITAO 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 77 % 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 3 % des ménages ne contiennent personne âgé 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 additionnelles, 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 selon le 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 et 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 et 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 et 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.

Taux de numéros de téléphone résolus et taux de succès selon la province - ancienne méthode

Provinces	Nombre de numéros de téléphone générés	Nombre de numéros de téléphone résolus	Taux de numéros de téléphone résolus (%)	Nombre total de ménages	Ménages avec données de liste valides	Taux d'achèvement des listes (%)	Taux de succès (%)
Terre-Neuve-et-Labrador	18 902	17 493	92,5	6 453	5 505	85,3	34,1
Île-du-Prince-Édouard	15 996	14 811	92,6	6 013	5 026	83,6	37,6
Nouvelle-Écosse	16 175	14 986	92,6	6 227	5 351	85,9	38,5
Nouveau-Brunswick	18 675	17 302	92,6	6 438	5 311	82,5	34,5
Québec	14 712	13 458	91,5	7 045	5 581	79,2	47,9
Ontario	14 510	13 201	91,0	5 975	4 424	74,0	41,2
Manitoba	14 721	13 559	92,1	5 974	4 852	81,2	40,6
Saskatchewan	14 540	13 232	91,0	6 289	4 931	78,4	43,3
Alberta	11 945	11 336	94,9	5 701	4 273	75,0	47,7
Colombie-Britannique	13 658	13 010	95,3	6 630	4 757	71,7	48,5
Canada	153 834	142 388	92,6	62 745	50 011	79,7	40,8

8.1 Taux de réponse des ménages – février à décembre 2008

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 13 446 (21,2 %) ménages non répondants dont 8 058 (12,7 % du total des ménages) ont refusé de participer.

Taux de réponse des ménages selon la province – nouvelle méthode

Provinces	Nombre total estimé de ménages	Nombre de ménages répondants	Taux de réponse des ménages (%)
Terre-Neuve-et-Labrador	6 620	5 483	82,8
Île-du-Prince-Édouard	6 081	5 010	82,4
Nouvelle-Écosse	6 296	5 333	84,7
Nouveau-Brunswick	6 535	5 291	81,0
Québec	7 182	5 564	77,5
Ontario	6 000	4 403	73,4
Manitoba	6 067	4 839	79,8
Saskatchewan	6 419	4 919	76,6
Alberta	5 555	4 254	76,6
Colombie-Britannique	6 525	4 739	72,6
Canada	63 281	49 835	78,8

En utilisant l'ancienne méthode 16 803 (25,2 %) des ménages étaient non répondants.

Taux de réponse des ménages selon la province - ancienne méthode

Provinces	Nombre total estimé de ménages	Nombre de ménages répondants	Taux de réponse des ménages (%)
Terre-Neuve-et-Labrador	6 894	5 483	79,5
Île-du-Prince-Édouard	6 452	5 010	77,6
Nouvelle-Écosse	6 624	5 333	80,5
Nouveau Brunswick	6 807	5 291	77,7
Québec	7 515	5 564	74,0
Ontario	6 366	4 403	69,2
Manitoba	6 382	4 839	75,8
Saskatchewan	6 827	4 919	72,1
Alberta	5 894	4 254	72,2
Colombie-Britannique	6 875	4 739	68,9
Canada	66 638	49 835	74,8

Taux de réponse des ménages selon le mois d'enquête - nouvelle méthode

Mois de l'enquête	Nombre total estimé de ménages	Nombre de ménages répondants	Taux de réponse des ménages (%)
février	5 910	4 538	76,8
mars	5 858	4 567	78,0
avril	5 846	4 642	79,4
mai	5 910	4 759	80,5
juin	5 864	4 662	79,5
juillet	5 720	4 420	77,3
août	5 668	4 416	77,9
septembre	5 614	4 461	79,5
octobre	5 638	4 452	79,0
novembre	5 646	4 506	79,8
décembre	5 608	4 412	78,7
Total	63 281	49 835	78,8

Taux de réponse des ménages selon le mois d'enquête - ancienne méthode

Mois de l'enquête	Nombre total de ménages	Nombre de ménages répondants	Taux de réponse des ménages (%)
février	6 189	4 538	73,3
mars	6 145	4 567	74,3
avril	6 148	4 642	75,5
mai	6 206	4 759	76,7
juin	6 204	4 662	75,1
juillet	6 046	4 420	73,1
août	5 978	4 416	73,9
septembre	5 921	4 461	75,3
octobre	5 910	4 452	75,3
novembre	5 965	4 506	75,5
décembre	5 925	4 412	74,5
Total	66 638	49 835	74,8

8.2 Taux de réponse des personnes – février à décembre 2008

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

- Le numéro de téléphone de la personne sélectionnée appartient à un ménage répondant.
- 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 15 ans ou plus au moment de l'interview (confirmé avec la personne sélectionnée).
- La personne sélectionnée a répondu au moins aux questions clés concernant les habitudes de fumer.

Dans 29 072 ménages, des données sur le ménage ont été recueillies mais personne n'a été sélectionné pour continuer l'ESUTC. (Pour plus d'information, voir (Tirage de l'échantillon) à la section 5.4.) Dans les ménages restants, on a sélectionné une personne dans 17 736 ménages et deux personnes dans 3 027 ménages. Le taux de refus au niveau des personnes était 4,4 %.

Taux de réponse des personnes selon la province

Provinces	Total des personnes sélectionnées	Total des personnes répondantes	Taux de réponse des personnes (%)
Terre-Neuve-et-Labrador	2 464	2 077	84,3
Île-du-Prince-Édouard	2 493	2 166	86,9
Nouvelle-Écosse	2 461	2 133	86,7
Nouveau-Brunswick	2 356	2 053	87,1
Québec	2 429	2 118	87,2
Ontario	2 214	1 860	84,0
Manitoba	2 427	2 151	88,6
Saskatchewan	2 465	2 181	88,5
Alberta	2 291	1 961	85,6
Colombie-Britannique	2 190	1 841	84,1
Canada	23 790	20 541	86,3

Taux de réponse des personnes selon le mois d'enquête

Mois de l'enquête	Total des personnes sélectionnées	Total des personnes répondantes	Taux de réponse des personnes (%)
février	2 255	1 955	86,7
mars	2 180	1 920	88,1
avril	2 256	1 963	87,0
mai	2 287	1 981	86,6
juin	2 177	1 900	87,3
juillet	2 135	1 780	83,4
août	2 101	1 807	86,0
septembre	2 071	1 750	84,5
octobre	2 098	1 804	86,0
novembre	2 135	1 860	87,1
décembre	2 095	1 821	86,9
Total	23 790	20 541	86,3

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

Groupes d'âge	Total des personnes sélectionnées	Total des personnes répondantes	Taux de réponses des personnes (%)
15 à 19	6 501	5 413	83,3
20 à 24	5 039	4 101	81,4
25 ans et plus	12 250	11 027	90,0
Total	23 790	20 541	86,3

Taux de réponse global selon la province – nouvelle méthode

Provinces	Taux de réponses des ménages (%)	Taux de réponses des personnes (%)	Taux de réponses global (%)
Terre-Neuve-et-Labrador	82,8	84,3	69,8
Île-du-Prince-Édouard	82,4	86,9	71,6
Nouvelle-Écosse	84,7	86,7	73,4
Nouveau-Brunswick	81,0	87,1	70,6
Québec	77,5	87,2	67,6
Ontario	73,4	84,0	61,6
Manitoba	79,8	88,6	70,7
Saskatchewan	76,6	88,5	67,8
Alberta	76,6	85,6	65,5
Colombie-Britannique	72,6	84,1	61,1
Canada	78,8	86,3	68,0

Taux de réponse global selon la province – ancienne méthode

Provinces	Taux de réponses des ménages (%)	Taux de réponses des personnes (%)	Taux de réponses global (%)
Terre-Neuve-et-Labrador	79,5	84,3	67,0
Île-du-Prince-Édouard	77,6	86,9	67,5
Nouvelle-Écosse	80,5	86,7	69,8
Nouveau-Brunswick	77,7	87,1	67,7
Québec	74,0	87,2	64,6
Ontario	69,2	84,0	58,1
Manitoba	75,8	88,6	67,2
Saskatchewan	72,1	88,5	63,8
Alberta	72,2	85,6	61,8
Colombie-Britannique	68,9	84,1	57,9
Canada	74,8	86,3	64,6

8.3 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 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.

8.4 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. S'il y a eu non-réponse totale, c'est parce que l'intervieweur a été incapable de communiquer avec le répondant ou que le répondant a refusé de participer à l'enquête. Les cas de non-réponse totale ont été traités en ajustant le poids des ménages ou des personnes qui ont répondu au questionnaire d'enquête de façon à le contrebalancer pour ceux qui n'y ont pas répondu.

8.5 Non-réponse partielle

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. Des codes dans le fichier de microdonnées indiquent les cas de non-réponse partielle, c'est-à-dire refus, ne sait pas.

8.6 Couverture

Tel qu'il est mentionné à la section 5.1 (Population visée), à peu près 8 % des ménages au Canada n'ont pas de ligne téléphonique terrestre. 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 ou possèdent seulement un téléphone cellulaire.

8.7 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 annuelle de 2002, l'on estime que 21,4 % des Canadiens fument actuellement la cigarette, et l'on constate que l'erreur-type de cette estimation est de 0,0039. Le coefficient de variation de l'estimation est donc calculé comme suit :

$$\left(\frac{0,0039}{0,214} \right) \times 100 \% = 1,8 \%$$

De plus amples renseignements sur le calcul du coefficient 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 des 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 techniques 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 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 de surveillance de l'usage du tabac au Canada (ESUTC) n'était pas autopondéré. Lorsqu'ils produisent des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent appliquer le poids d'enquête approprié.

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'ESUTC, 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'ESUTC.

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 de personnes qui fument actuellement des cigarettes ou la proportion de fumeurs quotidiens qui ont tenté de cesser de fumer constituent des exemples de telles estimations. Une estimation du nombre de personnes possédant une certaine caractéristique peut aussi être désignée une estimation d'un agrégat.

Exemples de questions catégoriques :

Q : Avez-vous fumé la cigarette au cours des 30 derniers jours?
R : Oui / Non

Q : Quelle était votre principale raison pour cesser de fumer?
R : Santé / Grossesse ou un bébé dans le ménage / Moins de stress dans la vie / Le coût des cigarettes / Fumer est moins acceptable socialement / Une autre raison

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 cigarettes fumées, les

samedis, par personne. Le numérateur (\hat{X}) est une estimation du nombre total de cigarettes fumées les samedis et son dénominateur (\hat{Y}) est le nombre de personnes ayant déclaré avoir fumé les samedis.

Exemples de questions quantitatives :

Q : Chez certains fumeurs, le nombre de cigarettes fumées dépend du jour de la semaine. Donc, 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

9.3.3 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 les proportions et les 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. Pour obtenir, par exemple, une estimation du nombre total de cigarettes fumées le samedi, multipliez la valeur déclarée à la question WP_Q10F (nombre de cigarettes fumées le samedi) par le poids final de l'enregistrement, puis additionnez cette valeur pour tous les enregistrements où la variable WP_Q10F < 96 (tous les répondants qui ont donné une réponse à ce champ).

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. Pour estimer, par exemple, le nombre moyen de cigarettes fumées le samedi,

- estimez le nombre total de cigarettes fumées le samedi (\hat{X}), tel qu'il est décrit ci-dessus,
- estimez le nombre de personnes (\hat{Y}) incluses dans cette catégorie en additionnant les poids finals de tous les enregistrements où la variable WP_Q10F < 96, puis
- divisez l'estimation a) par l'estimation b) (\hat{X} / \hat{Y}).

9.4 Lignes directrices pour l'analyse statistique

L'Enquête de surveillance de l'usage du tabac au Canada 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 des personnes 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'Enquête de surveillance de l'usage du tabac au Canada, 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'échantillons 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 pour l'Enquête de surveillance de l'usage du tabac au Canada - Fichier des ménages

La taille minimale des estimations pour les ménages au niveau provincial est indiquée dans le tableau ci-dessous. Les estimations plus petites que la taille minimale indiquée dans la colonne « Inacceptable » doivent être signalées de façon appropriée.

Tableau des seuils pour la diffusion des estimations – Fichier des ménages

Provinces	CV acceptable 0,0 à 16,5 %	CV médiocre 16,6 à 33,3 %	CV inacceptable > 33,3 %
Terre-Neuve-et-Labrador	1 500 et plus	500 à < 1 500	moins de 500
Île-du-Prince-Édouard	500 et plus	250 à < 500	moins de 250
Nouvelle-Écosse	3 000 et plus	500 à < 3 000	moins de 500
Nouveau-Brunswick	2 500 et plus	500 à < 2 500	moins de 500
Québec	23 500 et plus	6 000 à < 23 500	moins de 6 000
Ontario	44 000 et plus	11 000 à < 44 000	moins de 11 000
Manitoba	3 500 et plus	1 000 à < 3 500	moins de 1 000
Saskatchewan	3 000 et plus	1 000 à < 3 000	moins de 1 000
Alberta	12 000 et plus	3 000 à < 12 000	moins de 3 000
Colombie-Britannique	14 500 et plus	3 500 à < 14 500	moins de 3 500
Canada	24 500 et plus	6 000 à < 24 500	moins de 6 000

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

Le tableau suivant spécifie la taille minimale des estimations selon la province et le groupe d'âge. Les estimations plus petites que la taille minimale présentée dans la colonne « Inacceptable » doivent être signalées de façon appropriée.

Tableau des seuils pour la diffusion des estimations – Fichier des personnes

Provinces	Groupes d'âge	CV acceptable 0,0 à 16,5 %	CV médiocre 16,6 à 33,3 %	CV inacceptable > 33,3%
Terre-Neuve-et-Labrador	Tous	12 500 et plus	3 000 à < 12 500	moins de 3 000
	15 à 19	2 500 et plus	500 à < 2 500	moins de 500
	20 à 24	3 000 et plus	1 000 à < 3 000	moins de 1 000
	25+	14 500 et plus	3 500 à < 14 500	moins de 3 500
Île-du-Prince-Édouard	Tous	3 000 et plus	500 à < 3 000	moins de 500
	15 à 19	1 000 et plus	250 à < 1 000	moins de 250
	20 à 24	1 000 et plus	500 à < 1 000	moins de 500
	25+	3 500 et plus	1 000 à < 3 500	moins de 1 000
Nouvelle-Écosse	Tous	21 000 et plus	5 500 à < 21 000	moins de 5 500
	15 à 19	5 000 et plus	1 500 à < 5 000	moins de 1 500
	20 à 24	6 500 et plus	1 500 à < 6 500	moins de 1 500
	25+	24 500 et plus	6 000 à < 24 500	moins de 6 000
Nouveau-Brunswick	Tous	18 000 et plus	4 500 à < 18 000	moins de 4 500
	15 à 19	4 000 et plus	1 000 à < 4 000	moins de 1 000
	20 à 24	5 000 et plus	1 500 à < 5 000	moins de 1 500
	25+	20 000 et plus	5 000 à < 20 000	moins de 5 000
Québec	Tous	179 500 et plus	45 000 à < 179 500	moins de 45 000
	15 à 19	39 000 et plus	10 000 à < 39 000	moins de 10 000
	20 à 24	46 000 et plus	12 000 à < 46 000	moins de 12 000
	25+	202 500 et plus	51 000 à < 202 500	moins de 51 000
Ontario	Tous	349 500 et plus	88 000 à < 349 500	moins de 88 000
	15 à 19	64 000 et plus	16 500 à < 64 000	moins de 16 500
	20 à 24	95 000 et plus	25 500 à < 95 000	moins de 25 500
	25+	394 000 et plus	100 000 à < 394 000	moins de 100 000
Manitoba	Tous	26 500 et plus	6 500 à < 26 500	moins de 6 500
	15 à 19	7 000 et plus	2 000 à < 7 000	moins de 2 000
	20 à 24	8 000 et plus	2 000 à < 8 000	moins de 2 000
	25+	30 000 et plus	7 500 à < 30 000	moins de 7 500
Saskatchewan	Tous	20 000 et plus	5 000 à < 20 000	moins de 5 000
	15 à 19	6 000 et plus	1 500 à < 6 000	moins de 1 500
	20 à 24	8 500 et plus	2 000 à < 8 500	moins de 2 000
	25+	23 500 et plus	6 000 à < 23 500	moins de 6 000
Alberta	Tous	82 000 et plus	20 500 à < 82 000	moins de 20 500
	15 à 19	22 500 et plus	6 000 à < 22 500	moins de 6 000
	20 à 24	30 500 et plus	8 000 à < 30 500	moins de 8 000
	25+	97 000 et plus	24 500 à < 97 000	moins de 24 500
Colombie-Britannique	Tous	122 000 et plus	30 500 à < 122 000	moins de 30 500
	15 à 19	24 500 et plus	6 500 à < 24 500	moins de 6 500
	20 à 24	37 500 et plus	10 000 à < 37 500	moins de 10 000
	25+	146 000 et plus	37 000 à < 146 000	moins de 37 000
Canada	Tous	205 500 et plus	50 500 à < 205 500	moins de 50 500
	15 à 19	39 500 et plus	10 000 à < 39 500	moins de 10 000
	20 à 24	61 000 et plus	15 500 à < 61 000	moins de 15 500
	25+	238 000 et plus	59 000 à < 238 000	moins de 59 000

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 de surveillance de l'usage du tabac au Canada (ESUTC) pour le fichier des ménages.

Fichier des ménages

Provinces	Effet du plan	Taille de l'échantillon	Population
Terre-Neuve-et-Labrador	1,08	5 483	202 637
Île-du-Prince-Édouard	1,07	5 010	55 116
Nouvelle-Écosse	1,06	5 333	385 383
Nouveau-Brunswick	1,07	5 291	305 707
Québec	1,08	5 564	3 328 195
Ontario	1,10	4 403	4 840 546
Manitoba	1,04	4 839	459 064
Saskatchewan	1,10	4 919	389 315
Alberta	1,06	4 254	1 331 143
Colombie-Britannique	1,07	4 739	1 759 465
Canada	2,54	49 835	13 056 570

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 et le groupe d'âge qui ont été utilisés pour produire les tables de variabilité d'échantillonnage approximative de l'ESUTC pour le fichier des personnes.

Fichier des personnes

Provinces	Groupes d'âge	Effet du plan	Taille de l'échantillon	Population
Terre-Neuve-et-Labrador	Tous	1,70	2 077	427 504
	15 à 19	1,21	562	31 936
	20 à 24	1,19	409	30 744
	25+	1,24	1 106	364 824
Île-du-Prince-Édouard	Tous	1,52	2 166	115 009
	15 à 19	1,28	577	10 310
	20 à 24	1,31	371	9 508
	25+	1,21	1 218	95 191
Nouvelle-Écosse	Tous	1,60	2 133	783 196
	15 à 19	1,38	551	61 454
	20 à 24	1,31	420	62 431
	25+	1,21	1 162	659 311
Nouveau-Brunswick	Tous	1,63	2 053	628 391
	15 à 19	1,24	537	48 192
	20 à 24	1,34	430	48 787
	25+	1,16	1 086	531 412
Québec	Tous	1,66	2 118	6 422 438
	15 à 19	1,37	588	493 028
	20 à 24	1,26	435	479 182
	25+	1,15	1 095	5 450 228
Ontario	Tous	1,73	1 860	10 586 147
	15 à 19	1,14	528	867 750
	20 à 24	1,20	364	881 817
	25+	1,23	968	8 836 580
Manitoba	Tous	1,67	2 151	955 724
	15 à 19	1,25	538	86 440
	20 à 24	1,25	452	85 412
	25+	1,25	1 161	783 872
Saskatchewan	Tous	1,51	2 181	807 004
	15 à 19	1,31	571	76 233
	20 à 24	1,36	406	75 746
	25+	1,22	1 204	655 026
Alberta	Tous	1,59	1 961	2 841 559
	15 à 19	1,36	488	241 542
	20 à 24	1,37	402	274 404
	25+	1,27	1 071	2 325 613
Colombie-Britannique	Tous	1,70	1 841	3 714 637
	15 à 19	1,33	516	284 113
	20 à 24	1,25	336	311 067
	25+	1,32	989	3 119 457
Canada	Tous	4,24	20 541	27 281 608
	15 à 19	2,72	5 456	2 200 997
	20 à 24	3,04	4 025	2 259 097
	25+	3,17	11 060	22 821 514

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. Les utilisateurs intéressés à calculer les variances exactes pourront se procurer, sans frais, le fichier contenant les poids « bootstrap » ainsi que les programmes nécessaires pour générer les estimations de diverses statistiques.

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 devrait être considérée *inacceptable* et devrait être signalée de façon pertinente, 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 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. La proportion, par exemple, d'anciens fumeurs qui ont cessé pour des problèmes actuels de santé est plus fiable que le nombre estimé d'anciens fumeurs qui ont cessé pour des problèmes actuels de santé. (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., comme ses membres d'un sexe ou d'un groupe d'âge 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 ou 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 fumeurs et le numérateur, le nombre de fumeurs quotidiens.

Dans le cas où le numérateur n'est pas un sous-ensemble du dénominateur, comme dans l'exemple du rapport du nombre de fumeurs quotidiens comparativement au nombre de non-fumeurs, 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 annuel de 2002 et sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter. Veuillez noter que les données utilisées dans ces exemples diffèrent des véritables résultats de l'enquête et tiennent seulement lieu de guide.

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

Supposons qu'un utilisateur estime que durant la période de référence, 5 414 335 de personnes étaient des fumeurs actuels (DVSST1 = 1) au Canada. 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 – Tous les âges du fichier des personnes.

Enquête de surveillance de l'usage du tabac au Canada, 2002 - février à décembre - Fichier des personnes														
Tables de variabilité d'échantillonnage approximative pour le Canada - Tous les âges														
NUMÉRATEUR DU POURCENTAGE ('000)	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	197,2	196,3	195,3	192,3	187,1	181,9	176,4	170,8	165,0	159,0	152,8	139,5	108,0	62,4
2	139,4	138,8	138,1	135,9	132,3	128,6	124,8	120,8	116,7	112,5	108,0	98,6	76,4	44,1
3	113,8	113,3	112,7	111,0	108,0	105,0	101,9	98,6	95,3	91,8	88,2	80,5	62,4	36,0
4	98,6	98,1	97,6	96,1	93,6	90,9	88,2	85,4	82,5	79,5	76,4	69,7	54,0	31,2
5	88,2	87,8	87,3	86,0	83,7	81,3	78,9	76,4	73,8	71,1	68,3	62,4	48,3	27,9
...
75	*****	22,7	22,5	22,2	21,6	21,0	20,4	19,7	19,1	18,4	17,6	16,1	12,5	7,2
80	*****	21,9	21,8	21,5	20,9	20,3	19,7	19,1	18,5	17,8	17,1	15,6	12,1	7,0
85	*****	21,3	21,2	20,9	20,3	19,7	19,1	18,5	17,9	17,2	16,6	15,1	11,7	6,8
90	*****	20,7	20,6	20,3	19,7	19,2	18,6	18,0	17,4	16,8	16,1	14,7	11,4	6,6
95	*****	20,1	20,0	19,7	19,2	18,7	18,1	17,5	16,9	16,3	15,7	14,3	11,1	6,4
100	*****	19,6	19,5	19,2	18,7	18,2	17,6	17,1	16,5	15,9	15,3	13,9	10,8	6,2
125	*****	17,6	17,5	17,2	16,7	16,3	15,8	15,3	14,8	14,2	13,7	12,5	9,7	5,6
150	*****	16,0	15,9	15,7	15,3	14,8	14,4	13,9	13,5	13,0	12,5	11,4	8,8	5,1
200	*****	13,9	13,8	13,6	13,2	12,9	12,5	12,1	11,7	11,2	10,8	9,9	7,6	4,4
250	*****	12,4	12,4	12,2	11,8	11,5	11,2	10,8	10,4	10,1	9,7	8,8	6,8	3,9
300	*****	*****	11,3	11,1	10,8	10,5	10,2	9,9	9,5	9,2	8,8	8,1	6,2	3,6
350	*****	*****	10,4	10,3	10,0	9,7	9,4	9,1	8,8	8,5	8,2	7,5	5,8	3,3
400	*****	*****	9,8	9,6	9,4	9,1	8,8	8,5	8,3	8,0	7,6	7,0	5,4	3,1
450	*****	*****	9,2	9,1	8,8	8,6	8,3	8,1	7,8	7,5	7,2	6,6	5,1	2,9
500	*****	*****	8,7	8,6	8,4	8,1	7,9	7,6	7,4	7,1	6,8	6,2	4,8	2,8
750	*****	*****	*****	7,0	6,8	6,6	6,4	6,2	6,0	5,8	5,6	5,1	3,9	2,3
1000	*****	*****	*****	6,1	5,9	5,8	5,6	5,4	5,2	5,0	4,8	4,4	3,4	2,0
1500	*****	*****	*****	*****	4,8	4,7	4,6	4,4	4,3	4,1	3,9	3,6	2,8	1,6
2000	*****	*****	*****	*****	4,2	4,1	3,9	3,8	3,7	3,6	3,4	3,1	2,4	1,4
3000	*****	*****	*****	*****	*****	3,3	3,2	3,1	3,0	2,9	2,8	2,5	2,0	1,1
4000	*****	*****	*****	*****	*****	*****	2,8	2,7	2,6	2,5	2,4	2,2	1,7	1,0
5000	*****	*****	*****	*****	*****	*****	2,5	2,4	2,3	2,2	2,2	2,0	1,5	0,9
6000	*****	*****	*****	*****	*****	*****	*****	2,2	2,1	2,1	2,0	1,8	1,4	0,8
7000	*****	*****	*****	*****	*****	*****	*****	*****	2,0	1,9	1,8	1,7	1,3	0,7
8000	*****	*****	*****	*****	*****	*****	*****	*****	*****	1,8	1,7	1,6	1,2	0,7
9000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1,6	1,5	1,1	0,7
10000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1,5	1,4	1,1	0,6
12500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1,2	1,0	0,6
15000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	0,9	0,5

NOTA : Pour utiliser ces tables correctement, veuillez vous référer à la documentation reliée aux microdonnées.

- 2) L'agrégat estimé (5 414 335) 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 5 000 000.

- 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 2,5 %.
- 4) Le coefficient de variation approximatif de l'estimation est donc 2,5 %. Le résultat selon lequel il y avait 5 414 335 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1) de fumeurs actuels durant la période de référence, peut être publié sans réserve.

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

Supposons qu'un utilisateur estime à $2\,865\,929 / 12\,436\,728 = 23,0\%$ la proportion d'hommes qui fument actuellement au Canada durant la période de référence. 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 du fichier des personnes (voir ci-dessus). Le tableau au niveau du CANADA devrait être 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 fondé sur un sous-ensemble de la population totale (c'est-à-dire les hommes), il faut utiliser à la fois le pourcentage (23,0 %) et la portion numérateur du pourcentage (2 865 929) pour déterminer le coefficient de variation.
- 3) Le numérateur, 2 865 929, 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 3 000 000. 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 25,0 %.
- 4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 3,1 %, est le coefficient de variation à employer.
- 5) Le coefficient de variation approximatif de l'estimation est donc 3,1 %. Le résultat selon lequel 23,0 % des hommes fument actuellement, peut être publié sans réserve.

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

Supposons qu'un utilisateur estime à $2\,548\,406 / 12\,814\,359 = 19,9\%$ la proportion de femmes qui fument actuellement au Canada et à $2\,865\,929 / 12\,436\,728 = 23,0\%$ la proportion d'hommes qui fument actuellement au Canada. 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 pour le CANADA du fichier des personnes (voir ci-dessus) de la même façon que celle décrite dans l'exemple 2, donne un CV de l'estimation pour les femmes de 3,2 % et un CV de l'estimation pour les hommes de 3,1 %.
- 2) 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 (hommes), \hat{X}_2 est l'estimation 2 (femmes) 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 $\hat{d} = 0,230 - 0,199 = 0,031$ est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0,230)(0,031)]^2 + [(0,199)(0,032)]^2} \\ &= \sqrt{(0,00005) + (0,00004)} \\ &= 0,009\end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,009 / 0,031 = 0,290$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 29,0 %. La différence entre les estimations est considérée médiocre 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 E (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 qu'un utilisateur estime à 237 261 le nombre de femmes âgées de 15 à 19 ans qui fument actuellement et à 220 511 le nombre d'hommes âgés de 15 à 19 ans qui fument actuellement. 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 coefficient de variation de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation d'un rapport, où le numérateur de l'estimation (\hat{X}_1) est le nombre de femmes âgées de 15 à 19 ans qui fument actuellement. Le dénominateur de l'estimation (\hat{X}_2) est le nombre d'hommes âgés de 15 à 19 ans qui fument actuellement.
- 2) Reportez-vous à la table des coefficients de variation pour le CANADA – 15 à 19 ans du fichier des personnes.

Enquête de surveillance de l'usage du tabac au Canada, 2002 - février à décembre - Fichier des personnes														
Tables de variabilité d'échantillonnage approximative pour le Canada - 15 à 19 ans														
NUMÉRATEUR DU POURCENTAGE (000)	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	95,8	95,3	94,9	93,4	90,9	88,3	85,7	83,0	80,2	77,3	74,2	67,8	52,5	30,3
2	67,7	67,4	67,1	66,0	64,3	62,5	60,6	58,7	56,7	54,6	52,5	47,9	37,1	21,4
3	*****	55,0	54,8	53,9	52,5	51,0	49,5	47,9	46,3	44,6	42,9	39,1	30,3	17,5
4	*****	47,7	47,4	46,7	45,5	44,2	42,9	41,5	40,1	38,6	37,1	33,9	26,2	15,2
5	*****	42,6	42,4	41,8	40,7	39,5	38,3	37,1	35,9	34,6	33,2	30,3	23,5	13,6
6	*****	38,9	38,7	38,1	37,1	36,1	35,0	33,9	32,7	31,5	30,3	27,7	21,4	12,4
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
95	*****	*****	*****	9,6	9,3	9,1	8,8	8,5	8,2	7,9	7,6	7,0	5,4	3,1
100	*****	*****	*****	9,3	9,1	8,8	8,6	8,3	8,0	7,7	7,4	6,8	5,2	3,0
125	*****	*****	*****	*****	8,1	7,9	7,7	7,4	7,2	6,9	6,6	6,1	4,7	2,7
150	*****	*****	*****	*****	7,4	7,2	7,0	6,8	6,5	6,3	6,1	5,5	4,3	2,5
200	*****	*****	*****	*****	6,4	6,2	6,1	5,9	5,7	5,5	5,2	4,8	3,7	2,1
250	*****	*****	*****	*****	*****	5,6	5,4	5,2	5,1	4,9	4,7	4,3	3,3	1,9
300	*****	*****	*****	*****	*****	5,1	4,9	4,8	4,6	4,5	4,3	3,9	3,0	1,7
350	*****	*****	*****	*****	*****	*****	4,6	4,4	4,3	4,1	4,0	3,6	2,8	1,6
400	*****	*****	*****	*****	*****	*****	4,3	4,1	4,0	3,9	3,7	3,4	2,6	1,5
450	*****	*****	*****	*****	*****	*****	*****	3,9	3,8	3,6	3,5	3,2	2,5	1,4
500	*****	*****	*****	*****	*****	*****	*****	3,7	3,6	3,5	3,3	3,0	2,3	1,4
750	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,7	2,5	1,9
1000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,1	1,7
1500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	0,8

NOTA : Pour utiliser ces tables correctement, veuillez vous référer à la documentation reliée aux microdonnées.

- 3) Le numérateur de cette estimation de rapport est 237 261. Le chiffre qui s'en rapproche le plus est 250 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 5,6 %.
- 4) Le dénominateur de cette estimation de rapport est 220 511. Le chiffre qui s'en rapproche le plus est 200 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 6,4 %.
- 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 :

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0,056)^2 + (0,064)^2} \\ &= \sqrt{0,003136 + 0,004096} \\ &= 0,085\end{aligned}$$

- 6) Le rapport obtenu entre les femmes et les hommes âgés de 15 à 19 ans qui fument actuellement est 237 261 / 220 511, c'est-à-dire 1,08 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1). Le coefficient de variation de cette estimation est 8,5 %, ce qui fait qu'on peut la diffuser sans réserve.

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 des hommes qui fument actuellement (d'après l'exemple 2 à la section 10.1.1) serait calculé comme suit :

$$\hat{X} = 23,0 \% \text{ (ou exprimé sous forme de proportion } 0,230)$$

$$t = 2$$

$\alpha_{\hat{x}}$ = 3,1 % (0,031 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.

$$IC_{\hat{x}} = \{0,230 - (2) (0,230) (0,031), 0,230 + (2) (0,230) (0,031)\}$$

$$IC_{\hat{x}} = \{0,230 - 0,014, 0,230 + 0,014\}$$

$$IC_{\hat{x}} = \{0,216, 0,244\}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 21,6 % et 24,4 % des hommes fument actuellement.

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 a pas de différence entre la proportion d'hommes qui fument actuellement et la proportion de femmes qui fument actuellement. 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 0,009. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_d} = \frac{0,230 - 0,199}{0,009} = \frac{0,031}{0,009} = 3,44$$

Puisque $t = 3,44$ est plus grand que 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'Enquête de surveillance sur l'usage du tabac au Canada 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 cigarettes fumées le samedi serait supérieur au coefficient de variation de la proportion correspondante de fumeurs actuels. 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 des coefficients de variation – Fichier des ménages

Consulter le fichier ESUTC2008_Annuel_MN_CVTabF.pdf pour les tables de coefficient de variation pour le fichier des ménages de 2008.

10.6 Tables des coefficients de variation – Fichier des personnes

Consulter le fichier ESUTC2008_Annuel_PR_CVTabF.pdf pour les tables de coefficient de variation pour le fichier des personnes de 2008.

10.7 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'ESUTC, 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'ESUTC, 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'ESUTC, 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.

10.8 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. 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.

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

10.8.1 Autres progiciels

Une variable du poids d'enquête ayant une série correspondante de 250 variables de poids bootstrap moyens accompagne de nombreux fichiers de données de l'ESUTC afin qu'une approche entièrement fondée sur le plan d'enquête puisse être adoptée pour l'analyse des données.

L'approche de l'analyse fondée sur le plan d'enquête suppose d'abord l'utilisation de la variable du poids d'enquête en vue d'obtenir les estimations pondérées des quantités d'intérêt. On utilise ensuite des renseignements additionnels sur le plan d'enquête pour estimer les variances¹ (et covariances) des quantités estimées. Dans le cas des fichiers de microdonnées à grande diffusion (FMGD) de l'ESUTC, ces renseignements additionnels comprennent 250 variables de poids bootstrap moyens de l'enquête, où chaque poids bootstrap moyen est dérivé de 20 échantillons d'enquête bootstrap indépendants. On peut alors utiliser les estimations et estimations des variances fondées sur le plan d'enquête pour tirer les conclusions nécessaires dans le cadre de l'analyse.

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

Supposons que $\hat{\beta}$ soit l'estimation pondérée de la quantité d'intérêt, β , calculée au moyen de la variable du poids de l'enquête w , et que $\hat{\beta}^{(b)}$ soit une estimation obtenue exactement de la même façon, à l'exception de la substitution de la variable de poids bootstrap moyen b pour la variable de poids d'enquête w , $b=1,2,\dots,250$. Cela donne des estimations bootstrap (moyen) w de $\hat{\beta}^{(1)}, \dots, \hat{\beta}^{(250)}$. L'estimation habituelle du poids bootstrap moyen de la variance de $\hat{\beta}$ est donc

$$\hat{V}_B(\hat{\beta}) = \frac{20}{250} \sum_{b=1}^{250} (\hat{\beta}^{(b)} - \hat{\beta})^2 \quad (1)$$

Si $\hat{\beta}$ est un vecteur plutôt qu'une valeur unique, comme si $\hat{\beta}$ était l'ensemble des coefficients d'un modèle, alors la matrice des estimations des variances et covariances des éléments de $\hat{\beta}$ est $\hat{V}_B(\hat{\beta}) = \frac{20}{250} \sum_{b=1}^{250} (\hat{\beta}^{(b)} - \hat{\beta})(\hat{\beta}^{(b)} - \hat{\beta})'$.

(La valeur « 20 » comprise dans la formule provient du fait que chaque poids bootstrap moyen de l'ESUTC est créé à partir de 20 échantillons bootstrap. La valeur « 250 » figurant dans la formule provient du fait que nous avons 250 poids bootstrap moyens différents. Si le nombre d'échantillons bootstrap utilisés pour créer chaque variable de poids bootstrap moyen (20) ou le nombre de variables de poids bootstrap moyen (250) venait à changer, les valeurs figurant dans la formule (1) devraient être modifiées.)

La méthode bootstrap moyenne n'est qu'une méthode de répétition parmi d'autres que l'on peut utiliser pour obtenir des estimations de variances fondées sur le plan d'enquête à partir des données d'enquête. Même si plusieurs logiciels commerciaux d'analyse fondée sur le plan d'enquête offrent des méthodes de répétition pour l'estimation de variances, ils ne désignent généralement pas la méthode bootstrap moyenne comme faisant partie de leurs méthodes. Toutefois, en raison de la ressemblance de forme de l'estimation des variances pour le bootstrap moyen et pour la méthode de réplication particulière appelée BRR (répliques répétées équilibrées) avec un ajustement de Fay, on peut utiliser des programmes pouvant effectuer des estimations de variances au moyen de cette méthode avec les poids de répétition fournis par l'utilisateur pour obtenir des estimations des variances de poids bootstrap moyens. Plus particulièrement, dans ces logiciels, les 250 poids bootstrap moyens fournis dans les FMGD de l'ESUTC doivent être désignés comme 250 poids BRR, et le facteur d'ajustement de Fay doit prendre la valeur $1 - \sqrt{1/20} \approx 0.7764$.

Dans les sections ci-dessous, des directives seront données sur la mise en oeuvre de l'estimation de la variance des poids bootstrap moyens à partir des données des FMGD de l'ESUTC, au moyen de trois logiciels commerciaux différents qui peuvent effectuer une certaine analyse fondée sur le plan d'enquête pour le BRR avec un ajustement de Fay :

- Stata 9 ou 10,
- SUDAAN et
- WesVar.

Ces méthodes sont adaptées pour l'ESUTC d'après un article d'Owen Phillips intitulé *Comment utiliser les poids bootstrap avec Wes Var et SUDAAN* (n° 12-002-X20040027032 au catalogue) paru dans « *Le Bulletin technique et d'information des Centres de données de recherche (Index chronologique)* », automne 2004, vol.1 no 2, Statistique Canada, n° 12-002-XIF au catalogue. Dans tous les cycles de l'ESUTC où des poids bootstrap moyens sont fournis, les noms donnés à ces variables bootstrap dans la documentation de l'utilisateur sont de **wrpp0001** à **wrpp0250** pour les fichiers au niveau de la personne, et de **wrhp0001** à **wrhp0250** pour les fichiers au niveau du ménage. Le nom de la variable du poids d'enquête est soit **wtp** ou **wthp** respectivement.

Stata 9 ou 10

À partir de la version 9, le logiciel commercial Stata contient des méthodes de répétition additionnelles permettant d'effectuer des estimations de variances fondées sur le plan

d'enquête dans ses commandes d'analyse d'enquêtes. La méthode BRR avec ajustement de Fay fait partie des méthodes de répétition offertes; c'est cette méthode qui serait précisée dans l'analyse des données de l'ESUTC. Pour préciser cette méthode, voici la marche à suivre recommandée :

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

svyset [pweight=wtp], vce(brr) fay(.7764) brrweight(wrpp0001-wrpp0250) mse

L'énoncé **pweight=wtp** indique à Stata que le poids d'enquête (souvent appelé le poids de probabilité) est la variable wtp. L'option **vce(brr)** déclare que la méthode d'estimation de la variance à utiliser est BRR. L'option **fay(.7764)** déclare que la méthode d'estimation de la variance BRR doit utiliser un ajustement de Fay de 0,7764. L'option **brrweight(wrpp0001-wrpp0250)** indique que le nom des variables du poids BRR sont **wrpp0001**, **wrpp0002**, ..., **wrpp0250**. Cette option peut aussi être appelée **brrweight(wrpp0*)** à condition qu'il n'y ait pas de variable autre que les variables du poids bootstrap dont le nom commence par « wrpp0 ».

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

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

SUDAAN

SUDAAN est un progiciel commercial expressément élaboré par le Research Triangle Institute pour l'analyse de données tirées d'enquêtes-échantillons complexes et d'autres études par observation et expérimentales comprenant des données corrélées par grappes. La version exécutable SAS du logiciel est particulièrement utile aux personnes qui connaissent le SAS. Dans la version 9.0 et les versions ultérieures, toutes les procédures indiquées dans SUDAAN peuvent utiliser la méthode BRR avec un ajustement de Fay pour estimer des variances et des covariances.

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

- préciser **DESIGN=BRR** dans la déclaration de procédure;
- inclure l'énoncé du POIDS suivant pour définir la variable du poids d'enquête : **WEIGHT wtp;**
- inclure l'énoncé REPWGT pour indiquer le nom des variables de poids bootstrap moyen sur le fichier de données et pour fournir le nombre d'échantillons bootstrap utilisés pour produire chaque variable de poids bootstrap moyen (qui sert à calculer l'ajustement de Fay). Plus particulièrement, pour les FMGD de l'ESUTC, cet énoncé REPWGT aurait la forme suivante :

REPWGT wrpp0001-wrpp0250 / ADJFAY=20;

WesVar

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

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

- déplacer les variables de poids répétées (c.-à-d. wrpp0001 à wrpp0250) dans la boîte *Replicates (Répétitions)*;
- déplacer la variable du poids d'enquête (c.-à-d. wtp) dans la boîte *Full sample (Échantillon complet)*;
- pour le bootstrap moyen, préciser la *Method (Méthode)* de Fay, puis Fay_K=.7764;
- déplacer les variables de l'analyse dans la boîte *Variables (Variables)*, un identificateur unique de la boîte d'identité (facultatif) et sauvegarder le fichier.

Phillips (2004) illustre ces directives par un exemple utilisant les données de cycle 14 de l'Enquête Sociale Générale.

11.0 Pondération

Pour le fichier de microdonnées, des poids statistiques ont été placés pour chaque enregistrement pour représenter le nombre de ménages ou de personnes que chaque enregistrement échantillonné représente. Un poids a été calculé pour chaque ménage et, dans un fichier différent, un poids pour chaque personne.

La pondération pour 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 du champ de l'enquête, 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 15 ans et plus.

11.1 Procédures de pondération pour les fichiers des ménages et des 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.

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

Il y avait 153 834 numéros de téléphone dans l'échantillon avec des poids attribués.

2. Ajustement pour les numéros de téléphone non résolus

Il y a eu 8 909 numéros de téléphone qui n'ont pas été résolus, laissant 144 925 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 signalisant 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-mois, par genre de ligne prévu, et selon si le numéro a été envoyé aux intervieweurs ou pas.

Pour chaque province-mois-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 85 541 numéros de téléphone qui étaient hors du champ de l'enquête et 59 384 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 8 742 numéros de téléphone pour lesquels le nombre de lignes manquait toujours. Donc, il y avait 50 642 ménages dont l'information a été calculée ou imputée. L'ajustement a été fait au niveau province-mois.

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

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 desservait le ménage. L'ajustement a été fait au niveau province-mois.

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

6. Ajustement pour les ménages non répondants

Les répondants du ménage ont répondu aux questions 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 807 non-répondants. Donc 49 835 ménages pondérés dans le champ d'enquête ont été utilisés et ajustés par province-mois.

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

11.2 Procédures de pondération pour le fichier des ménages

7. Ajustement pour totaux externes connus de ménages par province-mois

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

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

Les poids des ménages, W_6 , obtenus après cette étape, sont finals et apparaissent sur le fichier de microdonnées du ménage.

11.3 Procédures de pondération pour le fichier des personnes

7. Éliminer les ménages où personne n'a été sélectionné

Il y avait 29 072 ménages où 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 ont été éliminés parce qu'ils n'avaient aucune information au niveau de la personne. Environ 70 % des répondants sélectionnés âgés de 25 ans et plus n'ont pas été retenus. Il y avait 20 763 ménages où nous avons sélectionné un répondant. Il y avait 17 736 ménages avec une personne sélectionnée et 3 027 avec deux personnes sélectionnées.

8. Calcul du poids de groupe

Tous les ménages répondants dans l'enquête ayant une liste complète des membres (c.-à-d. 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 suivants : 15 à 19, 20 à 24 et 25 ans 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 d'âge 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, $W_6 = W_5 * 3/2$.

9. Calcul du poids des ménages pour les personnes sélectionnées

Les $17\,736 + 2(3\,027) = 23\,790$ personnes sélectionnées sont associées avec des ménages répondants faisant partis du champ de l'enquête 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 sélectionné 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 à 19 et 20 à 24. Le facteur de sous-échantillonnage pour le groupe d'âge 25 ans et plus avait déjà été assigné et variait de 20,8 % à 32,1 %, selon la province.

11. Ajustement pour les individus non répondants

Le fichier des personnes comprend les enregistrements des répondants individuels qui ont donné l'information aux questions sur les habitudes de fumer et qui ont fourni une date de naissance correspondant à l'âge indiqué dans la liste des membres. Il y a eu 3 249 non-répondants.

Donc, nous avons utilisé 20 541 poids des individus dans le champ d'enquête et ajusté selon la

province, le groupe d'âge tiré de la liste des membres (15 à 19, 20 à 24, 25 à 44, 45 à 64, 65 ans et plus) et le sexe.

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

12. 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 15 ans et plus. Ceci est connu comme la post-stratification. Les totaux externes suivants ont été utilisés :

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

Pour le sommaire annuel :
totaux de population selon la province, le sexe et les groupes d'âge suivants :
15 à 19, 20 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. 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

Consulter le fichier ESUTC2008_C1_QuestF.pdf et ESUTC2008_C2_QuestF.pdf pour les questionnaires français utilisés en 2008.

13.0 Clichés d'enregistrement à valeurs univariées

13.1 Cliché d'enregistrement à valeurs univariées – Fichier des ménages

Consulter le fichier ESUTC2008_Annuel_MN_LvCds.pdf pour le cliché d'enregistrement à chiffres univariés pour le fichier des ménages de 2008.

13.2 Cliché d'enregistrement à valeurs univariées – Fichier des personnes

Consulter le fichier ESUTC2008_Annuel_PR_LvCds.pdf pour le cliché d'enregistrement à chiffres univariés pour le fichier des personnes de 2008.