



Guide de l'utilisateur des microdonnées

Enquête sur les activités et les préférences en matière de voyages

2006



Statistique
Canada

Statistics
Canada

Canada

Table des matières

1.0	Introduction	5
2.0	Contexte	7
3.0	Objectifs	9
4.0	Concepts et définitions	11
5.0	Méthodologie de l'enquête	13
5.1	Population visée	13
5.2	Stratification de l'échantillon	13
5.3	Échantillonnage par composition aléatoire	13
5.4	Plan de sondage	14
5.5	Taille de l'échantillon	15
6.0	Collecte des données	17
6.1	Réalisation des interviews	17
6.2	Supervision et contrôle de qualité	17
6.3	Non-réponse	17
7.0	Traitement des données	19
7.1	Saisie des données	19
7.2	Vérification	19
7.3	Imputation	20
7.4	Création de variables dérivées	20
7.5	Pondération	20
7.6	Suppression de renseignements confidentiels	20
8.0	Qualité des données	23
8.1	Taux de réponse	23
8.2	Erreurs relatives à l'enquête	26
8.2.1	Couverture de la base de sondage	27
8.2.2	Collecte des données	27
8.2.3	Traitement des données	28
8.2.4	Non-réponse	29
8.2.5	Mesure de l'erreur d'échantillonnage	29
9.0	Lignes directrices pour la totalisation, l'analyse et la diffusion de données	31
9.1	Lignes directrices pour l'arrondissement d'estimations	31
9.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation	32
9.3	Estimations catégoriques	32
9.3.1	Totalisation d'estimations catégoriques	32
9.4	Lignes directrices pour l'analyse statistique	33
9.5	Lignes directrices pour la diffusion de coefficients de variation	34
9.6	Seuils pour la diffusion des estimations pour l'Enquête sur les activités et les préférences en matière de voyages	35

10.0	Tables de variabilité d'échantillonnage approximative	37
10.1	Comment utiliser les tables de coefficients de variation pour des estimations catégoriques.....	39
10.1.1	Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques	41
10.2	Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance	44
10.2.1	Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance	45
10.3	Comment utiliser les tables de coefficients de variation pour effectuer un test t.....	45
10.3.1	Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t	46
10.4	Coefficients de variation pour des estimations quantitatives.....	46
10.5	Tables des coefficients de variation.....	46
11.0	Pondération	47
11.1	Procédures de pondération pour l'enquête téléphonique	47
11.2	Procédures de pondération pour l'enquête postale	49
12.0	Questionnaires	51
13.0	Cliché d'enregistrement à valeurs univariées	53

1.0 Introduction

Statistique Canada a mené l'Enquête sur les activités et les préférences en matière de voyages (EAPV) en 2006 avec l'appui et la collaboration de huit ministères et organismes provinciaux et territoriaux chargés du tourisme, ainsi que de la Commission canadienne du tourisme, de Parcs Canada, de Patrimoine canadien et du Partenariat de l'industrie touristique du Canada atlantique. Ce manuel a été produit pour faciliter la manipulation du fichier de microdonnées portant sur les résultats de l'enquête.

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

Statistique Canada

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

Ministère du tourisme de l'Ontario

Alex Athanassakos
Ministère du tourisme de l'Ontario
700, rue Bay, 15^e étage
Toronto (Ontario) M7A 2E1
Téléphone : (416) 314-7317
Télécopieur : (416) 314-7341
Courriel : Alex.Athanassakos@ontario.ca

2.0 Contexte

L'Enquête sur les activités et les préférences en matière de voyages (EAPV) de 2006 a été menée entre janvier et juin 2006 pour recueillir des renseignements sur les habitudes de voyage des Canadiens, leur participation aux activités récréatives et leur motivation à voyager.

Cette enquête en deux phases comprenait d'abord une interview téléphonique assistée par ordinateur (ITAO) pour distinguer les voyageurs et les non-voyageurs, puis un questionnaire papier de suivi, envoyé par la poste aux voyageurs.

Le questionnaire de l'EAPV de 2006 a été modifié par rapport à celui de l'enquête précédente, menée en 1999.

Les ministères provinciaux du tourisme, ainsi que des organismes et des ministères fédéraux, utilisent les données de l'EAPV. Parmi les autres utilisateurs, on trouve les médias, les entreprises, les cabinets d'experts-conseils, les universités et d'autres chercheurs qui s'intéressent aux voyageurs canadiens.

3.0 Objectifs

L'objectif global de l'enquête est de recueillir des renseignements sur les activités des voyageurs canadiens et leur motivation à voyager.

Les autres objectifs de l'Enquête sur les activités et les préférences en matière de voyages sont les suivants :

- recueillir des renseignements sur les voyages d'une nuit ou plus effectués à l'extérieur de la ville au cours des deux dernières années au Canada, aux États-Unis et à l'étranger;
- recueillir des renseignements sur les types d'activités récréatives et de divertissement auxquelles s'adonnent les voyageurs;
- tracer le profil des expériences de voyage et des facteurs de motivation selon les caractéristiques sociodémographiques;
- cerner les types d'activités de voyage qui incitent les gens à voyager;
- examiner le rapport entre les voyages, la participation aux activités et les destinations;
- recueillir des renseignements sur la planification des vacances;
- mieux comprendre les raisons de ne pas voyager.

4.0 Concepts et définitions

Ce chapitre donne un aperçu des concepts et des définitions d'intérêt pour les utilisateurs. Les utilisateurs sont priés de se reporter au chapitre 12.0 de ce document où figure une copie du (des) questionnaire(s) d'enquête réellement employé(s).

Membre du ménage

Toute personne qui, au moment de l'établissement de la liste des membres du ménage :

- considère ou est réputée considérer le logement comme son domicile habituel; ou
- habite le logement et n'a pas d'autre domicile habituel ailleurs.

Cette définition englobe :

- un conjoint, un partenaire ou un enfant temporairement absent en raison du travail ou des études, mais dont le logement est le domicile habituel, qu'il a habité pendant au moins 30 jours au cours des 12 derniers mois;
- des enfants en garde physique conjointe;
- une personne qui réside temporairement dans un établissement et qui a été absente de son domicile pendant moins de six mois;
- une personne qui demande le statut de réfugié;
- un étudiant qui fréquente l'école au Canada en vertu d'un visa d'étudiant;
- une personne qui réside au Canada en vertu d'un permis de travail.

Répondant sélectionné

Membre du ménage, âgé de 18 ans et plus, choisi au hasard pendant l'interview téléphonique pour répondre à l'enquête.

Voyageur

Pour les besoins de l'Enquête sur les activités et les préférences en matière de voyages, on entend par voyageur une personne ayant fait un voyage d'une nuit ou plus à l'extérieur de la ville au cours des deux dernières années.

Activité principale

Activité à laquelle le répondant consacre la plus grande partie de son temps.

Jours de vacances payées

Nombre de jours de congé payé qu'une personne accumule chaque année à un emploi rémunéré. La personne n'a pas nécessairement utilisé tous ses jours de vacances pendant l'année.

Personne travaillant à un emploi ou une entreprise

Personne qui est un travailleur rémunéré, un travailleur autonome dans son entreprise, sa profession ou son entreprise agricole ou un travailleur non rémunéré dans une entreprise familiale ou agricole. Cette définition englobe toute activité accomplie par le répondant contre rémunération ou bénéfice, y compris le travail à temps partiel, et le « paiement en nature » (paiement en biens ou en services plutôt qu'en argent). Elle exclut le travail ménager et le bénévolat, par exemple pour une église.

Travailleur autonome

Personne qui tire directement un revenu de son entreprise, de son métier ou de sa profession, au lieu de recevoir un salaire ou un traitement d'un employeur.

5.0 Méthodologie de l'enquête

L'enquête téléphonique a été menée de janvier à avril 2006 selon la méthode d'échantillonnage par composition aléatoire (CA). Lors de l'enquête postale de suivi, menée entre janvier et juin 2006, on a utilisé les adresses obtenues pendant l'enquête téléphonique pour contacter les voyageurs.

5.1 Population visée

La population cible de l'enquête téléphonique correspondait à l'ensemble des personnes âgées de 18 ans et plus de chacune des 10 provinces canadiennes, sauf celles qui résident à temps plein dans des établissements. L'enquête ayant été menée à l'aide d'un échantillon de numéros de téléphone, les ménages (et, par conséquent, les personnes qui vivent dans ces ménages) qui n'ont pas de ligne téléphonique terrestre ont été exclus de la population échantillonnée. Les personnes sans téléphone et celles qui possèdent uniquement un téléphone cellulaire ont donc été exclues. Les personnes sans ligne terrestre comptent pour moins de 6 % de la population cible. Toutefois, on a ajusté les estimations d'enquête par la pondération pour tenir compte des personnes sans ligne terrestre.

La population cible de l'enquête postale correspondait à l'ensemble des personnes âgées de 18 ans et plus de chacune des 10 provinces canadiennes, sauf celles qui résident à temps plein dans des établissements et les non-voyageurs, qui avaient fait un voyage d'une nuit ou plus à l'extérieur de la ville au cours des deux dernières années. On a identifié les voyageurs au moyen d'une question de sélection pendant l'interview téléphonique.

5.2 Stratification de l'échantillon

L'échantillon a été stratifié comme suit au niveau des régions métropolitaines de recensement (RMR) :

- dans les provinces de l'Atlantique : Halifax, Saint John, St. John's, Atlantique (autres);
- au Québec : Montréal, Québec, Gatineau, Québec (autres);
- en Ontario : Toronto, Ottawa, Hamilton, London, Kitchener, St. Catharines-Niagara, Windsor, Oshawa, Grand Sudbury, Kingston, Thunder Bay, Ontario (autres);
- au Manitoba : Winnipeg, Manitoba (autres);
- en Saskatchewan : Saskatoon, Regina, Saskatchewan (autres);
- en Alberta : Edmonton, Calgary, Alberta (autres);
- en Colombie-Britannique : Vancouver, Victoria, Abbotsford, Colombie-Britannique (autres).

5.3 Échantillonnage par composition aléatoire

On a sélectionné l'échantillon de l'EAPV au moyen de la composition aléatoire (CA), technique qui consiste à générer des numéros de téléphone au hasard, par ordinateur. La méthode utilise le concept des banques : tous les numéros de téléphone canadiens comportent un indicatif régional et un préfixe suivis de quatre chiffres. L'indicatif régional, le préfixe et les deux chiffres qui suivent définissent la banque des centaines. Par exemple, le numéro de téléphone 613-555-1234 appartient à l'indicatif régional 613, au préfixe 555 et à la banque 61355512.

La base de sondage CA comprend des banques de numéros en service compilées à partir des fichiers administratifs de compagnies de téléphone. Pour les besoins des enquêtes sociales, on entend par banque de numéros en service une banque contenant au moins un numéro de téléphone résidentiel en service. Toutes celles qui contiennent uniquement des numéros non attribués, hors service ou des numéros d'entreprise sont exclues de la base de sondage CA. Au Canada, il existe environ 269 000 banques de numéros en service (plus de 26 millions de numéros

possibles). Chacune des banques est attribuée à une province et, à l'intérieur d'une province, à une RMR ou à la partie hors RMR de la province. L'attribution est fondée sur l'indicatif régional et le préfixe de la banque.

On a sélectionné l'échantillon CA de l'EAPV en plusieurs étapes. La première consistait à sélectionner un vaste échantillon aléatoire simple avec remise de banques, à l'intérieur de chaque RMR et de chaque région hors RMR (voir la section 5.2). Pour chaque banque sélectionnée, on a choisi au hasard parmi les 100 possibilités, de 00 à 99, les deux derniers chiffres nécessaires pour composer le numéro de téléphone. On a ensuite stratifié les numéros de téléphone ainsi générés en trois strates : numéros résidentiels, numéros d'entreprise et numéros inconnus. La dernière étape consistait à sélectionner un échantillon aléatoire simple sans remise à partir de la strate résidentielle et un autre à partir de la strate « inconnu » à l'intérieur de chaque RMR et de chaque région hors RMR.

5.4 Plan de sondage

L'échantillon CA est un échantillon aléatoire simple stratifié de numéros de téléphone, sélectionné avec remise. On a stratifié l'échantillon par RMR (voir la section 5.2) et par type de ligne téléphonique (numéro résidentiel ou inconnu). On a procédé au filtrage des numéros de téléphone « inconnus » pour en retirer les numéros hors service avant d'envoyer l'échantillon à l'unité de l'interview téléphonique assistée par ordinateur (ITAO).

On a composé chaque numéro de téléphone de l'échantillon ITAO pour déterminer s'il permettait de joindre un ménage. Dans l'affirmative, on a demandé à la personne qui avait répondu au téléphone d'énumérer tous les membres du ménage et de préciser l'âge et le sexe de chacun. On a choisi au hasard un membre âgé de 18 ans et plus pour participer à l'enquête. L'unité finale d'échantillonnage est donc la personne choisie.

On a demandé à tous les répondants identifiés comme des voyageurs pendant l'interview téléphonique de remplir un questionnaire papier à retourner par la poste. Le plan d'échantillonnage de l'enquête postale est un plan à deux phases : la première est l'interview téléphonique et la deuxième est le questionnaire papier à l'intention des voyageurs seulement.

5.5 Taille de l'échantillon

L'échantillon CA initial comprenait 132 065 numéros de téléphone à l'échelle nationale, répartis par strates comme le montre le tableau ci-dessous. Pour accroître la proportion de numéros productifs dans l'échantillon initial, on a suréchantillonné les numéros de téléphone résidentiels par rapport aux numéros « inconnus ». Les tableaux de la section 8.1 présentent des données sur le nombre de répondants et les taux de réponse.

Taille de l'échantillon initial par strate

Région métropolitaine de recensement	Résidentiels	Inconnus	Total
St. John's	663	455	1 118
Halifax	2 169	1 357	3 527
Saint John	642	568	1 210
Atlantique (autres)	3 090	1 778	4 868
Québec	3 556	938	4 493
Montréal	6 613	2 174	8 787
Gatineau	3 475	1 037	4 512
Québec (autres)	5 491	2 181	7 672
Ottawa	3 202	1 531	4 734
Kingston	644	170	813
Oshawa	732	176	908
Toronto	12 426	6 262	18 688
Hamilton	2 994	1 146	4 141
St.Catharines-Niagara	924	401	1 325
Kitchener	2 279	840	3 119
London	2 705	1 089	3 794
Windsor	837	297	1 134
Grand Sudbury	817	420	1 237
Thunder Bay	539	293	831
Ontario (autres)	5 939	2 871	8 810
Winnipeg	2 699	1 126	3 826
Manitoba (autres)	2 664	1 278	3 942
Regina	2 229	899	3 128
Saskatoon	2 178	689	2 867
Saskatchewan (autres)	2 476	1 200	3 676
Calgary	3 163	863	4 026
Edmonton	3 119	766	3 885
Alberta (autres)	3 292	1 186	4 479
Abbotsford	845	193	1 037
Vancouver	6 372	1 748	8 120
Victoria	2 683	480	3 163
Colombie-Britannique (autres)	3 442	754	4 196
Canada	94 898	37 168	132 065

6.0 Collecte des données

La collecte des données de l'Enquête sur les activités et les préférences en matière de voyages (EAPV) a eu lieu de janvier à juin 2006. Elle comprenait deux phases : une enquête téléphonique pour distinguer les voyageurs des non-voyageurs et une enquête postale axée sur les voyageurs.

6.1 Réalisation des interviews

On a mené l'enquête téléphonique au moyen de l'interview assistée par ordinateur (IAO). La collecte des données de l'enquête téléphonique a eu lieu de janvier à la mi-avril 2006. Les interviews ont été menées aux bureaux régionaux de Statistique Canada à Halifax, Sherbrooke, Sturgeon Falls, Winnipeg et Edmonton.

Le personnel de Statistique Canada affecté à l'EAPV, soit les surveillants de projet, les intervieweurs principaux et les intervieweurs, ont participé à une vidéoconférence de formation pour se familiariser avec les objectifs et les concepts de l'enquête, le questionnaire IAO et les méthodes propres à l'EAPV. On a fourni aux intervieweurs un manuel de l'intervieweur à utiliser pendant la collecte des données.

La participation à l'enquête était volontaire. Les réponses par procuration au nom du répondant sélectionné n'étaient pas admises.

Après avoir tout mis en œuvre pour interviewer un répondant sélectionné, on attribuait au cas un code d'état final et on le retournait au bureau central.

La collecte des données de l'enquête postale a eu lieu de la mi-janvier à la mi-juin 2006. Cette enquête a été administrée par la Division des opérations et de l'intégration de Statistique Canada. Le personnel de cette division a effectué un suivi téléphonique auprès des voyageurs non répondants pendant la période de collecte.

6.2 Supervision et contrôle de qualité

Tous les intervieweurs travaillent sous la supervision d'un groupe d'intervieweurs principaux. Ceux-ci ont pour responsabilité de s'assurer que les intervieweurs connaissent bien les concepts de l'EAPV et les procédures qui y sont associées. Ils ont aussi pour fonction de contrôler régulièrement le travail des intervieweurs et d'examiner les documents complétés par ces derniers. Les intervieweurs principaux sont, à leur tour, supervisés par les gestionnaires du programme, affectés dans chacun des bureaux régionaux de Statistique Canada.

6.3 Non-réponse

Les intervieweurs ont pour consigne de faire tous les efforts raisonnables pour obtenir des interviews avec les membres sélectionnés des ménages admissibles. Lorsqu'une personne refuse au départ de participer, le bureau régional envoie à l'adresse du logement une lettre soulignant l'importance que revêtent l'enquête et la collaboration du ménage. Cette lettre est suivie d'un deuxième appel de l'intervieweur. Dans les cas où l'appel de l'intervieweur survient à un moment inopportun, on fixe un rendez-vous pour une visite de rappel à un moment plus opportun. Lorsqu'il n'y a personne dans le logement, l'intervieweur rappelle à de nombreuses reprises. Quelles que soient les circonstances, on ne remplace jamais des logements échantillonnés par d'autres logements en cas de non-réponse au questionnaire.

7.0 Traitement des données

Le principal produit de l'Enquête sur les activités et les préférences en matière de voyages (EAPV) est un fichier de microdonnées « épuré ». Ce chapitre présente un bref résumé des phases de traitement inhérentes à la production de ce fichier.

7.1 Saisie des données

Les intervieweurs saisissent directement les réponses aux questions de l'enquête au moment de l'interview à l'aide d'une version automatisée du questionnaire. L'emploi d'un questionnaire automatisé réduit les délais et coûts de traitement associés à la saisie des données, aux erreurs de transcription et à la transmission des données. On soumet les données des réponses à un cryptage pour en préserver le caractère confidentiel et elles sont transmises à Ottawa, au moyen d'une ligne protégée, pour y être traitées.

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

Tous les questionnaires postaux ont été retournés à la Division des opérations et de l'intégration, au bureau central, pour la saisie des données par un système informatisé de reconnaissance optique de caractères (ROC). Les enregistrements terminés ont été transférés électroniquement à la Division des enquêtes spéciales pour le traitement.

7.2 Vérification

La première étape du traitement d'enquête effectué au bureau central a été de remplacer les valeurs « hors limites » incluses dans le fichier de données par des blancs. Ce processus a été conçu pour faciliter les vérifications ultérieures.

Les erreurs dans le déroulement du questionnaire, où l'on a relevé des questions qui ne s'appliquaient pas au répondant (et auxquelles on n'aurait donc pas dû répondre) renfermaient des réponses constituaient le premier type d'erreurs traitées. Dans ces cas, une vérification par ordinateur a éliminé automatiquement les données superflues en suivant l'ordre du questionnaire dicté par les réponses à des questions antérieures et subséquentes, parfois.

Le second type d'erreurs traitées avait trait à un manque d'information dans les questions pour lesquelles le répondant aurait dû répondre. Pour ce type d'erreur, un code de non-réponse ou « non déclaré » était attribué à l'item.

7.3 Imputation

L'imputation consiste à attribuer des valeurs valides aux variables qui contiennent des données invalides ou manquantes. Il n'y avait pas lieu d'imputer la plupart des items du questionnaire téléphonique ou du questionnaire postal. On a attribué le code « Non déclaré » aux valeurs manquantes. Les enregistrements dont les données étaient jugées insuffisantes ont été retirés du traitement.

La seule variable qui a fait l'objet d'une imputation est la région métropolitaine de recensement (RMR). À l'étape de la sélection de l'échantillon, on a d'abord attribué des codes RMR en utilisant les six premiers chiffres du numéro de téléphone (l'indicatif régional et le préfixe). Puis, à l'aide du Fichier de conversion des codes postaux (FCCP) on a dérivé des codes RMR de meilleure qualité à partir des codes postaux recueillis lors de la collecte. On a mis en œuvre un processus de validation pour vérifier les codes postaux et les codes RMR dérivés. On a procédé à une imputation par donneur pour les enregistrements dont la RMR était jugée invalide (1 % des enregistrements) ou dont le code postal était manquant (5 % des enregistrements).

7.4 Création de variables dérivées

Un certain nombre de données élémentaires incluses dans le fichier de microdonnées ont été calculées en combinant des postes sur le questionnaire pour faciliter l'analyse des données. Il s'agit des groupes d'âge, du niveau de scolarité, du pays de naissance, de la composition du ménage et du nombre de jours de vacances.

7.5 Pondération

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

La phase de la pondération est une étape où l'on calcule ce nombre (ou poids) pour chaque enregistrement. Ce poids, qui figure dans le fichier de microdonnées, **doit** servir à calculer des estimations significatives à partir de l'enquête. Si, par exemple, le nombre de personnes ayant voyagé au cours des deux dernières années, doit être estimé, cette opération s'effectue en sélectionnant les enregistrements se référant aux personnes incluses à l'intérieur de l'échantillon qui présentent cette caractéristique et en additionnant les poids inscrits dans ces enregistrements.

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

7.6 Suppression de renseignements confidentiels

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

Le fichier maître de données de l'enquête comprend l'âge exact des répondants, tandis que le FMGD ne comprend que des regroupements d'âges.

Pour certaines variables sensibles au risque d'identification d'individus, le FMGD peut avoir subi de la suppression locale, c'est-à-dire que certaines valeurs présentes sur le fichier maître peuvent avoir été codées comme « non déclarées » sur le FMGD.

8.0 Qualité des données

8.1 Taux de réponse

Le tableau ci-dessous présente les comptes pour l'enquête téléphonique au niveau des provinces.

Comptes de l'enquête téléphonique

Région	Total	Envoi sur le terrain	Résolution sur le terrain	Estimation du nombre dans le champ de l'enquête	Ménages répondants	Personnes répondantes
Provinces de l'Atlantique	10 722	8 910	8 200	6 291	4 631	4 170
Québec	25 465	23 020	21 505	19 157	12 264	11 113
Ontario	49 534	43 910	40 698	33 989	20 085	17 961
Manitoba	7 767	6 721	6 062	5 382	3 498	3 189
Saskatchewan	9 671	8 277	7 464	6 948	4 620	4 270
Alberta	12 389	11 750	10 630	9 825	6 036	5 505
Colombie-Britannique	16 517	15 927	14 816	13 458	7 734	6 942
Canada	132 065	118 515	109 375	95 051	58 868	53 150

Les colonnes du tableau ci-dessus sont définies comme suit :

Total

Nombre total de numéros de téléphone sélectionnés dans l'échantillon composition aléatoire (CA) initial.

Envoi sur le terrain

Nombre de numéros de téléphone envoyés sur le terrain pour la collecte. La différence entre les colonnes « Total » et « Envoi sur le terrain » est le nombre de numéros de téléphone qu'on a retirés de l'échantillon après avoir effectué un filtrage des numéros hors service avant la collecte.

Résolution sur le terrain

Nombre de numéros de téléphone confirmés dans le champ de l'enquête (résidentiels) ou hors champ (numéros d'entreprise ou hors service) pendant la collecte.

Estimation du nombre de numéros dans le champ de l'enquête

$$= \sum_{\text{Strate}} \left(\frac{\text{Nombre de numéros dans le champ}}{\text{Nombre de numéros résolus sur le terrain}} \times \text{Nombre de numéros envoyés sur le terrain} \right)$$

Autrement dit, on a calculé au sein de chaque strate la proportion de numéros de téléphone résolus confirmés dans le champ de l'enquête et on a appliqué la même proportion aux numéros non résolus.

Réponse des ménages

Nombre de cas pour lesquels on dispose d'une liste complète des membres du ménage.

Réponse des personnes

Nombre de cas où le membre du ménage sélectionné pour participer à l'enquête a fourni suffisamment de données utilisables pendant l'interview téléphonique pour être considéré comme un répondant.

Le tableau ci-dessous présente les taux de réponse fondés sur les chiffres ci-dessus.

Taux de l'enquête téléphonique

Région	Taux de résolution global (%)	Taux de résolution sur le terrain (%)	Taux de succès global (%)	Taux de succès sur le terrain (%)	Taux de réponse des ménages (%)	Taux de réponse des personnes (%)	Taux de réponse global (%)
Provinces de l'Atlantique	93,4	92,0	58,7	70,6	73,6	90,0	66,3
Québec	94,1	93,4	75,2	83,2	64,0	90,6	58,0
Ontario	93,5	92,7	68,6	77,4	59,1	89,4	52,8
Manitoba	91,5	90,2	69,3	80,1	65,0	91,2	59,2
Saskatchewan	91,6	90,2	71,8	83,9	66,5	92,4	61,5
Alberta	91,0	90,5	79,3	83,6	61,4	91,2	56,0
Colombie-Britannique	93,3	93,0	81,5	84,5	57,5	89,8	51,6
Canada	93,1	92,3	72,0	80,2	61,9	90,3	55,9

Les taux du tableau ci-dessus ont été calculés comme suit :

$$\text{Taux de résolution global} = \frac{\text{Numéros éliminés} + \text{Numéros résolus sur le terrain}}{\text{Taille totale de l'échantillon}}$$

$$\text{Taux de résolution sur le terrain} = \frac{\text{Numéros résolus sur le terrain}}{\text{Numéros envoyés sur le terrain}}$$

$$\text{Taux de succès global} = \frac{\text{Estimation du nombre dans le champ}}{\text{Taille totale de l'échantillon}}$$

$$\text{Taux de succès sur le terrain} = \frac{\text{Estimation du nombre dans le champ}}{\text{Numéros envoyés sur le terrain}}$$

$$\text{Taux de réponse des ménages} = \frac{\text{Ménages répondants}}{\text{Estimation du nombre dans le champ}}$$

$$\text{Taux de réponse des personnes} = \frac{\text{Personnes répondantes}}{\text{Ménages répondants}}$$

$$\text{Taux de réponse global} = \frac{\text{Personnes répondantes}}{\text{Estimation du nombre dans le champ}}$$

Le tableau ci-dessous présente les comptes et les taux pour le questionnaire à retourner par la poste.

Comptes et taux de l'enquête postale

Région	Voyageurs (téléphone)	Envoi (papier)	Envoi en pourcentage (%)	Réponse (papier)	Taux de réponse (%)	Taux de réponse global (%)
Provinces de l'Atlantique	3 495	3 361	96,2	1 806	53,7	51,7
Québec	9 294	8 850	95,2	4 975	56,2	53,5
Ontario	15 569	15 253	98,0	8 153	53,5	52,4
Manitoba	2 683	2 599	96,9	1 561	60,1	58,2
Saskatchewan	3 786	3 598	95,0	2 064	57,4	54,5
Alberta	5 060	4 737	93,6	2 663	56,2	52,6
Colombie-Britannique	6 256	5 881	94,0	3 470	59,0	55,5
Canada	46 143	44 279	96,0	24 692	55,8	53,5

Les colonnes du tableau ci-dessus ont été définies comme suit :

Voyageurs (téléphone)

Nombre de voyageurs ayant répondu à l'enquête téléphonique.

Envoi (papier)

Nombre de questionnaires papier envoyés par la poste.

$$\text{Envoi en pourcentage} = \frac{\text{Envoi (papier)}}{\text{Voyageurs (téléphone)}}$$

Réponse (papier)

Nombre de questionnaires papier reçus avec suffisamment de données utilisables.

$$\text{Taux de réponse} = \frac{\text{Réponse (papier)}}{\text{Envoi (papier)}}$$

$$\text{Taux de réponse global} = \frac{\text{Réponse (papier)}}{\text{Voyageurs (téléphone)}}$$

Le tableau ci-dessous présente les comptes par régions métropolitaines de recensement (RMR).

Comptes par régions métropolitaines de recensement

Région métropolitaine de recensement	Réponse (téléphone)	Voyageurs (téléphone)	Non-voyageurs	Réponse (poste)
Halifax	1 408	1 231	177	648
Atlantique (autres)	2 762	2 264	498	1 158
Québec	2 072	1 786	286	990
Montréal	3 642	3 103	539	1 565
Gatineau	2 063	1 756	307	979
Québec (autres)	3 336	2 649	687	1 441
Ottawa	2 015	1 807	208	1 031
Kingston	437	387	50	222
Oshawa	399	342	57	180
Toronto	5 902	5 078	824	2 410
Hamilton	1 553	1 330	223	728
St.Catharines-Niagara	507	418	89	222
Kitchener	1 271	1 103	168	595
London	1 460	1 295	165	712
Windsor	431	372	59	181
Grand Sudbury	496	438	58	230
Thunder Bay	313	265	48	156
Ontario (autres)	3 177	2 734	443	1 486
Winnipeg	1 709	1 421	288	849
Manitoba (autres)	1 480	1 262	218	712
Regina	1 373	1 223	150	681
Saskatoon	1 386	1 227	159	694
Saskatchewan (autres)	1 511	1 336	175	689
Calgary	1 757	1 625	132	849
Edmonton	1 863	1 695	168	896
Alberta (autres)	1 885	1 740	145	918
Vancouver	2 996	2 671	325	1 384
Victoria	1 543	1 393	150	869
Colombie-Britannique (autres)	2 403	2 192	211	1 217
Canada	53 150	46 143	7 007	24 692

Le fichier de microdonnées à grande diffusion de l'Enquête sur les activités et les préférences en matière de voyages contient les 7 007 non-voyageurs qui ont répondu à l'enquête téléphonique et les 24 692 voyageurs qui ont répondu à l'enquête postale, pour un total de 31 699 enregistrements. Après avoir répondu à l'enquête postale, il y avait 609 répondants qui ont eu leur statut changer de voyageurs à non-voyageurs.

8.2 Erreurs relatives à l'enquête

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

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

Sur un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations calculées à partir de l'enquête. Toutefois, les erreurs systématiques contribuent à biaiser les estimations de l'enquête. Énormément de temps et d'efforts ont été consacrés à réduire les erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été prises à chacune des étapes du cycle de collecte et de traitement des données afin de contrôler la qualité des données. Ces mesures comprenaient le recours à des intervieweurs hautement qualifiés, une formation poussée des intervieweurs concernant les procédures d'enquête et du questionnaire, l'observation des intervieweurs en vue de cerner les problèmes liés à la conception du questionnaire ou à une mauvaise compréhension des instructions, des procédures visant à s'assurer que les erreurs de saisie des données étaient réduites au minimum ainsi que des vérifications de la qualité du codage et de contrôle ayant pour but d'attester la logique du traitement.

8.2.1 Couverture de la base de sondage

Comme nous l'avons mentionné à la section 5.1 (Population cible), moins de 6 % des ménages canadiens n'ont pas de ligne téléphonique terrestre. Les personnes qui vivent dans ces ménages peuvent présenter certaines caractéristiques dont les estimations d'enquête ne tiennent pas compte. Les utilisateurs doivent faire preuve de prudence lorsqu'ils analysent des sous-groupes de population présentant des caractéristiques qui peuvent être en corrélation avec l'absence de téléphone ou la possession d'un téléphone cellulaire seulement.

8.2.2 Collecte des données

La formation des intervieweurs pour l'Enquête sur les activités et les préférences en matière de voyages (EAPV) consistait à lire le Manuel du surveillant, le Manuel des procédures, le Manuel de l'intervieweur, la pratique avec des cas simulés de l'EAPV sur ordinateur et une discussion avec l'intervieweur principal des questions de l'intervieweur avant le début de l'enquête. L'information fournie aux intervieweurs comprenait une description du contexte et des objectifs de l'enquête ainsi qu'un glossaire de terminologie et une série de questions et de réponses. Les intervieweurs ont recueilli l'information pour l'EAPV après avoir recueilli l'information de l'EPA. La collecte des données téléphoniques de l'EAPV a eu lieu de janvier à avril 2006.

Dans l'enquête téléphonique et sur le questionnaire postal, on posait une question concernant les destinations de voyage au cours des deux dernières années (question TS_Q02 pour l'enquête téléphonique et question A01, colonne A, pour le questionnaire papier). Il y avait des incohérences dans la formulation des catégories de réponse (c.-à-d. *Europe incluant le Royaume-Uni et la Russie* sur le questionnaire papier, mais seulement *Europe incluant le Royaume-Uni* dans l'enquête téléphonique), ainsi que des catégories manquantes (c.-à-d. l'option *Autre(s) pays* existe dans l'enquête téléphonique, mais pas sur le questionnaire papier).

Les données des répondants présentaient également des incohérences entre les deux sources de données : on a comparé les réponses des répondants ayant déclaré des données à la question TS_Q02 de l'enquête téléphonique et à la question A01, colonne A, du questionnaire papier. Les comptes non pondérés sont présentés dans le tableau ci-dessous.

Par exemple, le nombre de répondants qui ont déclaré avoir voyagé dans leur province en réponse aux deux questions est de 17 630. Parmi ceux qui ont déclaré avoir voyagé dans leur province lors de l'enquête téléphonique, 10,6 % (2 083 / (17 630 + 2 083)) ont déclaré qu'ils n'avaient pas voyagé dans leur province sur le questionnaire papier. Parmi ceux qui ont déclaré avoir voyagé dans leur province sur le questionnaire papier, 12,1 % (2 438 / (17 630 + 2 438)) ont déclaré qu'ils n'avaient pas voyagé dans leur province lors de l'interview téléphonique.

Questions TS_Q02 (interview téléphonique) et A01, colonne A (questionnaire papier)

	TS_Q02 = Oui et A01 = Oui	TS_Q02 = Oui et A01 = Non		TS_Q02 = Non et A01 = Oui		TS_Q02 = Non et A01 = Non
		Nombre de répondants	%	Nombre de répondants	%	
Votre province	17 630	2 083	10,6	2 438	12,1	2 535
Autre province	12 751	820	6,0	2 994	19,0	8 121
États-Unis	10 333	716	6,5	2 063	16,6	11 574
Mexique	2 197	301	12,0	732	25,0	21 456
Amérique du Sud et Amérique Centrale	731	476	39,4	337	31,6	23 142
Caraïbes	2 555	416	14,0	953	27,2	20 762
Europe	3 258	461	12,4	485	13,0	20 482
Asie	853	156	15,5	290	25,4	23 387

Les incohérences entre les renseignements déclarés lors de l'interview téléphonique et sur le questionnaire papier indiquent clairement une erreur non due à l'échantillonnage. De plus, pour toutes les destinations sauf l'Amérique du Sud et l'Amérique Centrale, les comptes sont plus élevés sur le questionnaire papier que dans le cas de l'interview téléphonique; on peut en déduire que les répondants se rappellent un plus grand nombre de destinations lorsqu'ils remplissent un questionnaire que lorsqu'ils sont interviewés par téléphone.

Un autre aspect de la qualité des données lié à la collecte des données est le suivant : certaines variables ont été recueillies lors de l'interview téléphonique auprès de non-voyageurs et sur le questionnaire papier destiné aux voyageurs. Les estimations concernant ces variables reposent donc sur des données provenant de deux modes de collecte différents. Les utilisateurs de données doivent être conscients qu'à l'égard de ces variables, les différences dans le mode de collecte risquent d'avoir un impact non négligeable sur les analyses comparatives des voyageurs et des non-voyageurs.

8.2.3 Traitement des données

Le traitement des données de l'enquête téléphonique a été relativement simple, car les données ont été saisies au moyen de l'application d'une interview téléphonique assistée par ordinateur (ITAO), dans laquelle on avait programmé des vérifications et des cheminements pour améliorer la cohérence des données saisies. Le traitement des données s'est avéré beaucoup plus complexe dans le cas du questionnaire papier.

L'une des tâches liées au traitement des données du questionnaire papier consistait à établir des règles pour déterminer s'il fallait attribuer aux espaces vides du questionnaire la valeur « Non » ou « Non déclaré ». La section qui a posé la plus grande difficulté était la section A, questions A03 à A17, car elle ne comporte pas de catégorie « Aucune de ces réponses ». On a utilisé la règle suivante : s'il y a un « Oui » n'importe où dans les questions A03 à A17 ou si des données sont déclarées aux questions A01, A18, B01, B03 et C01, on convertit tous les

espaces vides des questions A03 à A17 à « Non ». L'interprétation erronée des espaces vides est une source éventuelle d'erreur non due à l'échantillonnage.

Les questions de type « cochez une seule réponse » du questionnaire papier ont causé des difficultés de traitement, certains répondants les ayant prises pour des questions de type « Cochez toutes les réponses qui s'appliquent » et ayant déclaré des réponses multiples. En général, le système de traitement a converti les réponses multiples aux questions de type « cochez une seule réponse » en réponses « Non déclaré ». Les questions A18, E10 et E12 ont présenté les taux les plus élevés de réponses multiples, soit, respectivement, 5,8 %, 8,6 % et 7,7 %.

Certaines données des répondants au questionnaire papier ne respectaient pas les directives concernant le cheminement du questionnaire. En général, le système de traitement a corrigé les incohérences de cheminement selon une approche descendante : on a supposé que les variables déclarées en premier dans le questionnaire étaient correctes et l'on a modifié les autres variables en conséquence.

8.2.4 Non-réponse

La non-réponse totale peut constituer une source importante d'erreur non due à l'échantillonnage dans de nombreuses enquêtes, selon la mesure dans laquelle les répondants et les non-répondants diffèrent à l'égard des caractéristiques d'intérêt. S'il y a eu non-réponse totale, c'est parce que l'intervieweur a été incapable de communiquer avec le répondant, que le répondant a refusé de participer à l'enquête ou qu'il n'a pas fourni suffisamment de données utilisables. Les non-réponses totales ont été traitées en ajustant le poids des personnes ou ménages qui ont répondu au questionnaire d'enquête de façon à compenser pour ceux qui n'y ont pas répondu.

Dans la plupart des cas, il y a eu non-réponse partielle au questionnaire d'enquête lorsque le répondant n'a pas compris ou a mal interprété une question, a refusé d'y répondre ou ne pouvait se rappeler l'information demandée. Dans le fichier de microdonnées, la non-réponse partielle est indiquée par des codes (Refus, Ne sait pas). Comme nous l'avons mentionné à la section 7.3, on a procédé à une imputation par donneur pour imputer la région métropolitaine de recensement dans le cas de 6 % des enregistrements.

8.2.5 Mesure de l'erreur d'échantillonnage

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

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

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

Par exemple, supposons que, d'après les résultats de l'enquête, l'on estime que 39,0 % des

personnes du Manitoba ont voyagé en Ontario au cours des deux dernières années et l'on constate que l'erreur-type de cette estimation est de 0,0155. Le coefficient de variation de l'estimation est donc calculé comme suit :

$$\left(\frac{0,0155}{0,39} \right) \times 100 \% = 4,0 \%$$

De plus amples renseignements sur le calcul du coefficient de variation, se trouvent au chapitre 10.0.

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

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

9.1 Lignes directrices pour l'arrondissement d'estimations

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

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

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

Le plan d'échantillonnage utilisé pour l'Enquête sur les activités et les préférences en matière de voyages (EAPV) n'était pas autopondéré. Lorsqu'ils produisent des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent appliquer le poids d'échantillonnage approprié.

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

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

9.3 Estimations catégoriques

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

Les estimations catégoriques sont des estimations du nombre ou du pourcentage de membres de la population visée par l'enquête possédant certaines caractéristiques ou faisant partie d'une catégorie définie. Le nombre de personnes du Manitoba ayant voyagé en Ontario au cours des deux dernières années au moment de l'enquête, ou la proportion de Manitobains ayant voyagé en Ontario constituent des exemples de telles estimations. Une estimation du nombre de personnes possédant une certaine caractéristique peut aussi être désignée une estimation d'un agrégat. La grande majorité des questions de l'EAPV étaient de nature nominale et, dans le cas contraire, elles ont été regroupées de manière à devenir des variables nominales.

Exemples de questions catégoriques :

Q : Au cours des 2 dernières années, avez-vous effectué des voyages d'une nuit ou plus à l'extérieur de la ville pour quelque raison que ce soit? Incluez les voyages d'une nuit ou plus que vous avez effectués dans une cabine, un chalet ou une maison de villégiature qui vous appartient ou qui appartient à un ami ou à un membre de votre famille.

R : Oui / Non

Q : Combien de voyages d'agrément ou de vacances d'une nuit ou plus à l'extérieur de la ville avez-vous effectués au cours des deux dernières années?

R : Aucun / Un / Deux / Trois / Quatre / Cinq ou plus

9.3.1 Totalisation d'estimations catégoriques

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

- a) additionnant les poids finals des enregistrements présentant la caractéristique qui nous intéresse pour le numérateur (\hat{X}),

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

9.4 Lignes directrices pour l'analyse statistique

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

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

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

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

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

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

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

9.5 Lignes directrices pour la diffusion de coefficients de variation

Avant de diffuser et/ou de publier toutes estimations établies à partir de l'EAPV, les utilisateurs devraient premièrement déterminer le niveau de qualité de cette estimation. Les niveaux de qualité sont *acceptable*, *médiocre* et *inacceptable*. Les erreurs d'échantillonnage et non dues à l'échantillonnage, dont il a été question au chapitre 8.0, influencent la qualité des données. Aux fins du présent document, cependant, on ne déterminera le niveau de qualité d'une estimation qu'à partir d'une erreur d'échantillonnage dont rend compte le coefficient de variation indiqué à l'intérieur du tableau qui figure ci-dessous. Les utilisateurs devraient néanmoins s'assurer de lire le chapitre 8.0 pour être plus pleinement informés des caractéristiques relatives à la qualité de ces données.

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

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

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

Lignes directrices relatives au niveau de qualité de l'estimation

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

9.6 Seuils pour la diffusion des estimations pour l'Enquête sur les activités et les préférences en matière de voyages

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

Par exemple, d'après le tableau, la qualité d'une estimation pondérée de 20 000 personnes possédant une caractéristique donnée dans les provinces de l'Atlantique est médiocre.

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

Région	CV acceptable 0,0 à 16,5 %	CV médiocre 16,6 à 33,3 %	CV inacceptable > 33,3 %
Provinces de l'Atlantique	54 100 et plus	13 600 à < 54 100	moins de 13 600
Québec	81 300 et plus	20 200 à < 81 300	moins de 20 200
Ontario	81 400 et plus	20 100 à < 81 400	moins de 20 100
Manitoba	33 700 et plus	8 500 à < 33 700	moins de 8 500
Saskatchewan	22 100 et plus	5 600 à < 22 100	moins de 5 600
Alberta	93 000 et plus	23 500 à < 93 000	moins de 23 500
Colombie-Britannique	125 400 et plus	31 700 à < 125 400	moins de 31 700
Canada	83 700 et plus	20 600 à < 83 700	moins de 20 600

Région métropolitaine de recensement	CV acceptable 0,0 à 16,5 %	CV médiocre 16,6 à 33,3 %	CV inacceptable > 33,3 %
Halifax	24 400 et plus	6 400 à < 24 400	moins de 6 400
Atlantique (autres)	60 000 et plus	15 200 à < 60 000	moins de 15 200
Québec	37 600 et plus	9 700 à < 37 600	moins de 9 700
Montréal	101 400 et plus	25 600 à < 101 400	moins de 25 600
Gatineau	17 600 et plus	4 600 à < 17 600	moins de 4 600
Québec (autres)	67 800 et plus	17 000 à < 67 800	moins de 17 000
Ottawa	36 000 et plus	9 200 à < 36 000	moins de 9 200
Kingston	31 800 et plus	9 800 à < 31 800	moins de 9 800
Oshawa	56 600 et plus	16 600 à < 56 600	moins de 16 600
Toronto	111 200 et plus	27 900 à < 111 200	moins de 27 900
Hamilton	46 600 et plus	12 200 à < 46 600	moins de 12 200
St.Catharines-Niagara	47 800 et plus	13 300 à < 47 800	moins de 13 300
Kitchener	33 300 et plus	8 800 à < 33 300	moins de 8 800
London	28 900 et plus	7 600 à < 28 900	moins de 7 600
Windsor	72 900 et plus	22 700 à < 72 900	moins de 22 700
Grand Sudbury	21 300 et plus	6 000 à < 21 300	moins de 6 000
Thmoins de Bay	24 000 et plus	7 200 à < 24 000	moins de 7 200
Ontario (autres)	85 100 et plus	21 500 à < 85 100	moins de 21 500
Winnipeg	35 800 et plus	9 300 à < 35 800	moins de 9 300
Manitoba (autres)	28 700 et plus	7 600 à < 28 700	moins de 7 600
Regina	15 700 et plus	4 200 à < 15 700	moins de 4 200
Saskatoon	19 400 et plus	5 200 à < 19 400	moins de 5 200
Saskatchewan (autres)	26 600 et plus	6 900 à < 26 600	moins de 6 900
Calgary	83 600 et plus	22 200 à < 83 600	moins de 22 200
Edmonton	83 000 et plus	22 200 à < 83 000	moins de 22 200
Alberta (autres)	97 500 et plus	26 200 à < 97 500	moins de 26 200
Vancouver	155 100 et plus	40 800 à < 155 100	moins de 40 800
Victoria	52 500 et plus	15 200 à < 52 500	moins de 15 200
Colombie-Britannique (autres)	93 400 et plus	24 300 à < 93 400	moins de 24 300
Canada	83 700 et plus	20 600 à < 83 700	moins de 20 600

10.0 Tables de variabilité d'échantillonnage approximative

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

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

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

Région	Effet du plan	Taille de l'échantillon	Population
Provinces de l'Atlantique	2,07	2 481	1 822 494
Québec	2,57	6 794	5 940 869
Ontario	2,44	10 545	9 671 592
Manitoba	2,35	2 067	843 107
Saskatchewan	2,25	2 548	706 325
Alberta	3,32	3 108	2 465 540
Colombie-Britannique	4,44	4 156	3 326 176
Canada	2,93	31 699	24 776 103

Région métropolitaine de recensement	Effet du plan	Taille de l'échantillon	Population
Halifax	2,01	825	297 463
Atlantique (autres)	1,85	1 656	1 525 031
Québec	2,51	1 276	559 702
Montréal	2,10	2 104	2 868 546
Gatineau	3,04	1 286	222 059
Québec (autres)	1,77	2 128	2 290 562
Ottawa	1,92	1 239	668 949
Kingston	2,77	272	117 138
Oshawa	1,76	237	264 517
Toronto	2,43	3 234	4 145 248
Hamilton	2,38	951	554 599
St.Catharines-Niagara	1,58	311	304 046
Kitchener	2,16	763	354 307
London	2,11	877	356 789
Windsor	2,57	240	258 365
Grand Sudbury	1,64	288	123 661
Thunder Bay	1,82	204	97 506
Ontario (autres)	1,91	1 929	2 426 467
Winnipeg	2,23	1 137	534 034
Manitoba (autres)	2,60	930	309 073
Regina	2,63	831	151 156
Saskatoon	2,85	853	177 889
Saskatchewan (autres)	1,79	864	377 280
Calgary	3,04	981	818 632
Edmonton	3,48	1 064	775 066
Alberta (autres)	3,65	1 063	871 842
Vancouver	4,49	1 709	1 764 224
Victoria	7,18*	1 019	256 193
Colombie-Britannique (autres)	3,00	1 428	1 305 759
Canada	2,93	31 699	24 776 103

* Dans le cas de la RMR (région métropolitaine de recensement) de Victoria, l'effet du plan de sondage est élevé parce que le taux de réponse dans la strate des numéros de téléphone « inconnus » était particulièrement faible : sur 480 numéros de téléphone de l'échantillon composition aléatoire initial, il n'y avait que 21 répondants.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Règle 4 : Estimations de rapports

Dans le cas où le numérateur est un sous-ensemble du dénominateur, le rapport devrait être converti en un pourcentage et la règle 2 appliquée. Cela s'appliquerait, par exemple, au cas où le dénominateur est le nombre de voyageurs et le numérateur, le nombre de voyageurs qui se rendent au Québec.

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

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

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

Règle 5 : Estimations de différences entre des rapports

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

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

Les exemples ci-dessous utilisent des données du fichier de l'EAPV de 2006 et sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter.

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

Supposons qu'un utilisateur estime que 1 919 960 de Québécois ont voyagé en Ontario au cours des deux dernières années. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Reportez-vous à la table des coefficients de variation pour le QUÉBEC.
- 2) L'agrégat estimé (1 919 960) ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, c'est-à-dire 2 000 000.
- 3) On trouve le coefficient de variation pour un agrégat estimé en se reportant à la première entrée autre que des astérisques sur cette rangée, c'est-à-dire 2,7 %.
- 4) Le coefficient de variation approximatif de l'estimation est donc 2,7 %. Le résultat selon lequel il y avait 1 919 960 de Québécois (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1) qui ont voyagé en Ontario au cours des deux dernières années, peut être publié sans réserve.

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

Supposons qu'un utilisateur estime que $16\,666 / 48\,100 = 34,6\%$ des femmes de Halifax âgées de 55 ans et plus regardent habituellement des émissions sportives à la télévision. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Reportez-vous à la table des coefficients de variation pour HALIFAX.
- 2) Parce que l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c'est-à-dire les femmes de Halifax âgées de 55 ans et plus), il faut utiliser à la fois le pourcentage (34,6 %) et la portion numérateur du pourcentage (16 666) pour déterminer le coefficient de variation.
- 3) Le numérateur, 16 666, ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, soit 17 000. De même, l'estimation du pourcentage ne figure dans l'en-tête d'aucune colonne; il faut donc utiliser la proportion qui s'en rapproche le plus, soit 35,0 %.
- 4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 16,6 %, est le coefficient de variation à employer.
- 5) Le coefficient de variation approximatif de l'estimation est donc 16,6 %. Le résultat selon lequel 34,6 % des femmes de Halifax âgées de 55 ans et plus regardent habituellement des émissions sportives à la télévision est considéré comme marginal. Ces estimations devraient être signalées par la lettre M (ou un autre identificateur semblable) et être accompagnées d'un avertissement mettant les

utilisateurs subséquents en garde contre les hauts taux d'erreur associés à l'estimation.

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

Supposons qu'un utilisateur estime que $152\,841 / 282\,871 = 54,0\%$ des hommes d'Edmonton ayant voyagé au cours des deux dernières années ont déclaré qu'il y avait bien des bonnes raisons de se rendre à Hawaï, alors que le pourcentage était de $234\,191 / 377\,196 = 62,1\%$ chez les femmes d'Edmonton. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

- 1) L'utilisation de la table des coefficients de variation d'EDMONTON de la même façon que celle décrite dans l'exemple 2, donne un CV de l'estimation pour les hommes de 9,2 % et un CV de l'estimation pour les femmes de 5,5 %.
- 2) En utilisant la règle 3, l'erreur-type d'une différence ($\hat{d} = \hat{X}_1 - \hat{X}_2$) est :

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

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

C'est-à-dire que l'erreur-type de la différence $\hat{d} = 0,540 - 0,621 = -0,081$ est :

$$\begin{aligned} \sigma_{\hat{d}} &= \sqrt{[(0,540)(0,092)]^2 + [(0,621)(0,055)]^2} \\ &= \sqrt{(0,002468) + (0,001167)} \\ &= 0,0603 \end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,0603 / 0,081 = 0,744$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 74,4 %. 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 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 4 : Estimations de rapports

Supposons qu'un utilisateur estime que 1 207 129 de femmes lisent des magazines consacrés aux affaires, aux finances et aux investissements pendant un mois normal, par rapport à 4 541 127 de femmes qui lisent des magazines de mode et d'esthétique pendant un mois normal. L'utilisateur est intéressé à comparer l'estimation sous la forme d'un rapport. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation d'un rapport, où le numérateur de l'estimation (\hat{X}_1) est le nombre de femmes qui lisent des magazines consacrés aux

affaires, aux finances et aux investissements pendant un mois normal. Le dénominateur de l'estimation (\hat{X}_2) est le nombre de femmes qui lisent des magazines de mode et d'esthétique pendant un mois normal.

- 2) Reportez-vous à la table des coefficients de variation pour le CANADA.
- 3) Le numérateur de cette estimation de rapport est 1 207 129. Le chiffre qui s'en rapproche le plus est 1 000 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 4,7 %.
- 4) Le dénominateur de cette estimation de rapport est 4 541 127. Le chiffre qui s'en rapproche le plus est 5 000 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 1,9 %.
- 5) Le coefficient de variation approximatif de l'estimation du rapport est donc donné par la règle 4, qui est :

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

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement.

C'est-à-dire que :

$$\begin{aligned} \alpha_{\hat{R}} &= \sqrt{(0,047)^2 + (0,019)^2} \\ &= \sqrt{0,00221 + 0,00036} \\ &= 0,051 \end{aligned}$$

- 6) Le rapport obtenu entre les femmes qui lisent des magazines consacrés aux affaires, aux finances et aux investissements pendant un mois normal et les femmes qui lisent des magazines de mode et d'esthétique pendant un mois normal est $1\,207\,129 / 4\,541\,127 = 0,266$ (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1). Le coefficient de variation de cette estimation est 5,1 %, ce qui fait qu'on peut la diffuser sans réserve.

Exemple 5 : Estimations de différences de rapports

Supposons qu'un utilisateur estime que le rapport entre les femmes qui lisent des magazines consacrés aux affaires, aux finances et aux investissements pendant un mois normal et les femmes qui lisent des magazines de mode et d'esthétique pendant un mois normal est $1\,207\,129 / 4\,541\,127 = 0,266$. Supposons qu'un utilisateur estime que le même rapport entre les hommes est $2\,443\,090 / 929\,580 = 2,63$. L'utilisateur est intéressé à comparer les deux rapports pour voir s'il y a une différence statistique entre ceux-ci. Comment l'utilisateur détermine-t-il le coefficient de variation de cette différence?

- 1) Tout d'abord calculez le coefficient de variation approximatif pour le rapport des femmes (\hat{R}_1) et le rapport des hommes (\hat{R}_2) tel qu'il est décrit dans l'exemple 4. Le CV approximatif pour le rapport des femmes est 5,1 % et 5,7 % pour celui des hommes.

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

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

où α_1 et α_2 sont les coefficients de variation de \hat{R}_1 et \hat{R}_2 respectivement. C'est-à-dire que l'erreur-type de la différence $\hat{d} = 0,266 - 2,63 = -2,364$ est :

$$\begin{aligned} \sigma_{\hat{d}} &= \sqrt{[(0,266)(0,051)]^2 + [(2,63)(0,057)]^2} \\ &= \sqrt{(0,000184) + (0,022473)} \\ &= 0,151 \end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,151 / (-2,364) = -0,064$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 6,4 %, ce qui fait qu'on peut la diffuser sans réserve.

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

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

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

En utilisant l'erreur-type d'une estimation, des intervalles de confiance pour des estimations peuvent être obtenues en partant de l'hypothèse qu'aux termes d'un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique donnée de la population se répartiront normalement autour de la valeur vraie de la population. Selon cette hypothèse, il y a environ 68 chances sur 100 que l'écart entre une estimation de l'échantillon et la valeur vraie pour la population soit inférieur à une erreur-type, environ 95 chances sur 100 que l'écart soit inférieur à deux erreurs-types et environ 99 chances sur 100 que l'écart soit inférieur à trois erreurs-types. Ces différents degrés de confiance sont désignés sous le nom de niveaux de confiance.

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

Des intervalles de confiance pour une estimation peuvent être calculés directement à partir des tables de variabilité d'échantillonnage approximative, en déterminant d'abord à

partir de la table appropriée le coefficient de variation de l'estimation \hat{X} , puis en utilisant la formule suivante pour le convertir à un intervalle de confiance ($IC_{\hat{x}}$) :

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

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

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

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

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

Un intervalle de confiance de 95 % pour la proportion estimée de femmes de Halifax âgées de 55 ans et plus qui regardent habituellement des émissions sportives à la télévision (d'après l'exemple 2 à la section 10.1.1) serait calculé comme suit :

$$\hat{X} = 34,6 \% \text{ (ou exprimé sous forme de proportion } 0,346)$$

$$t = 2$$

$$\alpha_{\hat{x}} = 16,6 \% \text{ (0,166 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.}$$

$$IC_{\hat{x}} = \{0,346 - (2) (0,346) (0,166), 0,346 + (2) (0,346) (0,166)\}$$

$$IC_{\hat{x}} = \{0,346 - 0,115, 0,346 + 0,115\}$$

$$IC_{\hat{x}} = \{0,231, 0,461\}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 23,1 % et 46,1 % des femmes de Halifax âgées de 55 ans et plus regardent habituellement des émissions sportives à la télévision.

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

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

elle sont identiques.

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

Si $t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ se situe entre -2 et 2, aucune conclusion à propos de la différence entre les caractéristiques n'est alors justifiée au niveau de signification de 5 %. Si, cependant, ce rapport est inférieur à -2 ou supérieur à +2, la différence observée est significative au niveau de 0,05. C'est-à-dire que la différence entre les estimations est significative.

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

Supposons que l'utilisateur désire tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y a pas de différence entre la proportion de d'hommes d'Edmonton ayant déclaré qu'il y avait bien des bonnes raisons de se rendre à Hawaï et la proportion de femmes. D'après l'exemple 3 à la section 10.1.1, il s'est avéré que l'erreur-type de la différence entre ces deux estimations était 0,0603. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0,540 - 0,621}{0,0603} = \frac{-0,081}{0,0603} = -1,34$$

Puisque $t = -1,34$ est entre 2 et -2 aucune conclusion à propos de la différence entre les deux estimations n'est alors justifiée au niveau de signification de 0,05 %.

10.4 Coefficients de variation pour des estimations quantitatives

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

En général cependant, le coefficient de variation d'un total quantitatif sera supérieur au coefficient de variation de l'estimation de la catégorie correspondante (c'est-à-dire l'estimation du nombre de personnes retenues dans l'estimation quantitative). S'il est impossible de diffuser l'estimation de la catégorie correspondante, on ne pourra pas non plus communiquer l'estimation quantitative. Si, par conséquent, le coefficient de variation de la proportion est inacceptable (rendant la proportion non diffusable), il en sera de même du coefficient de variation de l'estimation quantitative correspondante (rendant cette estimation quantitative non diffusable). Des estimations de la variance pour des variables précises peuvent être obtenues auprès de Statistique Canada, contre remboursement des frais.

10.5 Tables des coefficients de variation

Consulter le fichier EAPV2006_CVTabSF.pdf pour les tables de coefficient de variation pour le l'Enquête sur les activités et les préférences en matière de voyages de 2006.

11.0 Pondération

Le présent chapitre explique comment on a dérivé les poids (coefficients de pondération) pour l'Enquête sur les activités et les préférences en matière de voyages (EAPV). On a d'abord procédé à la pondération pour l'enquête téléphonique, puis pour le questionnaire papier à retourner par la poste.

11.1 Procédures de pondération pour l'enquête téléphonique

1. Calculer les poids déterminés par le plan d'échantillonnage

À chacun des 132 065 numéros de téléphone de l'échantillon, on a attribué un poids W_1 déterminé par le plan d'échantillonnage, égal à l'inverse de sa probabilité de sélection et calculé comme suit au sein de chaque strate :

$$W_1 = \left(\frac{\text{Nombre total de numéros échantillonnés possibles}}{\text{Nombre de numéros échantillonnés}} \right)$$

2. Éliminer les numéros de téléphone éliminés après sélection

On a retiré de la pondération les numéros de téléphone identifiés comme des numéros hors service avant la collecte des données. Il y en avait 13 550.

3. Ajustement pour les numéros de téléphone non résolus

Il y a eu 9 140 numéros de téléphone qui n'ont été résolus à l'étape de la collecte des données, c.-à-d. des numéros qui n'étaient ni dans le champ de l'enquête (numéros résidentiels) ni hors champ (numéros d'entreprise ou numéros hors service). On a ajusté les poids par strate, à l'intérieur de chaque province, comme suit :

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

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

Les numéros de téléphone identifiés comme des numéros d'entreprise, des numéros hors service ou hors du champ de l'enquête, tels de chalet, ont été éliminés. Il y en avait 19 341.

5. Ajustement pour les ménages non répondants

Il y avait 31 166 enregistrements confirmés dans le champ de l'enquête (numéros résidentiels) mais pour lesquels on ne disposait pas de la liste des membres du ménage. On a ajusté les poids par strate, à l'intérieur de chaque province, comme suit :

$$W_3 = \left(\frac{\sum W_2 \text{ pour les ménages répondants} + \sum W_2 \text{ pour les ménages non répondants}}{\sum W_2 \text{ pour les ménages répondants}} \right) \times W_2$$

6. Ajustement pour la sélection d'un membre du ménage

Pendant la collecte, on a sélectionné au hasard un membre du ménage âgé de 18 ans et plus pour participer à l'enquête. On a calculé les poids comme suit :

$$W_4 = \text{Nombre de membres du ménage admissible} \times W_3$$

7. Ajustement pour les personnes non répondantes

Il y avait 5 311 enregistrements comportant une liste complète des membres du ménage, mais où la personne sélectionnée pour participer à l'enquête n'a pas répondu (aucune donnée n'était déclarée en réponse à la première question, TS_Q01). On a modélisé la propension à répondre au moyen d'un modèle de régression logistique à l'intérieur de chaque région (Provinces de l'Atlantique, Québec, Ontario, Manitoba, Saskatchewan, Alberta, Colombie-Britannique), avec les variables explicatives sexe et groupe d'âge (18 à 24 ans, 25 à 34 ans, 35 à 44 ans, 45 à 54 ans, 55 à 64 ans, 65 ans et plus) du répondant sélectionné, taille du ménage (une, deux, trois, quatre personnes et plus), présence d'enfants dans le ménage (oui ou non), type de ligne téléphonique initial (numéro résidentiel ou inconnu) et langue de l'interview (français ou anglais). On a formé des groupes de non-réponse à l'intérieur de chaque région en fonction de la propension à répondre. Au sein de chaque groupe de non-réponse, on a ajusté les poids comme suit :

$$W_5 = \left(\frac{\sum W_4 \text{ pour les personnes répondante} + \sum W_4 \text{ pour les personnes non répondante}}{\sum W_4 \text{ pour les personnes répondante}} \right) \times W_4$$

8. Ajustement pour les enregistrements aux données insuffisantes

Il y avait 407 enregistrements comportant certaines données, mais pas suffisamment pour être utilisables. On a calculé des facteurs de correction à l'intérieur de chaque région (Provinces de l'Atlantique, Québec, Ontario, Manitoba, Saskatchewan, Alberta, Colombie-Britannique) à la question TS_Q01 afin de préserver la distribution du nombre de voyageurs. Au sein de chaque groupe de pondération, On a ajusté les poids comme suit :

$$W_6 = \left(\frac{\sum W_5 \text{ pour les enregistrements aux données suffisantes} + \sum W_5 \text{ pour les enregistrements aux données insuffisantes}}{\sum W_5 \text{ pour les enregistrements aux données suffisantes}} \right) \times W_5$$

9. Ajustement pour le nombre de lignes téléphoniques dans le ménage

On a ajusté les poids à la baisse dans le cas des ménages possédant plus d'une ligne téléphonique (avec des numéros différents) pour tenir compte du fait que ces ménages ont une plus grande probabilité d'être sélectionnés. On a divisé le poids de chaque répondant par le nombre de lignes résidentielles distinctes (jusqu'à concurrence de quatre) du ménage. Si le nombre de lignes était inconnu, on a imputé une valeur de un. On a calculé les poids comme suit :

$$W_7 = \left(\frac{W_6}{\text{Nombre de lignes téléphoniques du ménage}} \right)$$

10. Caler en fonction des totaux externes

On a procédé au calage en utilisant les chiffres de population de février 2006. On a utilisé deux ensembles de chiffres de population : ceux des personnes âgées de 18 ans et plus au niveau des régions métropolitaines de recensement (RMR) et ceux des groupes d'âge (18 à 24 ans, 25 à 34

ans, 35 à 44 ans, 45 à 54 ans, 55 à 64 ans, 65 ans et plus) par sexe au niveau des régions (Provinces de l'Atlantique, Québec, Ontario, Manitoba, Saskatchewan, Alberta, Colombie-Britannique). On a utilisé l'estimation par régression généralisée (GREG) pour caler les poids de manière que la somme des poids égale les chiffres de population dans les deux cas. Les poids calés sont désignés W_8 .

Les poids des 7 007 non-voyageurs ont été dérivés selon les étapes 1 à 10 de pondération.

11.2 Procédures de pondération pour l'enquête postale

On a dérivé un ensemble de poids distinct pour le questionnaire papier à retourner par la poste. On a utilisé les données de l'enquête téléphonique pour déterminer quelles variables expliquaient le mieux la non-réponse au questionnaire papier. On a procédé aux étapes suivantes :

11. Ajustement pour la non-réponse au questionnaire papier

On a corrigé la pondération de la non-réponse pour tenir compte des 21 451 voyageurs qui n'ont pas rempli le questionnaire papier. On a calculé des facteurs de correction à l'intérieur de chaque région (Provinces de l'Atlantique, Québec, Ontario, Manitoba, Saskatchewan, Alberta, Colombie-Britannique) par groupe d'âge (18 à 24 ans, 25 à 34 ans, 35 à 44 ans, 45 à 54 ans, 55 à 64 ans, 65 ans et plus), par sexe et par niveau de scolarité (études secondaires au plus, études universitaires, autres études postsecondaires). On a supprimé le niveau de scolarité pour les cellules de moins de 20 répondants au questionnaire papier ou de moins de 30 répondants à l'interview téléphonique. Au sein de chaque groupe de pondération, on a ajusté les poids comme suit :

$$W_9 = \left(\frac{\sum W_7 \text{ pour tous les voyageurs ayant répondu à l'interview}}{\sum W_7 \text{ pour tous les voyageurs ayant rempli le questionnaire}} \right) \times W_7$$

12. Caler en fonction de l'enquête téléphonique

On a utilisé les données des 46 143 voyageurs ayant répondu à l'interview téléphonique pour calculer des totaux de contrôle à l'aide du poids W_8 obtenu à l'étape 10 de la pondération. On a calculé les ensembles de totaux de contrôle suivants (à l'échelle régionale, sauf indication contraire) :

- nombre de voyageurs par groupe d'âge (18 à 24 ans, 25 à 34 ans, 35 à 44 ans, 45 à 54 ans, 55 à 64 ans, 65 ans et plus) par sexe;
- nombre de voyageurs âgés de 18 ans et plus au niveau des RMR;
- nombre de voyageurs nés au Canada (d'après la question DM_Q09);
- nombre de voyageurs dans les ménages comptant un, deux ou trois adultes et plus (d'après la liste);
- nombre de personnes ayant voyagé dans leur province (TS_Q02A);
- nombre de personnes ayant voyagé dans une autre province ou un autre territoire canadien (TS_Q02B);
- nombre de personnes ayant voyagé aux États-Unis (TS_Q02C);
- nombre de personnes ayant voyagé ailleurs (TS_Q02D à TS_Q02K).

On a calé les poids des 24 692 répondants à l'enquête postale de manière à produire les mêmes estimations que ces totaux de contrôle (encore une fois à l'aide de l'estimation GREG). On a imputé temporairement les enregistrements comportant des données manquantes dans les variables de calage. Les poids calés sont désignés W_{10} .

Il est à noter qu'on a calé les poids de manière que, par exemple, les estimations relatives à la question TS_Q02C (et non A01_Q14A) utilisant les 24 692 répondants à l'enquête postale et les poids W_{10} , soient égales aux estimations relatives à la question TS_Q02C utilisant les 46 143 voyageurs ayant répondu à l'enquête téléphonique et les poids W_8 . On n'a pas utilisé les questions A01_Q01A à A01_Q21B du questionnaire papier dans le calage parce qu'il y avait trop d'incohérences entre les données de l'interview téléphonique et celles du questionnaire papier, et l'utilisation des deux sources de données aurait pu entraîner des biais dans les poids.

Les poids des 24 692 voyageurs ont été dérivés selon les étapes 1 à 12 de pondération.

12.0 Questionnaires

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) :

EAPV2006_P_QuestF.pdf (Questionnaire postal)

EAPV2006_T_QuestF.pdf (Questionnaire téléphonique)

13.0 Cliché d'enregistrement à valeurs univariées

Consulter le fichier EAPV2006_LvCds.pdf pour le cliché d'enregistrement à chiffres univariés.