

ENQUÊTE SUR LES ACTIVITÉS ET LES PRÉFÉRENCES
EN MATIÈRE DE VOYAGES, 1999

GUIDE DE L'UTILISATEUR



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
4.1	Concepts et définitions de la composante de contrôle	11
4.2	Concepts et définitions de la composante du questionnaire	13
5.0	Méthodologie de l'enquête	17
5.1	Population	17
5.2	Plan de composition téléphonique aléatoire par l'élimination des banques non valides	17
5.3	Plan d'échantillonnage	18
5.4	Taille des échantillons	18
6.0	Collecte des données	19
7.0	Traitement des données	21
7.1	Saisie des données	21
7.2	Vérification et imputation	21
7.3	Codage des questions ouvertes	23
7.4	Création des variables dérivées	23
7.5	Pondération	24
7.6	Suppression de renseignements confidentiels	24
8.0	Qualité des données	25
8.1	Taux de réponse	25
8.2	Erreurs dans l'enquête	26
9.0	Lignes directrices pour les totalisations, l'analyse et la diffusion des données	29
9.1	Lignes directrices en matière d'arrondissement	29
9.2	Lignes directrices pour la pondération de l'échantillon en vue des totalisations	30
9.3	Lignes directrices pour l'analyse statistique	30
9.4	Lignes directrices pour la diffusion en fonction du coefficient de variation.	32
10.0	Tableaux des variabilités d'échantillonnage approximatives	35
10.1	Comment utiliser les tableaux des coefficient de variation pour les estimations de type nominal	37
10.2	Comment utiliser les tableaux des coefficient de variation pour calculer les limites de confiance	41
10.3	Comment utiliser les tableaux des coefficient de variation pour effectuer un test t.	43
10.4	Coefficients de variation des estimations quantitatives	44
10.5	Seuils de diffusion pour l'Enquête sur les activités et les préférences en matière de voyages	45
10.6	Tableaux des coefficient de variation	47

11.0	Pondération	49
12.0	Questionnaires et fiches de codage	53
13.0	Cliché d'enregistrement et variables uniques	55

1.0 Introduction

L'Enquête sur les activités et les préférences en matière de voyages (EAPV) a été effectuée par Statistique Canada pour le compte de la Commission canadienne du tourisme et quatre agences provinciales et territoriales de tourisme.

Il s'agit d'une enquête à participation volontaire qui a été effectuée en vertu de la *Loi sur la statistique*, Lois révisées du Canada, 1985, chapitre S19. Les plans de collecte de l'enquête sont enregistrés sous le numéro de collecte SQC/SSD-040-75154 et sous le numéro de fichier de renseignements personnels STC/P-PU-016.

Le présent document contient des renseignements sur l'accès aux données de l'enquête et leur utilisation.

Les personnes qui désirent obtenir de plus amples renseignements sur les ensembles de données ou leur utilisation doivent communiquer avec les personnes suivantes:

Statistique Canada

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

Alex Athanassakos
Ministère du Tourisme
Direction des politiques et recherches en matière de tourisme
700, rue Bay, 15^e étage
Toronto (Ontario) M7A 2E1
Téléphone : (416) 314-7317
Télécopieur : (416) 314-7341
Courriel : Alex.Athanassakos@mtr.gov.on.ca

2.0 Contexte

L'information recueillie porte sur : les endroits au Canada où les gens ont voyagé durant les deux dernières années et les endroits où les gens pensent voyager durant les deux prochaines années, les raisons pour lesquelles les non-voyageurs ne voyagent pas; la participation aux activités récréatives et de divertissement, les raisons pour lesquelles les gens voyagent au Canada, les types d'hébergement utilisés en voyage, les sources d'information utilisées pour planifier un voyage et les impressions des gens sur les différentes destinations canadiennes.

Les données de cette enquête ont été recueillies en vertu de la disposition du partage des données de la *Loi sur la statistique*. Des ententes pour le partage des données recueillies dans le cadre de cette enquête ont été conclues avec les ministères et organismes suivants :

1. La Commission canadienne du tourisme;
2. Le Ministère du développement économique, du commerce et du tourisme de l'Ontario;
3. La Saskatchewan Tourism Authority;
4. Le ministère de l'Industrie, du Commerce et du Tourisme du Manitoba; et
5. Le Yukon Government Department of Tourism.

Les données de l'Enquête sur les activités et les préférences en matière de voyages (EAPV) sont utilisées par quelques organismes et ministères fédéraux, la Commission canadienne du tourisme et des conseils provinciaux du tourisme. Parmi les autres utilisateurs des données, on trouve les médias, des entreprises, des experts-conseils, des universités et d'autres chercheurs qui s'intéressent aux voyageurs canadiens.

Les chercheurs et les experts-conseils du gouvernement, les entreprises privées, les universités et les médias utiliseront les résultats de l'enquête pour informer le public, pour élaborer de nouveaux programmes et pour déterminer les nouveaux besoins en services et infrastructures.

3.0 Objectifs

L'objectif global de l'enquête était de recueillir de l'information sur les activités et les préférences des Canadiens en matière de voyages.

Quelques-uns des objectifs supplémentaires de l'enquête en matière de données, sont de :

- recueillir des données sur les voyages avec nuitées effectués durant les deux dernières années au Canada, aux États-Unis et dans d'autres pays;
- recueillir des données sur les motifs des voyages avec nuitées au Canada, aux États-Unis et dans d'autres pays;
- recueillir des données sur les activités réalisées lors des voyages;
- recueillir des données sur les motifs pour entreprendre des voyages dans l'avenir.

4.0 Concepts et définitions

Nous décrivons dans ce chapitre les concepts et les définitions qui présentent un intérêt pour les utilisateurs. L'application de l'interview téléphonique assistée par ordinateur (ITAO) de l'Enquête sur les activités et les préférences en matière de voyages (EAPV) comportait deux composantes pour la collecte de données : la composante de contrôle et la composante du questionnaire.

La composante de contrôle servait à a) conserver un registre des appels pour chaque numéro de téléphone de l'échantillon, b) énumérer les membres des ménages de même que leur âge et leur sexe et c) choisir un répondant pour remplir le questionnaire téléphonique de l'EAPV.

Le questionnaire de l'EAPV, rempli par les répondants choisis, comportait deux composantes de contenu, l'une pour l'entrevue téléphonique, l'autre pour le questionnaire postal, si le répondant mentionnait lors de l'entrevue téléphonique, a) avoir voyagé au cours des deux dernières années ou b) avoir l'intention de voyager au cours des deux prochaines années.

Les concepts et les définitions de la composante de contrôle de l'EAPV sont décrits dans la section 4.1 alors que ceux propres à la composante du questionnaire de l'EAPV se trouvent à la section 4.2.

Consultez le chapitre 12.0 pour une copie des questions de l'enquête.

4.1 Concepts et définitions de la composante de contrôle

Logement

Un logement est un ensemble de pièces d'habitation qu'une personne ou un groupe de personnes habite ou peut habiter (p. ex., logement privé, appartement, duplex, etc.)

Ménage

Un ménage est une personne ou un groupe de personnes qui vit dans un logement.

Par exemple :

- une personne qui vit seule;
- un groupe de personnes non apparentées qui partagent le même logement;
- une famille.

Répondant du ménage

Le répondant du ménage est la personne qui donne les renseignements généraux sur les membres du ménage.

Membre du ménage

Toute personne qui, au cours de la période d'enquête :

- considère ou déclare considérer le logement en question comme son lieu de résidence habituel; ou
- habite le logement en question et n'a aucun autre lieu de résidence habituel.

Cela comprend :

- l'époux(se) ou conjoint(e) (en union libre ou de même sexe) qui habite habituellement dans le logement, mais peut se trouver ailleurs temporairement pour son travail ou ses études, par exemple un membre des Forces canadiennes peut être posté dans une autre région, mais le logement constitue son lieu de résidence habituel;
- un enfant qui vit temporairement ailleurs pour ses études ou un travail saisonnier, mais qui considère le logement comme son lieu de résidence habituel et qui y a passé au moins 30 jours au cours des 12 derniers mois;
- les enfants en garde « physique » partagée qui passent beaucoup de temps avec chaque parent (p. ex., une semaine avec un parent et une semaine avec l'autre) et ceux qui ont passé la nuit précédente dans ce logement;
- une personne qui habite temporairement dans un établissement institutionnel et qui s'est absentée de son logement depuis moins de six mois;
- une personne qui a demandé le statut de réfugié;
- un étudiant qui détient un visa d'étudiant et qui fréquente un établissement d'enseignement au Canada;
- une personne qui détient un permis de travail et qui demeure au Canada avec sa famille.

Personne choisie (répondant)

La personne choisie (répondant) est un membre du ménage, de 18 ans ou plus, choisie au hasard pour donner les renseignements du questionnaire.

Dans le champ de l'enquête

Sont considérés dans le champ de l'enquête les numéros de téléphone qui mettent en communication avec :

- des logements privés;
- des quartiers privés dans des logements collectifs.

Les logements collectifs comprennent les maisons de chambres, les résidences pour étudiants, les auberges, les résidences des groupes religieux, les camps militaires, etc.

Hors du champ de l'enquête

Sont considérés hors du champ de l'enquête les numéros de téléphone qui mettent en communication avec:

- une personne qui habite dans une institution, par exemple un centre d'hébergement, un hôpital, un établissement correctionnel, etc.;
- un logement collectif dans lequel dix personnes non apparentées ou plus utilisent le même numéro de téléphone.

4.2 Concepts et définitions de la composante du questionnaire

Définition d'un voyage avec nuitée

Tout voyage d'une nuit ou plus sans tenir compte de la distance de la demeure. Ce qui constitue un « voyage » est laissé à l'interprétation et à la perception du répondant. Par exemple, une nuit passée chez un voisin, un week-end au chalet, un séjour chez vos parents pendant que l'on repeint votre appartement, sont considérés comme des voyages si le répondant les perçoit ainsi.

Ne sont pas considérés des voyages:

- Le fait de prendre l'ambulance vers un hôpital ou une clinique;
- Les voyages effectués à titre d'opérateur ou de membre du personnel de l'équipage d'un véhicule (p. ex. chauffeur de camion, agent de bord, capitaine de bateau, etc.);
- Les déplacements effectués pour vous rendre au travail ou à l'école;
- Le déménagement d'une résidence à une autre (À noter qu'un voyage de recherche de logement avant le déménagement proprement dit est considéré comme un voyage.);
- Les voyages entrepris au moment où un répondant n'était pas un résident du Canada (les voyages doivent avoir le Canada comme point de départ);
- Les voyages qui s'étendent sur plus d'un an ou 365 jours.

Raison principale du voyage

Cette question détermine les raisons pour lesquelles le répondant a voyagé.

Si le répondant choisi a été accompagné lors de son voyage, n'inscrire que la raison principale pour laquelle le répondant choisi a fait le voyage.

Voyage d'agrément

Cette catégorie comprend les voyages effectués pour magasiner, à des fins touristiques, pour accompagner quelqu'un lors d'un voyage d'affaires ou pour assister à un événement sportif, à un concert, à un film ou à un festival de musique.

Cette catégorie comprend les voyages dont le motif principal est de magasiner ou de faire des achats qui ne sont pas à une distance plus rapprochée du domicile du répondant.

Dans certaines régions du Canada, les répondants font de courts voyages avec nuitée pour aller magasiner aux États-Unis. Il est important que tous ces voyages soient compris sous la catégorie « Voyage d'agrément ».

Visite d'amis ou de parents

Il s'agit d'un voyage dont le motif principal est de rendre visite à des amis ou des parents, même si le répondant choisi peut avoir pris part à d'autres activités lors du voyage.

Affaires

Le but principal du voyage du répondant choisi est l'exercice des fonctions professionnelles. Les voyages d'affaires comprennent les réunions, les conférences, les séminaires, la formation, les voyages d'inspection, les voyages de consultation, ou le fait d'assister à un congrès.

Autres raisons personnelles

Le motif principal du voyage est un rendez-vous chez un médecin, un avocat, un dentiste, la présence à des funérailles ou à une entrevue d'emploi, etc.

Les voyages dûs à une évacuation d'urgence sont inclus.

Un voyage lié à une activité religieuse sera inscrit en fonction du type d'événement auquel la personne a assisté. L'intervieweur doit chercher à faire préciser au répondant s'il s'agissait d'un voyage pour raisons personnelles, comme une retraite, ou d'un voyage d'agrément, comme une activité sociale organisée par l'Eglise.

Revenus

Le répondant doit déclarer toutes ses sources de revenus. Par exemple :

- ceux provenant de salaires et de traitements ou du travail autonome, y compris les pourboires, les commissions et les gratifications;
- ceux provenant du gouvernement comme les prestations d'assurance-emploi, le Régime de pensions du Canada, le Régime de rentes du Québec ou la sécurité de la vieillesse;
- ceux provenant d'intérêts, de dividendes, de placements ou de régimes de retraite particuliers
- d'autres sources telles les pensions alimentaires, les bourses d'études.

Le salaire total provenant de tous les emplois est inclus avant les déductions d'impôt sur le revenu, des cotisations à un régime de retraite, de l'assurance-médicale, etc.

Les gains provenant du jeu ou de la loterie, les sommes forfaitaires reçues en héritage pendant l'année, les remboursements d'impôt sur le revenu, les remboursements de prêts, les règlements forfaitaires d'assurance, les ristournes sur les taxes foncières, les remboursements de cotisations à un régime de retraite, les indemnités de grève et prestations de maladie versées par des syndicats, etc., sont exclus.

Niveau de scolarité

Cela a trait aux études secondaires, aux certificat, aux diplômes, aux grades ou aux autres études postsecondaires.

Études primaires et secondaires

Au Canada, l'éducation relève des gouvernements provinciaux et territoriaux. Chaque province et territoire a conçu son propre système d'éducation et les structures peuvent être différentes d'un secteur de compétence à l'autre.

Le tableau suivant illustre les similitudes et les différences en date de 1997-1998.

Province ou territoire	Nombre d'années pre-niveau 1	Primaire (par année)	Secondaire premier cycle/intermédiaire/moyen (par année)	Secondaire deuxième cycle (par année)	Secondaire deuxième cycle (par année)
Terre-Neuve et Labrador	1	1-6	7-9	10-12	
Île-du-Prince-Édouard		1-6	7-9	10-12	
Nouvelle-Écosse	1	1-6	7-9	10-12	
Nouveau-Brunswick (secteur anglais)	1	1-5	6-8	9-12	
Nouveau-Brunswick (secteur français)	1	1-8			9-12
Québec	2	1-6		Secondaire 1-5	
Ontario	2	1-8			9-12 (inclut le cours préuniversitaire de l'Ontario)
Manitoba	1	1-8			9-12
Saskatchewan	3	1-5	6-9		10-12
Alberta	1	1-6	7-9	10-12	
Colombie-Britannique	1	1-7			8-12
Yukon	1	1-7			8-12
Territoires du Nord Ouest ou Nunavut	1	1-6	7-9	10-12	

Source: Indicateurs de l'éducation au Canada, 1998-1999 n° 81-582-XPF au catalogue.

Certificat ou diplôme de formation professionnelle

Ce terme est utilisé pour le classement des cours de compétence qui préparent les stagiaires à des métiers qui ne sont pas professionnels ou semi-professionnels. Une école de formation professionnelle est un établissement d'enseignement public qui offre des cours préparant à l'emploi dans un métier précis, comme la conduite de matériel lourd, la mécanique automobile et le rembourrage. De nombreux collèges communautaires ou instituts techniques offrent des certificats ou des diplômes de métier.

Certificat ou diplôme de collège communautaire, de collège technique (collège d'enseignement général et professionnel (CEGEP) au Québec)

Il faut habituellement une ou deux années d'études pour obtenir un certificat ou un diplôme de collège communautaire ou de CEGEP.

NOTE : 1. Le Ryerson Polytechnical Institute et de nombreux collèges de la Colombie-Britannique décernent des diplômes et des certificats de niveau universitaire et collégial.

2. Les équivalences des anciens systèmes d'éducation sont les suivantes :

Collège biblique ou séminaire = baccalauréat (prêtrise) en théologie

Soins infirmiers = diplôme de collège communautaire

Autres études

Cela inclut :

- les diplômes, certificats ou permis obtenus par l'entremise d'associations professionnelles dans des domaines comme la comptabilité, le secteur bancaire ou l'assurance;
- les certificats non professionnels en santé (RCR, premiers soins, Croix-Rouge, etc.);
- l'éducation permanente, les cours d'intérêt personnel.

5.0 Méthodologie de l'enquête

L'enquête téléphonique a été réalisée du mois de mi-octobre à la mi-décembre 1999 avec la méthode de composition téléphonique aléatoire (CA). L'enquête postale, de novembre 1999 à mars 2000, a été effectuée avec les adresses obtenues au cours de l'enquête téléphonique pour communiquer avec les répondants.

Parce que l'enquête postale est complémentaire à l'enquête téléphonique, leurs conceptions sont semblables. Dans le document qui suit, nous débutons chaque sous-section en parlant de l'enquête téléphonique, puis nous enchaînons avec ce qui change dans l'enquête postale.

5.1 Population

La population cible de l'enquête téléphonique était toutes les personnes de 18 ans et plus dans chacune des dix provinces canadiennes, à l'exclusion des résidents à temps plein en établissement. Avec la CA, les ménages sans téléphone ont également été exclus. Toutefois, les personnes qui vivent dans de tels ménages représentent moins de 2 % de la population cible. Les estimations de l'enquête ont donc été pondérées afin de représenter les personnes sans téléphone.

La population cible de l'enquête postale était la même que celle de l'enquête téléphonique sauf que les non-voyageurs étaient exclus dans ce cas-ci. Les voyageurs sont les répondants qui ont affirmé, au cours de l'enquête téléphonique, avoir voyagé au cours des deux dernières années ou qui envisageaient fort probablement ou probablement voyager au cours des deux prochaines années.

5.2 Plan de composition téléphonique aléatoire par l'élimination des banques non valides

La composition téléphonique aléatoire est une méthode pour obtenir des données de répondants en appelant les ménages de façon aléatoire. La méthode que nous utilisons engage la notion de banques. Prenons par exemple le numéro de téléphone (613) 951-9999 avec l'indicatif régional. Les six premiers chiffres forment le préfixe et les huit premiers chiffres (en italique) forment la banque. Les deux sources d'information de numéros de téléphone que nous avons sont les fichiers de facturation que nous vendent les sociétés de téléphone, et la version électronique de tous les numéros de téléphone qui apparaissent dans les annuaires téléphoniques papier de toutes les provinces et territoires. Pour les numéros confidentiels, le fichier de facturation nous donne la banque en excluant les deux derniers chiffres pour préserver la confidentialité. Les numéros de téléphones cellulaires ont été évités dans ces listes parce que les préfixes qu'accordent les sociétés de téléphone aux téléphones cellulaires ne sont habituellement réservés qu'à cet usage. Nous utilisons ces deux sources pour déterminer les banques non valides, à la base de notre cadre de CA. Cette méthode est appelée élimination des banques non valides. On a assigné à chaque banque une strate dans sa province.

Un échantillon aléatoire de numéros de téléphone a été obtenu dans chaque strate (à partir des banques valides). Nous avons tenté de prédire le pourcentage de numéros composés qui rejoindraient les ménages (ce qu'on appelle « taux de succès »). Le taux de succès a été estimé à partir d'information d'enquête de CA antérieures, particulièrement le cycle le plus récent d'enquêtes sociales générales. Pour l'Enquête sur les activités et les préférences en matière de voyages (EAPV), nous avons anticipé un taux de succès de 43 %. Le taux de succès réel a été de 49 %.

Une fois établie la liste complète des membres du ménage, nous avons identifié l'échantillon des membres du ménage admissibles pour l'enquête, puis nous en avons choisi un au hasard pour l'enquête téléphonique. Lorsque la personne choisie était un voyageur (une voyageuse), nous lui demandions de remplir le questionnaire postal.

5.3 Plan d'échantillonnage

Le plan d'échantillonnage pour l'enquête téléphonique est un échantillonnage aléatoire simple stratifié de numéros de téléphones. En général, la strate représente les régions métropolitaines de recensement (RMR) de la province et les régions hors RMR. Nous avons effectué cette stratification en régions infraprovinciales pour être en mesure de dissocier les estimations finales par strate.

Le plan d'échantillonnage de l'enquête postale est un plan en deux phases, où la première phase est l'enquête téléphonique. Nous avons utilisé des données de l'enquête téléphonique pour définir les répondants de la deuxième phase comme voyageurs. Nous avons recensé tous les répondants de la deuxième phase.

Le plan d'échantillonnage est complexe parce que notre population cible est composée d'adultes, mais notre unité d'échantillonnage se compose de numéros de téléphone. C'est au moment de la pondération que nous avons calculé la sélection probabiliste initiale des gens de l'échantillon, laquelle est fondée sur la sélection probabiliste téléphonique initiale, le taux de succès, le nombre de téléphones par ménages et le nombre d'adultes par ménage.

5.4 Taille des échantillons

L'échantillon était constitué de 60 000 numéros de téléphone à l'échelle nationale. Dès que la communication était établie avec un ménage, la liste de tous les membres du ménage était dressée et les renseignements démographiques de base (âge et sexe) sur chacun d'eux étaient recueillis. De cette liste, nous avons choisi au hasard une personne de 18 ans ou plus pour l'enquête téléphonique. Au total, 29 123 personnes étaient admissibles à l'enquête téléphonique et 18 385 l'ont faite.

Nous avons demandé aux voyageurs qui ont fait l'enquête d'accepter et de remplir un questionnaire postal. Sur un échantillon total de 24 769 voyageurs éligibles, nous avons obtenu 5 740 questionnaires postaux remplis.

Le tableau de la section 8.1 donne plus de détails sur la répartition de la taille des échantillons et des taux de réponse par strate.

6.0 Collecte des données

L'enquête était un échantillonnage de composition téléphonique aléatoire (CA) de 18 385 Canadiennes et Canadiens et était divisée en deux phases : une enquête téléphonique pour retracer les voyageurs et les non-voyageurs, et une enquête postale qui devait être remplie par tous les voyageurs.

L'enquête téléphonique a été effectuée grâce à une application d'interview téléphonique assistée par ordinateur (ITAO) appelée Système d'exécution d'enquêtes assistées par ordinateur (CASES). La collecte des données de l'enquête téléphonique s'est déroulée sur une période d'environ dix semaines, de la mi-octobre à la mi-décembre 1999. L'ITAO a été faite par les bureaux régionaux de Statistique Canada de Montréal, Sturgeon Falls, Winnipeg et Vancouver. Dans la plupart des bureaux, les appels ont été effectués en deux quarts de travail, de 9 h à 21 h.

Tous les membres du personnel des bureaux régionaux qui ont pris part à l'enquête (c'est-à-dire les chargés de projet, les intervieweurs principaux et les intervieweurs) ont reçu une formation d'une journée à une journée et demie afin de les familiariser avec le but et les concepts de l'enquête, le questionnaire et les procédures de l'ITAO, ainsi que quelques techniques de base d'entrevue téléphonique. Un manuel du contenu et de l'intervieweur de l'Enquête sur les activités et les préférences en matière de voyages (EAPV) et des exemplaires imprimés des questionnaires ont été remis aux intervieweurs comme document de soutien lors de la collecte de données. La plupart des intervieweurs avaient déjà de l'expérience en entrevue téléphonique.

Les répondants ont été abordés dans la langue officielle principale de leur choix. On a fait appel, au besoin, à des intervieweurs bilingues.

La participation à l'enquête était volontaire. Si un répondant refusait de donner une partie ou la totalité de l'information demandée, le superviseur de l'intervieweur devait alors appeler le répondant une deuxième fois pour tenter d'obtenir l'information. Lorsque le répondant était temporairement absent ou que des difficultés d'ordre linguistique ou autre empêchaient l'enquête, l'intervieweur devait rappeler le répondant à une date ultérieure. Les réponses par procuration n'étaient pas admises.

Après avoir tout tenté pour interviewer un répondant choisi, l'intervieweur inscrivait le code d'état final de l'enquête sur le questionnaire de l'ITAO, en utilisant un des codes d'état finaux. Voir le tableau du section 8.1 pour plus de détails.

La collecte de données de l'enquête postale s'est étendue sur environ cinq mois, de novembre 1999 à mars 2000. L'enquête postale était également administrée par les bureaux régionaux de Statistique Canada de Montréal, Sturgeon Falls, Winnipeg et Vancouver. Dans la plupart des bureaux, un suivi téléphonique a été effectué en janvier et février 2000, de 9 h à 21 h.

7.0 Traitement des données

7.1 Saisie des données

Le Système d'exécution d'enquêtes assistées par ordinateur (CASES) a deux composantes principales, la gestion de cas et une autre qui est propre à chaque enquête.

Le système de gestion des cas contrôle l'affectation des cas et la transmission des données dans le cadre de l'enquête. Pour l'Enquête sur les activités et les préférences en matière de voyages (EAPV), un cas désignait un répondant choisi dans l'échantillon de composition téléphonique aléatoire (CA). Le système de gestion des cas enregistre automatiquement l'information relative à la « gestion de l'enquête » pour chaque communication (ou tentative de communication) auprès des répondants et il produit des rapports sur l'avancement du processus de collecte tout au long de l'entrevue.

La composante du CASES propre à l'enquête comprend une composante d'introduction avec les procédures pour communiquer auprès des répondants choisis pour l'entrevue. Une fois la communication établie, le CASES obtient les composantes du questionnaire pour l'entrevue avec le répondant. L'intervieweur pose les questions au répondant, puis entre les réponses dans l'ordinateur au fur et à mesure. Le programme de l'interview téléphonique assistée par ordinateur (ITAO) effectue également des vérifications en direct, qui permettent de corriger les erreurs potentielles de données pendant que le répondant est en ligne.

Une fois que les cas sont remplis, les fichiers de données contenant les enregistrements de toutes les données saisies sont transmis par voie électronique à Ottawa. Cela a permis la création d'un fichier de données brutes pour le traitement post-collecte de l'enquête.

Tous les questionnaires d'enquête postale remplis par les répondants ont été renvoyés aux quatre bureaux régionaux. Les questionnaires étaient ensuite groupés puis expédiés à la Division des opérations et de l'intégration à l'Administration centrale à Ottawa pour la saisie des données.

La saisie des données des questionnaires postaux a été faite par ordinateur. Au cours de ce processus, tout document contenant au moins un CASEID était saisi. Une version non vérifiée de tous les enregistrements a été envoyée électroniquement à la Division des enquêtes spéciales pour le traitement après collecte de l'enquête.

7.2 Vérification et imputation

Les données brutes obtenues à l'aide du CASES pour l'EAPV étaient dans l'ensemble plus épurées que celles recueillies auparavant lors d'entrevues avec papier et crayon. Néanmoins, il a fallu malgré tout procéder à une certaine vérification post-collecte, pour s'assurer que les utilisateurs du fichier de microdonnées final puissent clairement faire la distinction entre les réponses valides, la non-réponse et les sauts valides dans les questions de l'enquête.

Dans le cadre de la stratégie globale de vérification, une série de codes pouvant être appliqués à chaque question pour décrire les divers types de « non-réponse », a été élaborée. La stratégie de codage présentement en vigueur pour bon nombre d'enquêtes, y compris l'EAPV, se définit comme suit :

- 6, 96, 996, etc. pour les sauts valides (c.-à-d. la question ne s'applique pas au répondant)
- 7, 97, 997, etc. pour les réponses «ne sais pas» à une question
- 8, 98, 998, etc. pour les réponses «refus» à une question
- 9, 99, 999, etc. pour non déclaré (c.-à-d. le répondant peut appartenir ou non à la population cible de la question; se produit habituellement si la personne a répondu «ne sais pas» ou «refus» à une des questions définissant la population cible, ou si le répondant a abandonné avant la fin de l'enquête)

La préparation préliminaire du fichier de réponse ITAO/CASES comprenait le recodage de la valeur des réponses aux questions de type nominal : remplacer « x » par « 7 » (pour les réponses « ne sais pas »), « r » par « 8 » (pour « refus ») et les blancs par des codes de sauts valides (6, 96, 996, etc.), conformément à la stratégie de codage susmentionnée.

En outre, toutes les questions à choix multiples ont été converties en série de questions oui ou non.

Des parties de texte ont été supprimé là où des questions ouvertes ont été posées pour être remplacé par des indicateurs à un caractère. Un fichier d'entrée « autre, précisez » a été créé en prévision des étapes de traitement ultérieures lors du codage. Pour plus de détails, consultez la section 7.3.

Comme l'enchaînement des questions était déjà programmé dans l'ITAO, une réponse a été obtenue à chaque question figurant dans l'enchaînement, lors des entrevues complétées. Il était par conséquent impossible de sauter une question qui s'appliquait au répondant. De plus, le programme de l'ITAO vérifiait également chaque réponse saisie pour s'assurer qu'elle se situait à l'intérieur de la gamme valide établie pour la question. La vérification déterministe se fondait sur ces hypothèses : si un répondant était censé répondre à une question, alors toute réponse obtenue était valide – aucune vérification n'était faite. Si la question ne s'appliquait pas au répondant, alors un code de « saut valide » était attribué. S'il y avait des doutes à savoir si une question s'appliquait ou non à un répondant, le code « non déclaré » était attribué.

Comme dans le cas du traitement des enquêtes de CA précédentes, l'imputation ne convenait pas pour la plupart des questions, de sorte que le code «non déclaré» a généralement été attribué aux données manquantes. Les enregistrements contenant des données jugées insuffisantes ou incompatibles ont été supprimés du fichier.

De plus, tous les enregistrements de non-réponse ont été retirés du fichier de même que les enregistrements des répondants qui ont refusé de partager leurs réponses avec d'autres ministères gouvernementaux.

On a utilisé, dans l'enquête postale, la même série de codes que dans l'enquête téléphonique, pour définir les réponses «non déclaré» et «saut valide», etc. (tel que défini au début de cette section). Toutefois, il fallait aborder une approche générale différente pour la vérification. Les réponses (ou le manque de réponses) des questions 1 à 12 étaient vérifiées pour déterminer les réponses de certaines questions clés (Q1a, Q1b, Q2a, Q2b, Q3a et Q8a). Pour le reste des questions, une combinaison des méthodes de vérification descendante (de haut en bas) et ascendante (de bas en haut) a été utilisée. Deux erreurs principales dans l'enchaînement du questionnaire ont été corrigées.

Les premières erreurs étaient celles où on trouvait des réponses aux questions qui ne s'appliquaient pas aux répondants et qui n'auraient donc pas dû être posées. Dans ces cas, l'ordinateur corrigeait automatiquement l'erreur en éliminant les données superflues et en attribuant le code de « saut valide ».

Les secondes erreurs étaient celles où on trouvait un blanc aux questions qui auraient dû être répondues. Ces questions ont reçu le code «non déclaré».

De plus, certaines vérifications ont été faites pour assurer la cohérence entre les questions. Par contre, aucune vérification de cohérence n'a été faite entre les enquêtes téléphonique et postale et, contrairement à l'enquête téléphonique, les enregistrements des répondants ayant refusé le partage des données ont été prélevés du fichier de microdonnées.

7.3 Codage des questions ouvertes

Un certain nombre d'éléments de données de chaque questionnaire téléphonique ont été enregistrés sous forme de question ouverte. Certaines de ces entrées «Autre, précisez» ont été par la suite recodées.

Chaque entrée de texte a été manuellement vérifiée et pour chacune, on a décidé si elle appartenait ou non à l'une des catégories existantes de la question. Si on déterminait que oui, la réponse «Autre, précisez» était remplacée par «Non» et le texte retiré. La catégorie où la réponse était déplacée était remplacée par «Oui». Pour la question Q36, de nouvelles catégories ont été créées suite au recodage de certaines réponses de texte.

Il n'y avait pas de questions ouvertes dans l'enquête postale.

7.4 Création des variables dérivées

Un certain nombre d'éléments de données du fichier de microdonnées téléphoniques ont été dérivées par la combinaison d'éléments semblables du questionnaire pour créer des variables comme le nombre de personnes dans le ménage, le nombre de personnes dans la famille, le nombre de membres de moins de 18 ans dans le ménage, le nombre de membres de moins de 18 ans dans la famille, etc. Des descriptions plus complètes de chaque variable dérivée se trouvent au chapitre Cliché d'enregistrement.

Bon nombre de variables dérivées du fichier téléphonique ont été supprimées pour des raisons de confidentialité. Consultez le section 7.6 pour plus de renseignements. Il est possible d'obtenir des totalisations spéciales fondées sur le fichier principal.

Il n'y a pas de variables dérivées en tant que tel dans le fichier de microdonnées postales. Toutefois, certaines variables comme la province, la taille du ménage, etc. ont été prises dans l'enquête téléphonique pour faciliter l'utilisation du fichier de microdonnées postales.

7.5 Pondération

Le principe sur lequel repose l'estimation pour un échantillon probabiliste veut que chaque personne dans l'échantillon représente, en plus d'elle-même, plusieurs autres personnes qui ne font pas partie de l'échantillon. Par exemple, avec un échantillon aléatoire simple correspondant à 2 % de la population, chaque personne de l'échantillon représente 50 personnes dans la population. La phase de la pondération est l'étape où l'on calcule, pour chaque enregistrement, ce nombre. Ce poids de l'enquête figure dans le fichier de microdonnées et doit être utilisé pour produire des estimations significatives à partir des résultats de l'enquête. Par exemple, si l'on veut estimer le nombre de personnes qui ont voyagé au Canada au cours des deux dernières années, il faut sélectionner les enregistrements qui font référence aux personnes de l'échantillon qui présentent cette caractéristique et additionner les poids liés à ces enregistrements.

Puisque les enquêtes téléphonique et postale ont des populations cibles et des taux de réponse différents, chaque fichier de données fait l'objet d'une pondération séparée. Des précisions sur la méthode utilisée pour calculer ces poids se trouvent au chapitre 11.0.

7.6 Suppression de renseignements confidentiels

Il convient de noter que le fichier de microdonnées « à grande diffusion » décrit dans ce document diffère à bien des égards du fichier « principal » de l'enquête conservé par Statistique Canada. Ces différences sont le résultat de mesures prises pour protéger l'anonymat des répondants de l'enquête. Par exemple, les variables contenant des valeurs extrêmes ont été limitées (p.ex. le nombre de personnes d'un ménage), l'information de certaines variables a été ajoutée à des catégories plus générales (p.ex. âge) et certaines variables choisies de répondants particuliers ont été supprimées dans certains cas. Les utilisateurs qui désirent accéder à l'information exclue des fichiers de microdonnées peuvent acheter des totalisations personnalisées. Les estimations obtenues seront communiquées à l'utilisateur, à la condition qu'elles soient conformes aux lignes directrices pour l'analyse et la diffusion décrites au chapitre 9.0 de ce document.

8.0 Qualité des données

8.1 Taux de réponse

Le tableau qui suit présente un résumé du nombre de répondants choisis pour les enquêtes téléphonique et postale dans chaque strate infraprovinciale, de même que les taux de réponse atteints. Le taux de succès est le pourcentage de numéros de téléphones obtenus au hasard qui se sont révélés être des numéros résidentiels (au lieu de numéros commerciaux ou hors service). Le taux de réponse téléphonique est le pourcentage d'enquêtes téléphoniques complétées sur les numéros de téléphone résidentiels. Le taux de réponse postal est fondé sur une estimation du nombre de voyageurs qui ont un numéro de téléphone résidentiel. Par conséquent, la non-réponse de l'enquête téléphonique a eu un effet sur le taux de réponse de l'enquête postale.

Taux de réponse par strate pour les fichiers microdonnées à grande diffusion

Strate	Numéros de téléphone obtenus	Taux de succès (%)	Taux de réponse téléphonique (%)	Nombre de réponses téléphoniques	Taux de réponse postal (%)	Nombre de réponses postales
RMR de St. John's	1 662	42,2	64,8	455	23,9	141
RMR de Halifax	2 080	42,0	61,3	536	23,7	182
Atlantique Autre	2 225	40,2	62,3	557	23,5	163
RMR de Montréal	2 361	56,2	56,5	750	19,6	217
RMR de Québec	1 614	56,5	64,3	586	27,6	197
Québec Autre (excluant Hull)	1 458	51,2	67,1	501	26,2	140
RMR d'Ottawa-Hull	3 460	49,1	60,2	1 023	25,1	364
RMR de Toronto	5 787	44,6	67,1	1 730	21,0	444
RMR de Hamilton	2 872	50,5	69,2	1 004	26,4	313
RMR de London	2 318	51,0	52,4	619	21,9	227
RMR de Kitchener	2 390	55,9	52,4	700	21,0	246
RMR de St Catharines-Niagara	2 336	50,9	71,2	846	26,8	252
RMR de Windsor	2 211	57,2	48,3	610	16,0	178
RMR d'Oshawa	2 184	58,2	72,2	917	26,9	289
RMR de Sudbury	1 969	41,4	60,4	492	22,1	155
RMR de Thunder Bay	1 844	54,1	55,5	554	18,7	163
Ontario Autre	3 497	45,0	53,5	842	19,6	263
RMR de Winnipeg	1 944	47,6	73,4	679	26,4	199
Manitoba Autre	1 647	38,6	78,4	498	33,8	175
RMR de Saskatoon	1 196	51,8	73,0	452	27,4	156
RMR de Regina	1 316	44,4	69,2	404	24,9	129
Saskatchewan Autre	1 904	38,4	72,1	527	24,3	152
RMR de Edmonton	1 589	54,2	65,3	562	23,7	184
RMR de Calgary	1 692	51,4	63,6	553	21,7	175
Alberta Autre	1 380	43,8	71,9	435	26,4	144
RMR de Vancouver	2 036	51,4	60,4	632	20,2	188
RMR de Victoria	1 543	56,3	64,3	559	23,1	183
Colombie-Britannique Autre	1 166	48,3	64,3	362	23,2	121
Canada	59 681	48,8	63,1	18 385	23,2	5 740

8.2 Erreurs dans l'enquête

La qualité des données peut se diviser en deux aspects : les erreurs dues à l'échantillonnage et les erreurs non dues à l'échantillonnage. L'erreur d'échantillonnage survient lorsqu'on prend un échantillon au lieu d'un recensement et peut être quantifiée avec des outils statistiques normalisés. Les tableaux de variabilité d'échantillonnage qui se trouvent au chapitre 10.0 constituent une façon d'évaluer l'erreur d'échantillonnage.

Des erreurs qui ne sont pas liées à l'échantillonnage peuvent se produire à presque chaque phase d'une enquête. L'intervieweur peut mal comprendre les directives, le répondant peut se tromper en répondant aux questions, des réponses peuvent être mal inscrites sur le questionnaire et des erreurs peuvent être apportées au moment du traitement et de la totalisation des données. On a consacré beaucoup de temps et d'efforts en vue de réduire les erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été mises en oeuvre à chaque étape du cycle de collecte et de traitement des données, pour en contrôler la qualité. Ces mesures comprenaient la mise à l'essai du questionnaire auprès d'un groupe cible, la mise à l'essai des processus au moyen d'une enquête pilote, l'emploi d'intervieweurs hautement qualifiés, une formation poussée des intervieweurs portant sur les procédures d'enquête et le questionnaire, ainsi que des contrôles du codage et de la vérification pour assurer la logique du traitement.

Sur un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations établies à partir de l'enquête. Toutefois, les erreurs qui se produisent systématiquement introduiront des biais dans les estimations de l'enquête.

L'effet de la non-réponse sur les résultats de l'enquête est une source importante d'erreurs non dues à l'échantillonnage dans les enquêtes. Puisque la non-réponse est beaucoup plus élevée dans l'enquête postale que dans l'enquête téléphonique, nous suggérons aux utilisateurs, s'ils ont le choix, de s'appuyer davantage sur les estimations de l'enquête téléphonique.

L'importance de la non-réponse varie d'une non-réponse partielle (défaut de répondre à une seule ou à certaines questions) à une non-réponse totale. Il y a non-réponse totale quand l'intervieweur n'a pu communiquer avec le répondant, qu'aucun membre du ménage n'a pu donner l'information ou que le répondant a refusé de prendre part à l'enquête. On a traité les cas de non-réponse totale en corrigeant le poids des ménages qui ont répondu à l'enquête afin de compenser pour ceux qui n'y ont pas répondu.

Dans la plupart des cas, il y a non-réponse partielle à l'enquête lorsque le répondant n'a pas compris ou a mal interprété une question, qu'il a refusé de répondre à une question ou qu'il ne pouvait se rappeler l'information demandée.

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

La mesure de la valeur potentielle de l'erreur d'échantillonnage est fondée sur l'erreur-type des estimations découlant des résultats de l'enquête. L'erreur-type d'une estimation est habituellement exprimée en fonction de l'estimation à laquelle elle se rapporte. La mesure ainsi obtenue, appelée coefficient de variation (c.v.) d'une estimation, s'obtient en divisant l'erreur-type de l'estimation par l'estimation elle-même, et s'exprime en pourcentage de l'estimation.

Par exemple, supposons qu'à partir des résultats de l'enquête, on estime que 8,3 % des gens en Ontario ont voyagé en Alberta au cours des deux dernières années et qu'on établisse à 0,3 % l'erreur-type de cette estimation. Le coefficient de variation de l'estimation serait alors calculé comme suit :

$$c.v. = \left(\frac{0,003}{0,083} \right) \times 100 \% = 3,6 \%$$

Dans le cas de l'Enquête sur les activités et les préférences en matière de voyages (EAPV), nous pouvons constater une erreur non due à l'échantillonnage en se servant de la même question dans les questionnaires téléphonique et postal. Dans l'enquête téléphonique, elle était formulée ainsi :

Q5. Vous avez indiqué avoir fait au moins un voyage d'agrément au Canada au cours des deux dernières années. Dans quels provinces ou territoires avez-vous séjourné (province traversée ou province de destination) pour une nuitée ou plus durant votre (vos) voyage(s) d'agrément?

(10) Terre-Neuve et Labrador

⋮

(35) Ontario

Dans l'enquête postale, on a posé cette question aux répondants :

Q16B. Avez-vous fait au cours des 2 dernières années un voyage d'agrément d'une nuitée ou plus à un endroit quelconque du Nord ou du Sud de l'Ontario?

Le tableau qui suit montre les réponses à ces questions et les différences absolues entre les pourcentages estimés des deux enquêtes. Il est étonnant d'observer que le pourcentage estimé pour l'enquête postale est constamment plus élevé. Ce phénomène ne peut pas être attribué aux différences qu'il pourrait y avoir dans la définition des voyageurs. Nous croyons qu'il peut s'agir d'une erreur de mémoire des répondants. L'enquête téléphonique a été effectuée tout juste avant Noël et l'enquête postale a été effectuée surtout après Noël. Selon toute vraisemblance, les répondants ont tendance à mieux se rappeler les voyages faits au cours des quelques derniers mois plutôt que ceux faits auparavant, d'où le pourcentage constamment plus élevé pour l'enquête postale.

En raison de cette comparaison, l'utilisateur des données doit savoir que la réponse reçue ne répond pas nécessairement à la question posée. Bref, l'information dérivée de ces deux questions n'est pas très fiable. Il s'agit d'un cas où l'erreur d'échantillonnage est déformée par l'erreur non due à l'échantillonnage, en raison des erreurs de mémoire des répondants. Les utilisateurs sont donc avertis que la qualité des données ne peut être complètement résumée par la seule utilisation des estimations de variabilité d'échantillonnage.

Pourcentage estimé de voyages en Ontario au cours des deux dernières années parmi tous les voyageurs

Strate	Estimation de l'enquête téléphonique (%)	Estimation de l'enquête postale (%)	Différence (%)
RMR de Toronto	43,8	74,0	30,2
RMR de Thunder Bay	52,4	82,3	29,9
RMR de Hamilton	55,4	83,6	28,2
RMR d'Oshawa	55,9	81,1	25,2
RMR de St Catharines-Niagara	58,1	83,1	25,0
RMR de Winnipeg	32,3	56,8	24,5
RMR de Kitchener	61,4	83,8	22,4
RMR de London	64,2	84,9	20,7
RMR d'Ottawa	59,3	79,6	20,3
Ontario Autre	67,1	87,3	20,2
RMR de Sudbury	76,4	90,0	13,6
Atlantique Autre	26,3	39,8	13,5
RMR de Windsor	65,9	79,3	13,4
Saskatchewan Autre	9,1	22,4	13,3
Alberta Autre	10,8	23,1	12,3
RMR de Hull	31,3	43,0	11,7
RMR de Montréal	28,8	40,3	11,5
Manitoba Autre	20,8	32,3	11,5
Colombie-Britannique Autre	11,8	22,3	10,5
RMR de Halifax	31,6	40,4	8,8
RMR de Victoria	18,1	26,7	8,6
RMR de Calgary	17,5	24,9	7,4
RMR de Regina	16,6	23,2	6,6
Québec Autre (exclant Hull)	19,2	25,7	6,5
RMR de Saskatoon	13,9	19,8	5,9
RMR de Québec	26,2	31,6	5,4
RMR de St. John's	27,2	32,4	5,2
RMR de Edmonton	13,6	18,3	4,7
RMR de Vancouver	17,5	22,1	4,6
Canada	34,1	49,6	15,5

9.0 Lignes directrices pour les totalisations, l'analyse et la diffusion des données

Nous énonçons dans cette section les lignes directrices auxquelles doivent se conformer les utilisateurs qui font des totalisations et des analyses ou qui publient ou diffusent des données tirées des fichiers de microdonnées de l'enquête.

9.1 Lignes directrices en matière d'arrondissement

Afin que les estimations destinées à la publication ou à la diffusion et établies à partir des fichiers de microdonnées correspondent à celles qui sont établies par Statistique Canada, il est vivement recommandé aux utilisateurs de se conformer aux lignes directrices suivantes en matière d'arrondissement :

- a) Les estimations dans le corps principal d'un tableau statistique doivent être arrondies à la centaine près, selon la technique normale d'arrondissement qui prévoit ce qui suit : si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, alors le dernier chiffre à conserver ne change pas. Si toutefois le premier ou le seul chiffre à supprimer est compris entre 5 et 9, le dernier chiffre à conserver est alors élevé de 1. Prenons l'exemple suivant : pour l'arrondissement normal à la centaine près, si les deux derniers chiffres sont compris entre 00 et 49, ils sont remplacés par 00 et le chiffre qui précède (le chiffre de la centaine) est conservé tel quel. Si les deux derniers chiffres sont compris entre 50 et 99, ils sont remplacés par 00 et le chiffre qui précède est élevé de 1.
- b) Dans les tableaux statistiques, les sous-totaux et les totaux marginaux doivent être calculés à partir des chiffres bruts correspondants, puis être arrondis à la centaine près, selon la méthode normale d'arrondissement.
- c) Les moyennes, proportions, taux et pourcentages doivent être calculés à partir des composantes non arrondies (c'est-à-dire numérateurs et/ou dénominateurs), puis être arrondis à une seule décimale selon la méthode normale d'arrondissement. Selon cette méthode normale d'arrondissement à une décimale, la décimale retenue reste inchangée lorsque le dernier chiffre à éliminer se situe entre 0 et 4; par contre, la dernière décimale à retenir s'accroît d'une unité lorsque le dernier chiffre à éliminer se situe entre 5 et 9.
- d) Les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs composantes respectives non arrondies, puis être elles-mêmes arrondies à la centaine près (ou à la décimale unique), par la méthode normale d'arrondissement.
- e) Lorsque, pour des raisons d'ordre technique ou autre, il faut utiliser une technique d'arrondissement autre que la méthode normale, et que ceci donne lieu à des estimations à publier ou à diffuser qui diffèrent des estimations correspondantes publiées par Statistique Canada, on recommande vivement aux utilisateurs d'indiquer la raison justifiant le choix de cette autre méthode dans la publication ou le document diffusé.
- f) Les utilisateurs ne doivent en aucun cas publier ou autrement diffuser des estimations brutes. Les estimations non arrondies peuvent en effet laisser croire à un degré de précision supérieur à ce qu'il est en réalité.

9.3 Lignes directrices pour la pondération de l'échantillon en vue des totalisations

Le plan d'échantillonnage utilisé pour l'Enquête sur les activités et les motivations en matière de voyages (EAPV) n'était pas autopondéré. Pour produire des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent employer le poids d'échantillonnage approprié. Faute d'utiliser les poids voulus, les estimations calculées à partir des fichiers de microdonnées ne peuvent pas être considérées comme représentatives de la population observée et elles ne correspondront pas à celles de Statistique Canada.

Les utilisateurs doivent également prendre en note que certains progiciels produisent des estimations qui ne correspondent pas exactement à celles de Statistique Canada, à cause du traitement du champ de pondération.

Avant de décrire comment on peut totaliser et analyser les données de l'EAPV, il serait utile de décrire les deux principaux types d'estimations ponctuelles sur les caractéristiques de la population qui peuvent être produites à partir du fichier de microdonnées de cette enquête.

Les estimations de type nominal sont des estimations du nombre ou du pourcentage de personnes dans la population visée par l'enquête qui possèdent certaines caractéristiques ou qui font partie d'une catégorie définie. Le nombre de personnes du Manitoba qui ont voyagé en Ontario durant les deux dernières années ou la proportion de Manitobains qui ont voyagé en Ontario sont des exemples d'estimations de type nominal. On peut aussi dire que l'estimation du nombre de personnes qui possèdent une certaine caractéristique est l'estimation d'un agrégat. La grande majorité des questions de l'EAPV étaient catégoriques et, lorsqu'elles ne l'étaient pas, elles ont été regroupées pour devenir des variables catégoriques.

On peut obtenir une estimation du nombre de personnes possédant une caractéristique particulière (p.ex. qui ont voyagé durant les deux dernières années) à partir du fichier de microdonnées, en additionnant les poids finals de tous les enregistrements possédant la ou les caractéristiques qui nous intéressent. La façon de calculer les proportions et les ratios sous la forme X/Y est présentée au chapitre 10.0.

9.3 Lignes directrices pour l'analyse statistique

L'Enquête sur les activités et les préférences en matière de voyages est fondée sur un plan d'échantillonnage complexe, avec stratification et probabilités inégales de sélection des répondants. L'utilisation des données provenant d'enquêtes aussi complexes pose des difficultés pour les analystes parce que le plan d'enquête et les probabilités de sélection influencent les méthodes d'estimation et de calcul de la variance qui doivent être utilisées. Afin que les estimations et les analyses établies à partir de l'enquête soient exemptes de biais, il faut utiliser les facteurs de pondération.

Bien que bon nombre de méthodes d'analyse qui font partie des progiciels statistiques permettent l'utilisation des poids, la définition ou la signification du poids, dans ces méthodes, est différente de celles qui conviennent pour une enquête par sondage, de sorte que, même si les estimations faites au moyen de ces progiciels sont exactes dans bien des cas, les variances ainsi calculées n'ont pratiquement aucune signification. Les variances pour les estimations simples comme les totaux, les proportions et les ratios (pour les variables qualitatives) sont indiquées dans les tableaux des variabilités d'échantillonnage ci-joints.

Il existe, pour d'autres techniques d'analyse (par exemple la régression linéaire, la régression logistique et l'analyse de la variance), une méthode qui rend plus significative la variance calculée par les logiciels standards et qui consiste à incorporer les probabilités inégales de sélection. Cette méthode a pour effet de transformer les poids de manière à ce que la valeur du poids moyen soit égale à un (1).

Supposons par exemple que l'on veuille faire une analyse de l'ensemble des répondants de sexe masculin. La marche à suivre pour le rééchantillonnage des poids est la suivante :

- a) sélectionner tous les répondants dans le fichier qui ont déclaré SEXE = homme;
- b) calculer le poids MOYEN de ces enregistrements, en faisant la somme des poids initiaux de ces enregistrements obtenus du fichier de microdonnées, puis en divisant le résultat obtenu par le nombre de répondants qui ont déclaré SEXE = homme;
- c) pour chacun de ces répondants, calculer un poids RÉ- ÉCHELONNÉ égal au poids initial de la personne, divisé par le poids MOYEN;
- d) procéder à l'analyse pour ces répondants, en utilisant le poids RÉ-ÉCHELONNÉ.

Cependant, comme on ne tient toujours pas compte de la stratification du plan d'échantillonnage, les estimations de la variance ainsi calculées seront probablement sous-estimées.

Le calcul d'estimations de la variance véritablement significative exige une connaissance détaillée du plan de l'enquête. Cette information ne peut toutefois pas être communiquée pour ce fichier de microdonnées, pour des raisons de protection de la confidentialité des données. Pour atténuer le problème, il est possible d'utiliser les effets du plan de sondage donnés dans le premier tableau de la chapitre 10.0. Pour cela, il faut trouver le champ où l'inférence doit être faite et multiplier l'estimation de variance (considère habituellement un échantillon aléatoire simple) par l'effet du plan de sondage. Cela ajustera de façon brute en considération du fait que le plan de sondage comprend une stratification et que nous faisons des ajustements pour la non-réponse. Les variances qui tiennent compte du plan d'échantillonnage complet peuvent cependant être calculées par Statistique Canada pour bon nombre de statistiques selon la formule du recouvrement des coûts.

9.4 Lignes directrices pour la diffusion en fonction du coefficient de variation

Avant de diffuser ou de publier des estimations établies à partir de l'Enquête sur les activités et les préférences en matière de voyages, les utilisateurs doivent d'abord déterminer le niveau de qualité des estimations, lequel peut être *acceptable*, *marginal* ou *inacceptable*. La qualité des données dépend à la fois des erreurs d'échantillonnage et des erreurs non dues à l'échantillonnage (voir chapitre 8.0). Cependant, aux fins de la présente analyse, le niveau de qualité d'une estimation sera déterminé uniquement en fonction de l'erreur d'échantillonnage définie par le coefficient de variation indiqué dans le tableau qui suit. Les utilisateurs doivent néanmoins lire la chapitre 8.0, pour bien comprendre les caractéristiques qui influent sur la qualité de ces données.

Il s'agit en premier lieu de déterminer le nombre de répondants qui contribuent au calcul de l'estimation. Si ce nombre est inférieur à 30, l'estimation pondérée doit être considérée comme étant de qualité inacceptable.

Pour les estimations pondérées basées sur des échantillons composés de 30 répondants ou plus, les utilisateurs doivent calculer le coefficient de variation de l'estimation et suivre les lignes directrices ci-après. Ces lignes directrices s'appliquent aux estimations pondérées arrondies.

Toutes les estimations peuvent être diffusées. Cependant, celles de qualité marginale ou inacceptable doivent être accompagnées d'une mise en garde à l'intention des utilisateurs subséquents.

Lignes directrices relatives au niveau de qualité

Niveau de qualité de l'estimation	Lignes directrices
Acceptable	<p>Caractéristiques de l'estimation : Taille de l'échantillon égale ou supérieure à 30 et coefficient de variation faible (entre 0,0 % à 16,5 %)</p> <p>Aucune mise en garde n'est requise.</p>
Marginal	<p>Caractéristiques de l'estimation : Taille de l'échantillon égale ou supérieure à 30 et coefficients de variation élevés (entre 16,6 % à 33,3 %)</p> <p>Les estimations doivent être marquées de la lettre M (ou autre identificateur similaire). Elles doivent aussi être accompagnées d'une mise en garde informant les utilisateurs subséquents sur le niveau d'erreur élevé, associé à ces estimations.</p>
Inacceptable	<p>Caractéristiques de l'estimation : 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 I (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 I) 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>

10.0 Tableaux des variabilités d'échantillonnage approximatives

Nous avons créé deux séries de tableaux des variabilités d'échantillonnage approximatives afin de déterminer les coefficients de variation qui seraient applicables à une grande variété d'estimations de type nominal établies à partir de ces fichiers de microdonnées et auxquels l'utilisateur pourrait facilement avoir accès. Ces tableaux de référence permettent à l'utilisateur d'obtenir un coefficient de variation approximatif, basé sur la taille de l'estimation calculée à partir des données de l'enquête.

Les coefficients de variation (c.v.) sont calculés par la formule d'estimation de la variance pour la méthode d'échantillonnage aléatoire simple, en incorporant un facteur qui reflète la nature stratifiée du plan d'échantillonnage et la non-réponse. Ce facteur, désigné effet du plan, a été déterminé d'abord en calculant les effets du plan de sondage pour un vaste éventail de caractéristiques, puis en choisissant parmi celles-ci une valeur conservatrice à utiliser dans les tableaux de référence qui s'appliquerait ensuite à l'ensemble des caractéristiques.

Tous les coefficients de variation indiqués dans les tableaux des variabilités d'échantillonnage approximatives sont approximatifs et donc non officiels. Les estimations de la variance réelle s'appliquant à des variables précises peuvent être obtenues de Statistique Canada, selon une formule de recouvrement des coûts. L'utilisation d'estimations exactes de la variance permettrait aux utilisateurs de diffuser des estimations qui, sans cela, ne pourraient pas être diffusées, car les coefficients de variation de ces estimations se situeraient à l'intérieur de l'intervalle de «confiance».

Il est important de se rappeler que si le nombre d'observations sur lequel une estimation est basée est inférieur à 30, l'estimation pondérée ne devrait pas être diffusée quelle que soit la valeur du coefficient de variation qui s'y rapporte. Cette exclusion est rendue nécessaire par le fait que les formules utilisées pour estimer la variance ne donnent pas les mêmes résultats avec des petits échantillons.

Le tableau qui suit indique, selon les divers niveaux géographiques (Canada, provinces ou RMR), les effets du plan de sondage, la taille des échantillons et les effectifs de la population qui ont été utilisés dans la production des tableaux de variabilités d'échantillonnage approximatives.

Données d'entrée des tableaux des variabilités d'échantillonnage approximatives

Niveau géographique	Enquête téléphonique			Enquête postale		
	Effet du plan	Taille de l'échantillon	Taille de la population	Effet du plan	Taille de l'échantillon	Taille de la population
Canada	2,21	18 385	23 261 571	2,91	5 740	19 618 857
Atlantique	2,32	1 548	1 829 572	2,77	486	1 455 265
Québec	1,64	2 103	5 755 306	2,09	640	4 547 941
Ontario	2,39	9 071	8 900 616	3,70	2 808	7 522 137
Manitoba	1,28	1 177	828 107	1,73	374	688 964
Saskatchewan	1,42	1 383	724 612	1,59	437	642 229
Alberta	1,33	1 550	2 177 501	1,76	503	1 980 387
Colombie-Britannique	1,82	1 553	3 045 857	2,41	492	2 781 933
St. John's RMR / Atlantique autre	2,05	1 012	1 553 109			
Halifax RMR	1,46	536	276 463			
Montréal RMR	1,29	750	2 694 544			
Québec RMR	1,53	586	575 614			
Québec Autre (excluant Hull)	1,29	501	2 275 061			
Ottawa-Hull RMR	1,32	1 023	857 198			
Toronto RMR	1,25	1 730	3 608 933			
Hamilton RMR	1,23	1 004	517 168			
London RMR	1,28	619	331 351			
Kitchener RMR	1,28	700	318 984			
St Catharines-Niagara RMR	1,31	846	288 348			
Windsor RMR	1,30	610	228 587			
Oshawa RMR	1,31	917	216 040			
Sudbury/Thunder Bay RMR	1,30	1 046	235 638			
Ontario Autre	1,34	842	2 508 455			
Winnipeg RMR	1,29	679	539 559			
Manitoba Autre	1,21	498	288 548			
Saskatoon/Regina RMR	1,30	856	324 433			
Saskatchewan Autre	1,29	527	400 179			
Edmonton RMR	1,38	562	713 881			
Calgary RMR	1,31	553	700 911			
Alberta Autre	1,39	435	762 709			
Vancouver RMR	1,41	632	1 538 851			
Victoria RMR	1,32	559	256 921			
Colombie-Britannique Autre	1,35	362	1 250 085			

10.1 Comment utiliser les tableaux des coefficient de variation pour les estimations de type nominal

Les règles qui suivent devraient permettre à l'utilisateur de déterminer les coefficients de variation approximatifs, à partir des tableaux de la variabilité d'échantillonnage approximative, pour des estimations du nombre, de la proportion ou du pourcentage de la population interrogée possédant une caractéristique particulière, ainsi que pour les ratios et les différences entre ces estimations.

Règle n° 1 : Estimations du nombre possédant une caractéristique particulière (agrégats)

Le coefficient de variation dépend seulement de la taille de l'estimation. Sur le tableau de la variabilité d'échantillonnage, trouver pour la région géographique concernée le nombre estimé, indiqué dans la colonne à l'extrême gauche du tableau (intitulée «Numérateur du pourcentage») et suivre les astérisques (le cas échéant) jusqu'au premier chiffre indiqué. Ce chiffre est le coefficient de variation approximatif.

Règle n° 2: Estimations de la proportion ou du pourcentage qui possède une caractéristique particulière

Le coefficient de variation d'une proportion ou d'un pourcentage dépend à la fois de la taille de cette proportion ou pourcentage, ainsi que de la taille du total sur lequel est basé cette proportion (ou pourcentage). Les proportions ou pourcentages estimatifs sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque cette proportion ou ce pourcentage est basé sur un sous-groupe de la population. À titre d'exemple, la proportion des « voyageurs en Colombie-Britannique durant les deux dernières années » est plus fiable que le nombre estimé de « voyageurs en Colombie-Britannique durant les deux dernières années ». (À noter que, dans les tableaux, la valeur du c.v. diminue de gauche à droite).

Lorsque la proportion ou le pourcentage est basé sur la population totale de la région géographique couverte par le tableau, le c.v. de la proportion ou du pourcentage est égal au c.v. du numérateur de la proportion ou du pourcentage; dans un tel cas, la règle n° 1 peut être appliquée.

Lorsque la proportion ou le pourcentage est basé sur un sous-ensemble de la population totale (p. ex. les personnes appartenant à un sexe ou à un groupe d'âge particulier), il faut alors se référer à la proportion ou au pourcentage (en haut du tableau) et au numérateur de la proportion ou du pourcentage (à gauche du tableau). Le point d'intersection de la rangée et de la colonne appropriées correspond au coefficient de variation.

Règle n° 3 : Estimations de la différence entre des agrégats ou des pourcentages

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

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

où \bar{X}_1 est l'estimation 1, \bar{X}_2 est l'estimation 2 et α_1 et α_2 sont les coefficients de variation de \bar{X}_1 et \bar{X}_2 respectivement. Le coefficient de variation de \bar{d} est représenté par $\sigma_{\bar{d}}/\bar{d}$. Cette formule est exacte pour la différence entre des caractéristiques distinctes et non corrélées, mais n'est qu'approximative dans les autres cas.

Règle n° 4 : Estimations des ratios

Lorsque le numérateur est un sous-ensemble du dénominateur, le ratio doit être exprimé en pourcentage et la règle n° 2 doit alors s'appliquer. Un tel cas se produirait, par exemple, si le dénominateur est le nombre de voyageurs et le numérateur est le nombre de non-voyageurs.

Lorsque le numérateur n'est pas un sous-ensemble du dénominateur, par exemple le ratio entre le nombre de voyageurs et le nombre de non-voyageurs, l'écart-type du ratio des estimations est à peu près égal à la racine carrée de la somme des carrés de chaque coefficient de variation examiné séparément, multipliée par R. En d'autres mots, l'erreur-type d'un ratio ($R = \bar{X}_1 / \bar{X}_2$) est égale à :

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

où α_1 et α_2 sont les coefficients de variation de \bar{X}_1 et \bar{X}_2 respectivement. Le coefficient de variation de R est représenté par σ_R/R . La formule aura tendance à surestimer l'erreur s'il existe une corrélation positive entre \bar{X}_1 et \bar{X}_2 mais à la sous-estimer si cette corrélation entre \bar{X}_1 et \bar{X}_2 est négative.

Règle n° 5 : Estimations des différences de ratios

Ici, les règles 3 et 4 sont combinées. Le c.v. pour les deux ratios est d'abord déterminé par la règle n° 4, puis le c.v. de leur différence est calculé suivant la règle n° 3.

Exemples de l'utilisation des tableaux des c.v. pour les estimations de type nominal. Nous avons inclus quelques exemples de situations réelles, pour aider les utilisateurs à appliquer les règles précitées.

Exemple 1 : Estimation du nombre possédant une caractéristique particulière (agrégats)

Supposons qu'un utilisateur estime, à partir des données de l'enquête téléphonique, que 1 142 000 Québécois(es) ont fait un voyage d'agrément en Ontario durant les deux dernières années. Comment l'utilisateur peut-il déterminer le coefficient de variation de cette estimation?

- (1) Consulter le tableau des c.v. pour le Québec. L'estimation globale (1 142 000) ne figure pas dans la colonne de gauche («Numérateur du pourcentage»); il faut donc utiliser le chiffre qui s'en rapproche le plus, en l'occurrence 1 000 000.
- (2) Le coefficient de variation d'une estimation globale correspond à la première entrée sur cette rangée qui diffère d'un astérisque, soit 6,0 %.
- (3) Le coefficient de variation approximatif de l'estimation est donc 6,0 %. Par conséquent, la conclusion selon laquelle 1 142 000 Québécois(es) ont fait un voyage d'agrément en Ontario durant les deux dernières années peut être publiée sans restriction.

Exemple 2 : Estimations de la proportion ou du pourcentage possédant une caractéristique particulière

Supposons que l'utilisateur estime à partir des données de l'enquête téléphonique que 19,7 % des Haligoniennes de 60 ans et plus (4 042 sur 20 519) ont loué des cassettes vidéo sur une base régulière. Comment l'utilisateur détermine-t-il ici le coefficient de variation de cette estimation?

- (1) Consulter le tableau sous Halifax. Comme l'estimation est un pourcentage basé sur un sous-ensemble de la population totale (c.-à-d. les Haligoniennes de 60 ans et plus), il faut utiliser à la fois le pourcentage (19,7 %) et le numérateur du pourcentage (4 042) pour déterminer le coefficient de variation.
- (2) Le numérateur, 4 042, ne figure pas dans la colonne de gauche («Numérateur du pourcentage»); il faut donc utiliser le chiffre qui s'en rapproche le plus, en l'occurrence 4 000. De même, l'estimation du pourcentage ne figure comme aucune en-tête de colonne, de sorte qu'il faut là aussi utiliser le chiffre le plus proche, soit 20,0 %.
- (3) Le chiffre qui se trouve à l'intersection de la rangée et de la colonne utilisées, 38,8 % est le coefficient de variation à utiliser.
- (4) Le coefficient de variation approximatif de l'estimation est donc 38,8 %. Par conséquent, la conclusion selon laquelle 19,7 % des Haligoniennes de 60 ans et plus louent des cassettes vidéo sur une base régulière est considérée inacceptable et Statistique Canada recommande de ne pas publier cette estimation. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre I (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 3 : Estimations des différences entre des agrégats ou des pourcentages

Supposons qu'un utilisateur estime à partir des données de l'enquête téléphonique que 51,4 % des Edmontoniens (179 168 sur 348 702) ont affirmé qu'il est très probable ou assez probable qu'ils fassent un voyage d'agrément aux États-Unis d'une nuitée ou plus au cours des deux prochaines années alors que ce pourcentage est de 37,8 % pour les Edmontoniennes (138 114 sur 365 179). Comment l'utilisateur détermine-t-il ici le coefficient de variation de la différence entre ces deux estimations?

- (1) En utilisant le tableau du c.v. sous Edmonton de la même manière qu'à l'exemple 2, on obtient le c.v. de l'estimation pour les hommes (6,6 %) et le c.v. de l'estimation pour les femmes (8,4 %).
- (2) Selon la règle n° 3, l'erreur-type d'une différence ($\bar{d} = X_1 - X_2$) est égale à :

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

où X_1 est l'estimation 1 (hommes), X_2 est l'estimation 2 (femmes) et α_1 et α_2 sont les coefficients de variation de X_1 et X_2 respectivement.

Par conséquent, l'erreur-type de la différence $\bar{d} = (0,514 - 0,378) = 0,136$ est égale à :

$$\begin{aligned}\sigma_{\bar{d}} &= \sqrt{[(0,514)(0,066)]^2 + [(0,378)(0,084)]^2} \\ &= \sqrt{(0,011508) + (0,0010081)} \\ &= 0,046\end{aligned}$$

- (3) Le coefficient de variation de \bar{d} est représenté par $\sigma_{\bar{d}}/\bar{d} = 0,046/0,136 = 0,338$
- (4) Le coefficient de variation approximatif de la différence entre les estimations est donc de 33,8 %. La différence entre les estimations est considérée inacceptable et Statistique Canada recommande de ne pas publier cette estimation. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre (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 : Estimation de ratios

Supposons que l'utilisateur estime à partir des données de l'enquête téléphonique que 21 060 Winnipegaises âgées de 60 ans et plus lisent régulièrement des revues de mode ou de décoration comme *Chatelaine*, *Vogue* et *Women's Day*, comparativement à 5 319 de leurs homologues masculins. L'utilisateur désire comparer l'estimation des femmes à celle des hommes, en les présentant sous forme de ratio. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- (1) Il faut se rappeler qu'il s'agit d'une estimation par ratio, où le numérateur de l'estimation (= X_1) correspond au nombre de Winnipegaises qui lisent régulièrement des revues de mode ou de décoration et où le dénominateur (= X_2) est le nombre Winnipegais avec les mêmes caractéristiques.
- (2) Consulter le tableau sous Winnipeg.

- (3) Le numérateur de cette estimation par ratio est 21 060; le chiffre qui s'en rapproche le plus est 21 000. Le coefficient de variation pour cette estimation correspond à la première entrée sur cette rangée qui diffère d'un astérisque, soit 21,5 %.
- (4) Le dénominateur de cette estimation par ratio est 5 319; le chiffre qui s'en rapproche le plus est 5 000. Le coefficient de variation de cette estimation correspond à la première entrée sur cette rangée qui diffère d'un astérisque, en l'occurrence 45,0 %.
- (5) Le coefficient de variation approximatif de l'estimation par ratio est obtenu par la règle n° 4, donc :

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

où α_1 et α_2 sont les coefficients de variation de X_1 et X_2 respectivement.

Donc : ,

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0,215)^2 + (0,450)^2} \\ &= \sqrt{0,046225 + 0,2025} \\ &= 0,499\end{aligned}$$

- (6) Le ratio entre le nombre de femmes et le nombre d'hommes qui lisent régulièrement des revues de mode ou de décoration comme *Chatelaine*, *Vogue* et *Women's Day* est donc 21 060 / 5 319, soit 4,0 %. Le coefficient de variation de cette estimation étant de 49,9 %, celle-ci ne peut pas être diffusée peut importe les circonstances. Il devrait plutôt être effacé et remplacé par des tirets.

10.2 Comment utiliser les tableaux des coefficients de variation pour calculer les limites de confiance

Bien que les coefficients de variation soient largement utilisés, il existe une mesure significative plus intuitive de l'erreur d'échantillonnage -- l'intervalle de confiance d'une estimation. L'intervalle de confiance est un énoncé qui précise le niveau de confiance que la valeur réelle pour la population se situe à l'intérieur d'une fourchette de valeurs précise. À titre d'exemple, un intervalle de confiance de 95 % peut être décrit comme suit :

Si l'échantillonnage de la population est répété de façon indéfinie et que chaque échantillon produit un nouvel intervalle de confiance pour une estimation, alors dans 95 % des échantillons, l'intervalle contiendra la vraie valeur de la population.

Les intervalles de confiance d'une estimation peuvent être déterminés à partir de l'erreur-type d'une estimation, en présumant que, lors d'un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique donnée de la population sont distribuées normalement autour la vraie valeur de la population. Selon cette hypothèse, les chances sont d'environ 68 sur 100 que la différence entre une estimation et la vraie valeur de la population soit inférieure à une erreur-type, d'environ 95 sur 100 que cette différence soit inférieure à deux erreurs-

types et d'environ 99 sur 100 qu'elle soit inférieure à trois erreurs-types. C'est ce qu'on appelle les niveaux de confiance.

Les intervalles de confiance pour une estimation, \hat{X} , sont généralement exprimés par deux chiffres, le premier étant inférieur à l'estimation et l'autre supérieur, sous la forme $(\hat{X}-k, \hat{X}+k)$, où la valeur de k dépend du niveau de confiance recherché et de l'erreur d'échantillonnage de l'estimation.

Les intervalles de confiance d'une estimation peuvent être calculés directement des tableaux de la variabilité d'échantillonnage approximative, en déterminant d'abord, à partir du bon tableau, le coefficient de variation de l'estimation \hat{X} , puis en utilisant la formule qui suit pour convertir le coefficient en intervalle de confiance (CI) :

$$CI_X = [\hat{X} - t\hat{X}\alpha_{\hat{X}}, \hat{X} + t\hat{X}\alpha_{\hat{X}}]$$

où α_X est le coefficient de variation déterminé de \hat{X} et

- t = 1, pour un intervalle de confiance au niveau de 68 %
- t = 1,6, pour un intervalle de confiance au niveau de 90 %
- t = 2, pour un intervalle de confiance au niveau de 95 %
- t = 2,6, pour un intervalle de confiance au niveau de 99 %

Nota : Les lignes directrices en matière de diffusion, qui s'appliquent à l'estimation, s'appliquent également à l'intervalle de confiance. Si l'estimation ne peut être diffusée, il en va de même de l'intervalle de confiance.

Exemple de l'utilisation des tableaux des c.v. pour calculer les limites de confiance

L'intervalle de confiance de 95 %, pour la proportion estimée de Britanno-Colombiens de 18 ans et plus qui ont fait un voyage d'affaire d'une nuitée ou plus au cours des deux dernières années, serait calculé comme suit :

$$\hat{X} = 17,5 \% \text{ (ou exprimé en proportion } = 0,175)$$

$$t = 2$$

$\alpha_X = 2,3 \% (0,023 \text{ exprimé en proportion})$ est le coefficient de variation de cette estimation, tel que déterminé à partir des tableaux.

$$CI_X = \{0,175 - (2) (0,175) (0,023), 0,175 + (2) (0,175) (0,023)\}$$

$$CI_X = \{0,175 - 0,008, 0,175 + 0,008\}$$

$$CI_X = \{0,167, 0,183\}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 16,7 % et 18,3 % des Britanno-Colombiens de 18 ans et plus qui ont fait un voyage d'affaire d'une nuitée ou plus au cours des deux dernières années (selon l'année de référence de l'enquête).

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

Les erreurs-types peuvent aussi être utilisées pour vérifier des hypothèses, une technique qui permet d'établir une distinction entre les paramètres de la population établis à partir d'estimations. Il peut s'agir de nombres, de moyennes, de pourcentages, de ratios, etc. De plus, les tests peuvent être effectués à différents niveaux de signification, le niveau de signification désignant ici la probabilité de conclure que les caractéristiques sont différentes alors qu'elles sont en fait identiques.

Supposons que X_1 et X_2 sont les estimations des deux caractéristiques à l'étude. Supposons également que l'erreur-type de la différence $X_1 - X_2$ est égale à σ_d .

Si $t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_d}$ se situe entre -2 et 2, alors aucune conclusion sur la différence entre les

caractéristiques n'est justifiée au niveau de signification de 5 %. Par contre, si ce ratio est inférieur à -2 ou supérieur à +2, la différence observée est alors significative au niveau de 0,05. En d'autres mots, ceci signifie que les caractéristiques sont significatives.

Exemple de l'utilisation des tableaux des c.v. pour effectuer un test t

Supposons que nous voulions vérifier, au niveau de confiance de 5 %, l'hypothèse selon laquelle il n'y a aucune différence entre la proportion d'Edmontoniens qui ont affirmé qu'il est très probable ou assez probable qu'ils fassent un voyage d'agrément aux États-Unis d'une nuitée ou plus au cours des deux prochaines années et la proportion d'Edmontoniennes. Si l'on se reporte à l'exemple 3, présenté au section 10.1, l'erreur-type de la différence entre ces deux estimations est égale à = 0,138. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_d} = \frac{0,514 - 0,378}{0,046} = \frac{0,136}{0,046} = 2,96.$$

Comme $t = 2,96$ est supérieur à 2, il faut en conclure qu'il existe une différence significative entre les deux estimations, au niveau de signification de 0,05.

10.4 Coefficients de variation des estimations quantitatives

Il faudrait produire des tableaux spéciaux pour déterminer l'erreur d'échantillonnage des estimations quantitatives. Cependant, comme la plupart des variables pour l'Enquête sur les activités et les préférences en matière de voyages sont de type nominal, ceci n'a pas été fait.

En règle générale, toutefois, le coefficient de variation d'un total quantitatif est supérieur au coefficient de variation de l'estimation de la catégorie correspondante (c.-à-d. l'estimation du nombre de personnes ayant servi à produire l'estimation quantitative). Si l'estimation de la catégorie correspondante ne peut être diffusée, l'estimation quantitative ne peut l'être non plus. À titre d'exemple, le coefficient de variation du nombre total de voyages faits au Yukon durant les deux dernières années serait supérieur au coefficient de variation de la proportion correspondante des personnes qui ont voyagé au Yukon durant les deux dernières années. Par conséquent, si le coefficient de variation de la proportion ne peut être diffusé, alors le coefficient de variation de l'estimation quantitative correspondante ne peut l'être non plus.

Les coefficients de variation de ces estimations peuvent être calculés au besoin, pour une estimation précise, à l'aide d'une technique dite de pseudo-réplication. Celle-ci consiste à diviser les enregistrements du fichier de microdonnées en sous-groupes (ou échantillons répétés) et à déterminer la variation de l'estimation, d'un échantillon à un autre. Les utilisateurs qui désirent calculer les coefficients de variation pour des estimations quantitatives peuvent contacter Statistique Canada pour obtenir des conseils sur la façon appropriée de répartir les enregistrements en échantillons répétés et sur les formules à utiliser pour ces calculs.

10.5 Seuils de diffusion pour l'Enquête sur les activités et les préférences en matière de voyages

Le tableau qui suit indique la valeur minimale de l'estimation, au niveau des provinces pour les ménages. Les estimations inférieures à celles indiquées à la colonne « Inacceptable » doivent être identifiées de la manière appropriée.

Tableau des seuils de diffusion (c.-à-d. taille minimum d'une estimation) pour l'enquête téléphonique

Niveau géographique	CV inacceptable > 33,3 %	CV médiocre 16,6 à 33,3 %	CV acceptable 0,0 à 16,5 %
Canada	Moin de 25 200	25 200 à < 102 300	102 300 et plus
Atlantique	Moin de 24 400	24 400 à < 95 500	95 500 et plus
Québec	Moin de 40 200	40 200 à < 160 300	160 300 et plus
Ontario	Moin de 21 100	21 100 à < 85 300	85 300 et plus
Manitoba	Moin de 8 000	8 000 à < 31 800	31 800 et plus
Saskatchewan	Moin de 6 600	6 600 à < 26 300	26 300 et plus
Alberta	Moin de 16 700	16 700 à < 66 500	66 500 et plus
Colombie-Britannique	Moin de 31 900	31 900 à < 125 700	125 700 et plus
St. John's RMR/Atlantique autre	Moin de 27 900	27 900 à < 107 600	107 600 et plus
Halifax RMR	Moin de 6 600	6 600 à < 25 100	25 100 et plus
Montréal RMR	Moin de 41 200	41 200 à < 160 100	160 100 et plus
Québec RMR	Moin de 13 200	13 200 à < 50 400	50 400 et plus
Québec Autre (excluant Hull)	Moin de 51 600	51 600 à < 196 600	196 600 et plus
Ottawa-Hull RMR	Moin de 9 900	9 900 à < 38 800	38 800 et plus
Toronto RMR	Moin de 23 400	23 400 à < 93 300	93 300 et plus
Hamilton RMR	Moin de 5 700	5 700 à < 22 300	22 300 et plus
London RMR	Moin de 6 100	6 100 à < 23 400	23 400 et plus
Kitchener RMR	Moin de 5 200	5 200 à < 20 100	20 100 et plus
St Catharines-Niagara RMR	Moin de 4 000	4 000 à < 15 500	15 500 et plus
Windsor RMR	Moin de 4 300	4 300 à < 16 600	16 600 et plus
Oshawa RMR	Moin de 2 700	2 700 à < 10 800	10 800 et plus
Sudbury/Thunder Bay RMR	Moin de 2 600	2 600 à < 10 300	10 300 et plus
Ontario Autre	Moin de 35 500	35 500 à < 138 500	138 500 et plus
Winnipeg RMR	Moin de 9 100	9 100 à < 35 200	35 200 et plus
Manitoba Autre	Moin de 6 200	6 200 à < 23 600	23 600 et plus
Saskatoon/Regina RMR	Moin de 4 400	4 400 à < 17 100	17 100 et plus
Saskatchewan Autre	Moin de 8 600	8 600 à < 33 000	33 000 et plus
Edmonton RMR	Moin de 15 500	15 500 à < 59 100	59 100 et plus
Calgary RMR	Moin de 14 700	14 700 à < 56 100	56 100 et plus
Alberta Autre	Moin de 21 400	21 400 à < 80 100	80 100 et plus
Vancouver RMR	Moin de 30 400	30 400 à < 116 600	116 600 et plus
Victoria RMR	Moin de 5 400	5 400 à < 20 500	20 500 et plus
Colombie-Britannique Autre	Moin de 40 700	40 700 à < 150 600	150 600 et plus

Tableau des seuils de diffusion (c.-à-d. taille minimum d'une estimation) pour l'enquête postale

Niveau géographique	CV inacceptable > 33,3 %	CV médiocre 16,6 à 33,3 %	CV acceptable 0,0 à 16,5 %
Canada	Moins de 89 300	89 300 à < 358 700	358 700 et plus
Atlantique	Moins de 71 100	71 100 à < 251 900	251 900 et plus
Québec	Moins de 130 100	130 100 à < 487 100	487 100 et plus
Ontario	Moins de 88 300	88 300 à < 347 300	347 300 et plus
Manitoba	Moins de 27 600	27 600 à < 100 100	100 100 et plus
Saskatchewan	Moins de 20 400	20 400 à < 75 700	75 700 et plus
Alberta	Moins de 60 600	60 600 à < 225 500	225 500 et plus
Colombie-Britannique	Moins de 117 700	117 700 à < 424 200	424 200 et plus

10.6 Tableaux des coefficients de variation

Veillez consulter les fichiers indiqués ci-après pour obtenir les tableaux des c.v. de l'Enquête sur les activités et les préférences en matière de voyages :

EAPV1999_M_CVTabF.pdf (enquête postale)

EAPV1999_T_CVTabF.pdf (enquête téléphonique)

11.0 Pondération

Nous décrivons dans ce chapitre le calcul des poids d'enquête de l'Enquête sur les activités et les préférences en matière de voyages (EAPV). La pondération a été faite pour l'enquête téléphonique, puis pour l'enquête postale. La stratégie de base pour l'ajustement de la non-réponse demeure la même dans les deux enquêtes. Les étapes de l'enquête importantes pour la pondération étaient les suivantes :

- 1) Obtenir un échantillon aléatoire simple de numéros de téléphone dans 28 strates géographiques;
- 2) Déterminer si le numéro de téléphone est un numéro de téléphone résidentiel en fonction;
- 3) Obtenir la liste des membres du ménage;
- 4) Obtenir une entrevue téléphonique avec une personne du ménage choisie aléatoirement;
- 5) Faire remplir le questionnaire postal aux voyageurs qui ont répondu à l'enquête téléphonique.

Afin de prendre en considération l'information que nous obtenons à chaque étape, nous avons ajusté la probabilité initiale de sélection aux poids d'enquête. Pour chacune des étapes, nous avons apporté les ajustements suivants :

- 1) Ajustement de la non-réponse des ménages au niveau régional;
- 2) Ajustement de la taille du ménage et du nombre de téléphones du ménage au niveau du ménage;
- 3) Ajustement de la non-réponse individuelle au niveau de la taille du ménage;
- 4) Ajustement de la non-réponse à l'enquête postale en utilisant un groupement âge / sexe / statut socio-économique / pays d'origine.

À noter que le niveau du Bureau régional est défini comme étant Montréal, Sturgeon Falls, Winnipeg et Vancouver, et que le niveau de la taille du ménage était les ménages d'une personne par rapport aux ménages de deux personnes ou plus.

Au départ du moins, il y avait peu d'information sur les ajustements, mais à mesure que nous passons d'une étape à l'autre, il y eut de plus en plus de données. Lorsque nous avons fait l'ajustement de la non-réponse de l'enquête postale, nous avons pris en considération le potentiel de toutes les variables téléphoniques pour l'ajustement de l'enquête postale. À la fin, nous avons défini les statuts socio-économiques « inférieur », « moyen » et « supérieur » fondés sur l'éducation et le revenu déclarés. Les trois groupements d'âge et de sexe ont été fondés sur les schémas différentiels de réponse. Les répondants nés à l'extérieur du pays ont été regroupés. En fait, nous avons créé des groupes de réponses homogènes pour minimiser le biais mesurable de non-réponse.

Après avoir mis en application ces ajustements, nous avons fait la stratification a posteriori des totaux de la population. Pour l'enquête téléphonique, nous avons fait la stratification a posteriori de façon séquentielle (1) des totaux des strates âge/sexe, et (2) des totaux âge/sexe provinciaux dérivés des totaux du Recensement de 1996 de la Division de la démographie de Statistique Canada. Pour l'enquête postale, nous avons stratifié a posteriori le nombre estimé de voyageurs par strate, fondé sur les taux de voyage des strates. Nous sommes parvenus à deux poids d'enquête, un pour l'enquête téléphonique et un pour l'enquête postale.

Les strates sont les 28 champs infraprovinciaux que l'on retrouve dans la plupart des tableaux de ce document. Nous avons utilisé les groupements d'âge 18-34 ans, 35-49 ans et 50 ans et plus pour la première stratification a posteriori des strates. Pour la répartition âge/sexe infra provinciale, nous n'avons pas utilisé la répartition propre à chaque province mais plutôt le groupement des provinces de l'Atlantique, du Québec, de l'Ontario, du Manitoba, de la Saskatchewan, de l'Alberta et de la Colombie-Britannique. La répartition des âges était la suivante : 18-24 ans, 25-29 ans, 30-34 ans, 35-39 ans, 40-44 ans, 45-49 ans, 50-54 ans, 55-59 ans, 60-64 ans, 65-69 ans et 70 ans et plus.

Maintenant que nous avons expliqué le processus avec des mots, passons maintenant au traitement mathématique de la pondération. Nous utiliserons les notations suivantes :

h	indiquera la strate, où $h = 1, 2, \dots, 28$;
π_h	sera la probabilité initiale de sélection de numéros de téléphone dans la strate h ;
i	sera le ménage dont la liste des membres est complète;
t_i	sera le nombre de téléphones dans le ménage i ;
n_i	sera le nombre de personnes éligibles dans le ménage i ;
j_1	sera l'ajustement de la classe du ménage par bureau régional, $j_1 = 1, 2, 3, 4$;
a_{j_1}	sera l'ajustement de la non-réponse du ménage avec l'indice j_1 ;
j_2	sera l'ajustement de la classe de la personne par la taille du ménage, $j_2 = 1, 2$;
a_{j_2}	sera l'ajustement de la non-réponse de la personne avec l'indice j_2 ;
$I(\cdot)$	sera un indicateur variable égal à 1 si les conditions sont remplies, sinon égal à 0;
w^*	sera les poids séquentiels de l'enquête selon le nombre d'astérisques;
w^{tel}	sera la pondération finale de l'enquête téléphonique pour les réponses téléphoniques complétées;
w^{mail}	sera la pondération finale de l'enquête postale pour les réponses postales complétées.

Pour la stratification a posteriori à deux étapes de l'enquête téléphonique,

k_1	sera la première strate téléphonique de second niveau, $k_1 = 1, 2, 3, \dots, 174$;
Mk_1	sera le total de la première strate téléphonique de second niveau avec l'indice k_1 ;
k_2	sera la deuxième strate téléphonique de second niveau, $k_2 = 1, 2, 3, \dots, 154$; et
Mk_2	sera le total de la deuxième strate téléphonique de second niveau avec l'indice k_2 .

Enfin, pour l'échantillon de l'enquête postale,

l_1	sera l'ajustement de la classe postale pour l'enquête postale;
al_1	sera l'ajustement de la non-réponse de l'enquête postale avec l'indice l_1 ;
l_2	sera la strate postale de second niveau;
MI_2	sera le total de la strate postale de second niveau avec l'indice l_2 .

Commençons en définissant l'ajustement de la non-réponse du ménage (les sommes pour les formules ci-dessous sont pour tous les numéros de téléphone admissibles), la pondération initiale individuelle (qui prend en considération les probabilités initiales d'échantillonnage, le nombre de téléphones par ménage et le nombre de membres admissibles dans le ménage) et l'ajustement de la non-réponse individuelle.

$$a_{j_1} = \frac{\sum \pi_h^{-1} I (RO=j_1 \text{ \& } \text{numéro de téléphone admissible})}{\sum \pi_h^{-1} I (RO=j_1 \text{ \& } \text{numéro de téléphone admissible \& } \text{liste du ménage obtenue})}$$

$$w_{hi} = \left(\frac{1}{\pi_h} \right) \left(\frac{n_i}{t_i} \right)$$

$$a_{j_2} = \frac{\sum_h \sum_i \sum_{j_1} w_{hi} a_{j_1} I (\text{taille du ménage}=j_2 \text{ et liste du ménage})}{\sum_h \sum_i \sum_{j_1} w_{hi} a_{j_1} I (\text{taille du ménage}=j_2 \text{ et liste du ménage et interview téléphonique complète})}$$

La stratification a posteriori constitue les deux dernières étapes pour l'obtention de la pondération téléphonique. La première de ces deux étapes est d'ajuster le poids pour que la pondération totale des réponses téléphoniques s'ajoute aux totaux de contrôle des strates. Nous savons, toutefois, que les totaux de contrôle de niveau infraprovincial, dérivés des estimations démographiques, ne sont pas aussi précis que les totaux de contrôle âge/sexes provinciaux. C'est pourquoi nous faisons la stratification a posteriori des totaux des strates et nous utilisons ensuite les totaux de contrôle âge/sexes provinciaux pour effectuer la stratification a posteriori téléphonique finale.

Les étapes de la stratification a posteriori sont de simples ajustements du ratio des totaux de contrôle estimés avec les totaux de contrôle réels. Afin de simplifier la notation, nous avons réduit la pondération téléphonique en ne gardant que les indices i et j , correspondant au ménage répondant dans une strate. Les deux étapes de stratification a posteriori apparaissent donc comme suit :

$$\hat{M}_{k_1} = \sum_h \sum_i \sum_{j_1} \sum_{j_2} w_{hi} a_{j_1} a_{j_2} I (\text{strate}=k_1)$$

$$\hat{M}_{k_2} = \sum_h \sum_i \sum_{j_1} \sum_{j_2} \sum_{k_1} w_{hi} a_{j_1} a_{j_2} \left(\frac{M_{k_1}}{\hat{M}_{k_1}} \right)$$

$$w_{hi}^{tel} = w_{hi} a_{j_1} a_{j_2} \left(\frac{M_{k_1}}{\hat{M}_{k_1}} \right) \left(\frac{M_{k_2}}{\hat{M}_{k_2}} \right) \quad \forall i \in \text{échantillon téléphonique répondant dans la strate } h.$$

Pour la pondération postale, nous avons ajusté la non-réponse postale avec les groupes de réponses homogènes décrits plus tôt. Nous avons ensuite utilisé les totaux dérivés du contrôle des voyageurs par strate M_{l_2} , pour en arriver à la pondération finale. Les sommes suivantes sont pour tous les répondants de l'enquête téléphonique.

Les indices I_1 et I_2 ont encore une fois été réduits pour de raisons de clarté à l'étape finale. À noter que la définition

$$a_{I_1} = \frac{\sum_h \sum_i w_{hi}^{tel} I(\text{socio-éco} * \text{âge/sexe} * \text{pays d'origine}=I_1 \ \& \ \text{voyageur})}{\sum_h \sum_i w_{hi}^{tel} I(\text{socio-éco} * \text{âge/sexe} * \text{pays d'origine}=I_1 \ \& \ \text{voyageur} \ \& \ \text{réponse postale})}$$

$$M_{I_2} = \sum_h \sum_i w_{hi}^{tel} I(\text{strate}=I_2 \ \& \ \text{voyageur au téléphone})$$

$$\hat{M}_{I_2} = \sum_h \sum_i \sum_{I_1} w_{hi}^{tel} a_{I_1} I(\text{strate}=I_2 \ \& \ \text{réponse postale})$$

$$w_{hi}^{mail} = w_{hi}^{tel} a_{I_1} \left(\frac{M_{I_2}}{\hat{M}_{I_2}} \right) \quad \forall i \in \text{échantillon répondants poste dans la strate } h.$$

de voyageur dans les équations ci-dessus est une personne qui a affirmé être un voyageur lors de l'interview téléphonique (d'où « voyageur au téléphone »).

12.0 Questionnaires et fiches de codage

Veillez consulter les fichiers indiqués ci-après pour obtenir les questionnaires des microdonnées de l'Enquête sur les activités et les préférences en matière de voyages (EAPV) :

EAPV1999_P_Quest_F.pdf (Questionnaire postal)

EPAV1999_T_Quest_F.pdf (Questionnaire téléphonique)

13.0 Cliché d'enregistrement et variables uniques

Veillez consulter les fichiers indiqués ci-après pour obtenir les fichiers des microdonnées et les guides de codage des microdonnées de l'Enquête sur les activités et les préférences en matière de voyages (EAPV) :

EAPV1999_P_LvCds.pdf (Livres de codes postale)

EAPV1999_T_LvCds.pdf (Livres de codes téléphonique)