

Guide de l'utilisateur des microdonnées

Enquête nationale sur les voyages

2020



Statistique
Canada

Statistics
Canada

Canada

Comment obtenir d'autres renseignements

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

Statistique Canada
150 Promenade Tunney's Pasture, Ottawa (Ontario) K1A 0T6
Téléphone : 1-800-263-1136
Courriel : STATCAN.infostats-infostats.STATCAN@canada.ca

Comment accéder aux produits ou les commander

L'Enquête nationale sur les voyages (ENV) produit deux types de fichiers de microdonnées : les fichiers maîtres et les fichiers de microdonnées à grande diffusion (FMGD).

Fichiers maîtres

Les fichiers maîtres contiennent toutes les variables et tous les enregistrements de l'enquête recueillis au cours d'une période de collecte. Ces fichiers sont accessibles à Statistique Canada pour usage interne, dans les Centres de données de recherche (CDR) de Statistique Canada et peuvent aussi faire l'objet de demandes de totalisations personnalisées.

Centre de données de recherche

Le Programme des CDR permet aux chercheurs d'utiliser les données d'enquête contenues dans les fichiers maîtres dans un environnement sécuritaire situé dans plusieurs universités de tout le Canada. Les chercheurs doivent soumettre des propositions de recherche qui, une fois acceptées, leur donneront accès aux CDR. Pour plus de renseignements, consultez la page Web suivante : <http://www.statcan.gc.ca/rdc-cdr/index-fra.htm>

Totalisations personnalisées

Une autre méthode d'accès aux fichiers maîtres consiste à offrir à tous les utilisateurs de faire appel au personnel du Service à la clientèle du **Centre canadien de la statistique du tourisme et des transports** pour produire des totalisations personnalisées. Ce service est offert en recouvrement des coûts. Il permet aux utilisateurs qui ne savent pas se servir de logiciels de totalisation d'obtenir des résultats personnalisés. Le service filtre les résultats avant de les diffuser pour les rendre conformes aux normes de confidentialité et de fiabilité. Pour plus de renseignements, communiquez avec Statistique Canada.

Fichiers de microdonnées à grande diffusion

Les fichiers de microdonnées à grande diffusion sont élaborés à partir des fichiers maîtres suivant une technique qui vise à concilier l'impératif de confidentialité des répondants et la nécessité de produire des données d'utilité maximale. Les FMGD doivent répondre à des normes sévères de sécurité et de confidentialité, conformément à la *Loi sur la statistique*, avant d'être diffusés. Pour s'assurer du respect de ces normes, un comité formé de hauts gestionnaires de Statistique Canada soumet chaque FMGD à un processus officiel d'examen et d'approbation. Les variables les plus susceptibles de permettre l'identification d'une personne sont supprimées du fichier ou agrégées en catégories moins détaillées.

Veuillez noter que des différences entre les estimations diffusées par Statistique Canada dans ses produits statistiques et les estimations reproduites par les utilisateurs à l'aide des fichiers de microdonnées à grande diffusion pourraient être observées, les estimations diffusées par Statistique Canada étant produites à l'aide des fichiers maîtres.

Pour commander un exemplaire de FMGD, communiquez avec Statistique Canada à l'adresse mentionnée ci-haut.

L'Initiative de démocratisation des données

Le Programme de l'Initiative de démocratisation des données (IDD) permet aux étudiants et chercheurs d'utiliser les fichiers de microdonnées à grande diffusion dans plusieurs universités de tout le Canada. Pour plus de renseignements, consultez la page web suivante : <http://www.statcan.gc.ca/fra/idd/idd>

L'impact de la pandémie de COVID-19 sur le FMGD ENV 2020

En raison de la pandémie de COVID-19, la collecte de l'Enquête nationale sur les voyages a été temporairement suspendue en 2020, ce qui a entraîné l'absence de données disponibles pour les mois de mars, avril, mai et juin. **Ce FMGD ne couvrant donc que huit mois, il ne permet pas de calculer des totaux annuels.** De plus, en raison du faible nombre de répondants ayant déclaré des voyages à l'étranger aux troisième et quatrième trimestres, toutes les destinations de voyage pour ces trimestres ont été codées comme « inconnue ». Ce n'est qu'au premier trimestre, qui ne couvre que deux mois (janvier et février), que les noms des pays individuels sont publiés.

Table des matières

1.0	Description de l'enquête	5
2.0	Concepts et définitions	6
2.1	Concepts et définitions de l'Enquête nationale sur les voyages	6
2.2	Développement du contenu	13
3.0	Méthodologie d'enquête	13
3.1	Population cible et population de l'enquête	13
3.2	Plan d'échantillonnage	13
4.0	Collecte des données	14
5.0	Traitement des données	14
5.1	Saisie des données	14
5.2	Vérification	15
5.3	Codage des questions ouvertes	17
5.4	Création de variables dérivées	17
5.5	Imputation	17
5.6	Contrôle de la divulgation	18
6.0	Qualité des données	18
6.1	Erreurs non dues à l'échantillonnage	18
6.2	Erreurs d'échantillonnage	20
7.0	Pondération	21
7.1	Procédure de pondération pour l'ENV	21
8.0	Lignes directrices pour la totalisation, l'analyse et la diffusion de données	22
8.1	Lignes directrices pour l'arrondissement d'estimations	22
Annexe A – Estimation de la variance pour les fichiers maîtres		23
Annexe B – Estimation de la variance pour les fichiers de microdonnées à grande diffusion		24
Annexe C – Différences entre l'Enquête sur les voyages des résidents du Canada et l'Enquête nationale sur les voyages		34

1.0 Description de l'enquête

L'Enquête nationale sur les voyages (ENV) a été élaborée pour remplacer entièrement l'Enquête sur les voyages des résidents du Canada (EVRC numéro d'enregistrement 3810) et pour remplacer la composante des résidents canadiens de l'Enquête sur les voyages internationaux (EVI numéro d'enregistrement 3152). L'Enquête nationale sur les voyages recueille des renseignements concernant les voyages intérieurs et internationaux des résidents du Canada.

L'Enquête nationale sur les voyages fournit des statistiques sur les activités des résidents du Canada en ce qui a trait au tourisme intérieur et international. Elle a été élaborée pour mesurer le volume des activités, les caractéristiques et les répercussions économiques du tourisme. Pour le Système de la comptabilité nationale, l'enquête mesure l'ampleur des voyages intérieurs au Canada, du point de vue de la demande.

Les principaux utilisateurs des données de l'ENV sont Statistique Canada, Destination Canada, les provinces et les bureaux de tourisme. Les médias, les entreprises, les consultants de même que les chercheurs comptent parmi les autres utilisateurs.

Ce guide a été rédigé pour faciliter l'utilisation du FMGD de l'enquête.

Toute question au sujet de l'ensemble de données ou de son utilisation devrait être adressée à :

Statistique Canada
150 Promenade Tunney's Pasture, Ottawa (Ontario) K1A 0T6
Téléphone : 1-800-263-1136
Courriel : STATCAN.infostats-infostats.STATCAN@canada.ca

2.0 Concepts et définitions

2.1 Concepts et définitions de l'Enquête nationale sur les voyages

Cette section explique les concepts et définitions reliés au sujet d'intérêt. Les utilisateurs peuvent également se référer au questionnaire de l'ENV et aux livres de codes inclus dans la trousse de diffusion de l'enquête. Les livres de codes présentent l'information complète sur chacune des variables sur le fichier de microdonnées à grande diffusion. L'information fournie pour chaque variable est : nom de la variable, la description ou la définition, la liste des codes avec leur description ou l'étendue de la valeur que la variable peut avoir ainsi que leurs fréquences.

2.1.1 Types de voyages couverts par l'enquête

Types de voyages

Les voyages couverts par l'ENV sont :

- les voyages intérieurs (origine et destination au Canada) sans composante internationale
- les voyages intérieurs (origine et destination au Canada) avec une composante internationale
- les voyages internationaux (origine au Canada et destination à l'extérieur du Canada) avec une composante intérieure
- les voyages internationaux (origine au Canada et destination à l'extérieur du Canada) sans composante intérieure

Les voyages au Canada peuvent être considérés comme des entités avec des composantes provinciales. Un voyage peut couvrir une ou plusieurs provinces. Les voyages intérieurs peuvent donc être classifiés comme suit:

- voyages intérieurs (origine et destination au Canada) :
 - intra-provinciaux
 - interprovinciaux
- voyages intérieurs (origine et destination au Canada) avec une composante internationale :
 - intra-provinciaux
 - interprovinciaux

Durée du voyage : de même jour ou avec nuitée(s)

La durée d'un voyage, définie comme « de même jour » ou « avec nuitée(s) », est déterminée en considérant le nombre total de nuitées au cours du voyage, que les nuitées aient été au Canada ou à l'extérieur du Canada.

Les voyages sont classés comme-suit sur le FMGD :

Classification des voyages
1 International de même jour
2 International avec nuitées
3 Intérieur de même jour - interprovincial
4 Intérieur avec nuitées - interprovincial
5 Intérieur de même jour – intra-provincial
6 Intérieur avec nuitées – intra-provincial

Composantes d'un voyage : visites

Lors de l'analyse des visites, les voyages intérieurs, les voyages internationaux et les composantes extérieures et intérieures des différents voyages sont inclus.

Les voyages sont divisés en trois types de composantes : l'origine du voyage, les endroits visités au cours du voyage et la destination principale du voyage (ou point de sortie du Canada pour les voyages internationaux). L'origine et la destination principale du voyage sont toujours incluses comme composantes et sont uniques. Les endroits visités au cours du voyage sont inclus si au moins une nuitée est passée dans ces endroits. Il y aura autant de composantes « endroits visités » qu'il y a d'endroits visités avec nuitées (les endroits visités excluent la destination principale). Des dépenses peuvent être assignées à tous les types de composantes. Par contre, les nuitées et les types d'hébergements ne peuvent être assignés qu'aux endroits visités et à la destination principale du voyage s'il y a lieu.

Des composantes aéroports sont ajoutées si le répondant rapporte avoir utilisé un avion commercial pour une destination canadienne et que les frais de transport commerciaux sont alloués à un transporteur canadien. Il y aura une composante aéroport pour chaque composante origine, endroit visité, et destination. Des dépenses en transport sont assignées aux composantes aéroports mais aucune nuitée et type d'hébergement ne le sont.

Le lieu de chaque composante d'un voyage est codé géographiquement au niveau du pays (Canada ou autre pays), province, région touristique (RT), région métropolitaine de recensement (RMR), division de recensement (DR) et sous-division de recensement (SDR). Au cours d'un voyage, on compte une visite à un certain niveau géographique si au moins un endroit visité ou la destination principale sont rapportés dans cette géographie. L'origine et les aéroports ne comptent pas comme des visites. Par exemple, un voyage intérieur constituera une – et une seule – visite-province au Québec si au moins un endroit visité au Québec est rapporté ou la destination principale est au Québec. Par défaut, tous les voyages intérieurs inclus dans l'analyse des visites comptent comme une visite-Canada puisqu'ils ont tous au moins un endroit visité au Canada ou la destination principale rapportée est au Canada. Ce concept de *visite-géographie* s'applique à tous les niveaux géographiques de l'enquête : pays, province, regroupement de provinces (provinces maritimes, territoires), région touristique, région métropolitaine de recensement et division de recensement.

Selon ce concept de visite, une visite à un certain niveau géographique peut compter pour plus d'une visite à un niveau géographique plus petit. Par exemple, une visite-province pourrait représenter deux visites-région touristique si des endroits ont été visités dans deux des régions touristiques de cette province. L'exemple suivant illustre quelques cas au niveau province et région touristique. Les mêmes concepts s'appliquent aux RMR et DR également.

Durée de la visite : de même jour ou avec nuitée(s)

Pour déterminer la durée de la visite, le nombre de nuitées passées au cours de la visite est utilisé, peu importe le niveau géographique de la visite analysé (pays, province, région touristique, etc.). S'il y a des nuitées passées au cours d'une visite, alors la durée de la visite est avec nuitées, sinon la durée est de même jour.

Dépenses associées aux voyages

Pour chaque voyage rapporté, les dépenses encourues au cours du voyage sont collectées, et ce, pour différents postes de dépenses (par exemple, les frais d'hébergement commercial, de location d'un véhicule, de transport local). Au niveau des voyages, les dépenses rapportées sont comptabilisées à la destination principale du voyage. Les dépenses rapportées pour un voyage sont également réallouées aux différentes composantes du voyage (origine, visite, destination ou aéroport) par poste de dépense. Certains types de dépenses ne peuvent être alloués à certaines composantes.

2.1.2 Principales définitions

Activités : Activités entreprises par les *voyageurs* au cours du *voyage*, et non de leur *visite*. Ainsi, on ne peut pas toujours associer une activité à un endroit précis. Par exemple, une personne peut avoir effectué au cours du même voyage une visite à Vancouver et une visite à Whistler et déclarer avoir fait du ski alpin comme activité. Il est impossible de savoir avec certitude si le voyageur a fait du ski alpin à Vancouver, à Whistler ou aux deux endroits. Dans le cas d'un *voyage de même jour*, il s'agira de l'activité la plus importante réalisée pendant le voyage et pour un *voyage d'une nuit ou plus*, il s'agira de toutes les activités réalisées pendant le voyage.

Destination (principale) : Lieu déclaré par le voyageur à titre de destination d'un *voyage*. Si un *voyageur* visite plus d'un lieu au cours d'un *voyage*, la destination principale est le lieu donné par le *voyageur* lorsqu'il répond à cette question.

Dépenses (déclarées) : Les dépenses déclarées sont les dépenses engagées par le voyageur et tous les membres du ménage ayant effectué le *voyage* ainsi que les dépenses engagées par d'autres personnes qui n'ont pas pris part au voyage. Les dépenses des membres d'autres ménages qui ont participé au voyage sont exclues. Les dépenses sont déclarées en dollars canadiens et incluent toutes les taxes et tous les pourboires. Elles sont associées à l'ensemble du voyage.

Les dépenses déclarées ne comprennent pas les articles suivants : les aliments achetés avant le voyage et devant être consommés pendant le voyage, les articles achetés en vue de leur revente ou de leur utilisation à des fins commerciales (y compris les articles utilisés sur des fermes), les véhicules, notamment les voitures, roulottes, embarcations, les dépenses en immobilisations, notamment les biens immobiliers, œuvres d'art, articles rares et actions, l'argent comptant donné à des parents ou à des amis au cours d'un voyage ne servant pas à payer des biens ou des services touristiques, ainsi que les dons faits à des institutions.

Catégories des dépenses déclarées :

- **Dépenses d'hébergement** : Dépenses totales pour les nuitées passées dans un hôtel, un motel, un lieu de villégiature, une cabine, un chalet loué ou commercial, un terrain de camping, etc. au cours du voyage. Ces dépenses visent principalement les frais de location. Toutefois, dans le cas de voyages dans des chalets privés ou chez des parents ou des amis, elles peuvent également comprendre toute somme versée aux propriétaires pour l'utilisation des installations d'hébergement.
- **Dépenses de location de véhicule** : Montant des dépenses relatives aux frais d'utilisation, de location, de tarification versés pour l'utilisation de tout véhicule tel une automobile, un camion, une motocyclette, une bicyclette, une embarcation, une autocaravane, une motoneige, etc. (incluant les assurances) au cours du voyage. Le montant déclaré comprend toutes les taxes et tous les pourboires applicables.
- **Dépenses de fonctionnement de véhicule** : Montant des dépenses relatives au fonctionnement d'un véhicule privé ou loué, notamment l'essence, les réparations et les frais de stationnement au cours du voyage. Le montant déclaré comprend toutes les taxes et tous les pourboires applicables.
- **Dépenses de transport local** : Montant des dépenses de transport au sein d'une ville ou d'une région métropolitaine, c'est-à-dire de transport interurbain, pendant le voyage. Sont compris sous cette rubrique les frais de taxi, d'autobus, de métro et d'autres dépenses comme celles visant les excursions en autocar au cours du voyage, dans le lieu de visite et/ou de destination. Le montant déclaré comprend toutes les taxes et tous les pourboires applicables.
- **Dépenses de transport commercial** : Montant des dépenses de transport entre différentes zones urbaines et/ou rurales à partir de la destination ou jusqu'à celle-ci pendant le voyage. Ces dépenses de transport couvrent le prix des billets pour des déplacements interurbains par avion, par bateau, par aéroglisseur, par train, par autocar et par traversier. Le montant déclaré comprend toutes les taxes et tous les pourboires applicables.

▪ **Dépenses liées aux aliments et boissons achetés dans les restaurants ou les bars :** Montant des dépenses au titre des repas et des boissons achetés dans des restaurants, des bars, des cafétérias, des comptoirs de restauration rapide et de mets à emporter, et des minibars que l'on trouve dans certaines chambres d'hôtel ou de motel, peu importe le lieu où ils ont été consommés au cours du voyage. Le montant déclaré comprend toutes les taxes et tous les pourboires applicables.

▪ **Dépenses liées aux aliments et boissons achetés dans des magasins pendant le voyage :** Montant dépensé pour des aliments et boissons achetés dans des magasins locaux au cours du voyage, peu importe le lieu où ils sont consommés. Par exemple, les produits alimentaires achetés en vue d'être ramenés chez soi ou consommés pendant le voyage sont inclus dans cette catégorie. Le montant déclaré comprend toutes les taxes et tous les pourboires applicables.

▪ **Dépenses de loisirs :** Montant dépensé au titre des activités de loisirs ou sportives et de la location d'équipement pour de telles activités au cours du voyage. Sont compris dans cette catégorie les droits d'entrée sur les terrains de golf, les billets de remonte-pente, la location de chaussures de quilles et la location d'équipement sportif tel que des skis, des bâtons de golf et des canots. Sont exclus les achats d'abonnements visant des activités de loisirs (p. ex. les billets de remonte-pente pour la saison). Le montant déclaré comprend toutes les taxes et tous les pourboires applicables.

▪ **Dépenses de divertissements :** Somme d'argent dépensée pour les activités culturelles ou de divertissement et pour les attractions au cours du voyage. Cette catégorie comprend les dépenses pour les droits d'entrée au théâtre, à des galeries d'art, à des boîtes de nuit ou événements sportifs (p. ex. une partie de hockey). Elle comprend également les dépenses liées au jeu de hasard, le prix d'entrée pour des activités culturelles ou de loisirs (p. ex. cinémas, musées, parcs thématiques, pistes de karting), les excursions en bateau et les vols en montgolfière. Il faut exclure les coûts liés à l'achat d'abonnements à des activités de loisirs ou des spectacles (p. ex. un abonnement au théâtre). Le montant déclaré comprend toutes les taxes et tous les pourboires applicables.

▪ **Dépenses d'achat de vêtements :** Sont compris tous les achats de vêtements, de chaussures ou d'accessoires, qu'il s'agisse de cadeaux ou d'articles personnels. Le montant déclaré comprend toutes les taxes et tous les pourboires applicables.

▪ **Autres dépenses :** Montant des dépenses au titre d'autres articles au cours du voyage. Sont compris sous cette rubrique les souvenirs, les articles ménagers, les droits d'inscription à des cours, des conventions ou des congrès, les droits de douane, les cartes postales et les timbres, les frais d'assurance, les médicaments, les livres, le matériel d'artiste et les pellicules pour appareils-photo, les frais de garde d'enfants, de garde de domicile et de garde d'animaux, les frais de téléphone, les frais de location d'hébergement (p. ex. une salle de séminaire, une salle de formation, etc.), l'achat d'abonnement à des activités de loisirs ou des spectacles (p. ex. les billets de remonte-pente pour la saison, les abonnements au théâtre). Le montant déclaré comprend toutes les taxes et tous les pourboires applicables.

Dépenses réallouées : Renvoie au procédé par lequel toutes les dépenses déclarées par les voyageurs sont réparties de nouveau aux régions géographiques particulières où les sommes ont été dépensées. L'ENV fait appel à un modèle de réallocation des dépenses qui permet d'attribuer les sommes de chacune des catégories de dépenses aux régions géographiques conformément à des règles précises. Les niveaux géographiques en fonction desquels les dépenses sont réallouées dans le cadre de l'ENV sont les suivants : pays, province, région touristique, région métropolitaine de recensement et division de recensement.

Distance du voyage : Terme désignant la distance (aller simple), exprimée en kilomètres, entre l'origine et la destination du voyage.

Durée : Un voyage débute lorsque le *voyageur* quitte sa résidence habituelle et se termine lorsqu'il y revient. La durée du voyage est calculée selon le nombre de nuitées que le *voyageur* passe en dehors de sa résidence habituelle.

Forfaits : Les forfaits comprennent les montants dépensés pour deux éléments ou plus d'un *voyage*, mais dans la plupart des cas les coûts ne sont pas indiqués de manière distincte. Ces éléments peuvent être une association de *transport* et d'*hébergement* ou un *hébergement* et des *aliments* et *boissons*, etc. Les forfaits sont généralement achetés par l'entremise d'une agence de voyages ou d'une organisation sociale.

Hébergement : Type d'hébergement où le voyageur a passé la nuit pendant le *voyage*. Les données sur le type d'hébergement sont recueillies pour chacune des nuitées du *voyage*. Cela comprend l'hébergement commercial (hôtel, motel, lieu de villégiature, terrain de camping, etc.) et sans frais (domicile d'un ami ou d'un parent, chalet particulier, autre type d'hébergement sans frais).

Mois de référence : Mois au cours duquel le *voyage* s'est terminé, entre le premier et le dernier jour du mois civil. Le *voyage* peut avoir commencé avant le mois de référence.

Moyen de transport principal : Moyen de transport utilisé pour parcourir la distance la plus longue au cours du *voyage*. Si le répondant indique qu'il a utilisé deux moyens de transport pour parcourir des distances équivalentes, c'est le premier moyen de transport utilisé qui est enregistré.

Nuitée-personne : Nuit passée à l'extérieur du domicile par une personne en *voyage*. Lorsque deux personnes en voyage passent trois nuitées à l'extérieur de chez elles, on compte six nuitées-personne.

Nuitée-visite-personne : Lorsque la personne qui entreprend un *voyage* passe une nuit au Canada ou un autre pays à l'extérieur du foyer (par exemple, on compte trois nuitées-visites-personnes lorsqu'une personne entreprend un *voyage* de trois nuitées à l'extérieur du foyer). Le nombre total de nuitées-visites-personnes correspond au nombre de nuitées passées à l'extérieur du foyer au **Canada ou dans un autre pays** par chaque personne ayant fait ces *voyages*. Les niveaux géographiques selon lesquels les nuitées-visites-personnes peuvent être estimées dans l'ENV sont les suivants : pays, province, région touristique, région métropolitaine de recensement (RMR) et division de recensement (DR).

Organisation mondiale du tourisme des Nations Unies (OMTNU) : L'Organisation mondiale du tourisme (OMT) est un organisme spécialisé des Nations Unies et la principale institution internationale dans le domaine du tourisme. Elle sert de forum mondial pour les questions de politiques du tourisme et représente une source concrète de savoir-faire en matière de tourisme.

Elle regroupe 156 pays, 6 territoires et quelque 500 membres affiliés représentant le secteur privé, des établissements d'enseignement, des associations de professionnels du tourisme et des autorités touristiques locales.

Origine : L'origine correspond au point de départ d'un *voyage*. Le voyage doit avoir débuté dans l'une des dix provinces.

Raison principale : Raison principale pour laquelle le voyageur a entrepris le *voyage*, sans tenir compte des raisons qu'ont les autres personnes du ménage pour participer au même voyage. Il s'agit de la raison sans laquelle le voyage n'aurait pas eu lieu.

Région touristique : Les frontières des régions touristiques (RT) sont définies par chaque ministère du Tourisme provincial et non par Statistique Canada. Chaque année, ces ministères ont la possibilité de modifier leurs frontières des RT, que Statistique Canada applique ensuite à l'ENV. Les régions touristiques sont créées à partir de la SDR.

Revenu du ménage : Revenu total du ménage provenant des salaires et traitements, pourboires, commissions, pensions, intérêts, loyers, etc., avant impôt et déductions, pour toutes les personnes habitant le logement, pour l'année précédant l'année de référence.

Tourisme : La définition utilisée est celle adoptée par l'Organisation mondiale du tourisme et la Commission statistique des Nations Unies, et se lit comme suit : « les activités déployées par les personnes au cours de leurs voyages et de leurs séjours dans les lieux situés en dehors de leur environnement habituel pour une période consécutive qui ne dépasse pas une année, à des fins de loisirs, pour affaires et autres motifs. »

Trimestre : Trimestre pendant lequel le voyage s'est terminé. Le premier trimestre va de janvier à mars, le deuxième trimestre va d'avril à juin, le troisième trimestre va de juillet à septembre et le quatrième trimestre va d'octobre à décembre.

Visite : Lieu de séjour au Canada ou un autre pays. Il s'agit du lieu où le voyageur a passé au moins une nuit dans le cas d'une visite avec nuitée ou de la destination du voyage dans le cas d'une visite de même jour. Si le répondant s'est rendu dans le même lieu à deux reprises au cours d'un même voyage, on ne compte qu'une seule visite du lieu en question.

Visite ou voyage de même jour : Sont déclarés voyages de même jour les voyages au cours desquels le voyageur a parcouru une distance de 40 kilomètres ou plus (pour l'aller) et est rentré à son domicile dans la journée.

Visite ou voyage interprovincial : Voyage ou visite effectué dans une province différente de la province d'origine du voyage.

Visite ou voyage intra-provincial : Voyage ou visite effectué au sein de la province d'origine du voyage.

Visite-pays : Chaque personne qui fait une visite dans un pays, le Canada ou un autre, est enregistrée comme ayant fait une *visite-personne* au Canada.

Visite-personne : Visite effectuée par une *personne voyageant* seule ou en groupe. Une *visite* effectuée par quatre personnes voyageant ensemble constitue quatre visites-personnes.

Visite-province : Toutes les personnes faisant une visite dans une province sont enregistrées comme ayant fait une *visite-personne* dans cette province. Veuillez noter que l'on compte deux visites-provinces lorsqu'un voyageur visite deux provinces au cours du même voyage, mais une seule *visite-pays*.

Voyage : Un voyage doit avoir *débuté* dans l'une des dix provinces; il doit s'être terminé au cours du *mois de référence* et durer moins de 365 jours/nuitées. L'ENV collecte tous les *voyages d'une nuit ou plus* dont la destination se situe au Canada ou dans un autre pays, tous les *voyages de même jour* d'une distance de 40 km et plus avec une destination au Canada ou à l'international. Ces voyages sont considérés comme « à l'extérieur de la ville » par le voyageur.

Les voyages qui font partie du champ de l'enquête comprennent :

- tous les *voyages* aux fins de loisirs ou de vacances;
- tous les *voyages* effectués pour rendre visite à des amis ou des parents;
- tous les *voyages* d'affaires ou liés au travail, **à l'exception** des déplacements courants faisant partie intégrante d'un emploi;
- tous les *voyages* effectués pour d'autres raisons, **à l'exception** des voyages dont la raison principale est l'achat régulier d'épicerie ou d'articles ménagers, un déménagement dans une nouvelle résidence ou école (ou aider un ami à déménager), de se rendre à l'école, à un rendez-vous régulier chez le médecin ou chez le dentiste, est liée à l'observance ou aux services de pratiques religieuses habituelles, de se rendre à des funérailles et les *voyages* effectués dans le cadre de diverses corvées domestiques, comme aller chercher quelqu'un au stade.

Voyage intérieur ou visite intérieure : On considère comme *voyage* intérieur ou *visite* intérieure tout *voyage* ou toute *visite* dont la destination est au Canada.

Voyage international : Un voyage est considéré comme international lorsque la destination du voyage se situe à l'extérieur du Canada.

Voyage ou visite d'une nuit ou plus : Voyage ou visite pendant lesquels le répondant s'est absenté de son domicile au moins une nuitée.

Voyage-personne : Voyage effectué par une personne. Si cette personne fait plus d'un voyage et/ou voyage en compagnie d'autres membres adultes du même ménage, le nombre de voyages-personnes correspondra au nombre de voyages et de personnes ayant fait ces voyages. Si quatre membres d'un même ménage voyagent ensemble, on estime qu'il s'agit de quatre voyages-personnes. Deux voyages effectués par une même personne correspondent à deux voyages-personnes.

Voyageur : Toute personne âgée de 18 ans ou plus qui effectue un voyage. Toute personne qui n'effectue pas de voyage peut être désignée par le terme de non-voyageur.

2.1.3 Description des ensembles de données de l'ENV

Chaque mois, trois fichiers maîtres (ou ensembles) de données sont créés : un fichier personne, un fichier voyage et un fichier visite. Lorsque le FMGD est créé, les fichiers personnes, les fichiers voyages et les fichiers visites sont agglomérés ensemble pour créer des fichiers annuels pour chacun.

Fichier de données des personnes

L'ensemble de données (ou fichier) des personnes renferme les données sur les voyageurs et les non-voyageurs, les premiers étant les personnes de 18 ans et plus qui ont fait au moins un voyage s'étant terminé au cours du mois de référence, les seconds, celles qui n'ont pas fait de voyage s'étant terminé au cours du mois de référence. À chaque répondant à l'ENV correspond un enregistrement dans le fichier des personnes, que le répondant ait effectué ou non un voyage au cours du mois de référence ou de l'année de référence.

Le fichier de microdonnées des personnes contient des renseignements sociodémographiques de base touchant les voyageurs comme les non-voyageurs. Il permet d'établir des profils sociodémographiques simples et de calculer la fréquence des voyages. Il permet, par exemple, de comparer l'âge, le sexe ou le niveau de scolarité des voyageurs et des non-voyageurs, ou encore de déterminer le pourcentage de la population âgée de 55 ans et plus qui a voyagé.

Le fichier de microdonnées des personnes ne fournit pas de renseignements sur le nombre de voyages ou de voyages-personnes; il fournit plutôt des données sur le nombre de voyageurs et de non-voyageurs. Si une personne a fait plus d'un voyage au cours de la période de référence, elle ne sera considérée qu'une seule fois à titre de voyageur.

Fichier de données des voyages

L'ensemble de données des voyages renferme les caractéristiques des voyages, comme par exemple l'origine et la destination du voyage, la raison du voyage, les dépenses, etc.. Pour chaque répondant sur le fichier personne, il y a un enregistrement de voyage pour chaque voyage déclaré. Si un répondant n'a pas déclaré de voyage, il n'y a pas d'enregistrement de voyage pour ce répondant sur le fichier voyage.

Fichier de données des visites

L'ensemble de données des visites présente des renseignements sur les régions visitées par les voyageurs, que celles-ci aient constitué la destination principale ou un arrêt d'une nuit au cours du

voyage. Ce fichier des visites renferme, outre les renseignements sur les visites, des données sur les dépenses réallouées des ménages.

Le concept de visite sert à déterminer le nombre de visites-personnes pour un lieu donné. Le fichier des visites contient au moins deux enregistrements de visites effectuées dans le cadre de chacun des voyages-personnes apparaissant dans le fichier des voyages (origine et destination).

2.2 Développement du contenu

Le contenu du questionnaire électronique de l'Enquête nationale sur les voyages a été tiré de l'Enquête sur les voyages des résidents du Canada et de l'Enquête sur les voyages internationaux, qui étaient fondées sur des consultations auprès de plusieurs organismes/ministères provinciaux du tourisme. Les responsables du Système de comptabilité nationale de Statistique Canada ont participé à la conception du questionnaire.

Le questionnaire a fait l'objet d'un essai cognitif sous forme d'interviews approfondies, dans les deux langues officielles du Canada, menées par le Centre de ressources en conception de questionnaires de Statistique Canada. L'objectif de l'étude qualitative était de vérifier une nouvelle introduction à l'enquête et différentes définitions de voyages. Il y a eu deux enquêtes pilotes. La première enquête pilote menée en février-mars 2016 visait à évaluer la sélection des répondants au moyen de lettres. La méthode habituelle de sélection aléatoire consistait à sélectionner un ménage, puis à sélectionner un répondant au moyen de l'application. L'essai pilote a fourni des renseignements sur la capacité du membre du ménage à interpréter une des méthodes de sélection aléatoire décrites dans la lettre et à s'y conformer.

La deuxième enquête menée en août 2017 visait à évaluer l'application de réponse en ligne et à estimer le taux de participation. L'enquête pilote a aussi servi à évaluer plusieurs méthodes de suivi auprès des non-répondants, incluant l'envoi de lettres, des appels de courtoisie et des appels pour offrir de remplir le questionnaire au téléphone.

3.0 Méthodologie d'enquête

3.1 Population cible et population de l'enquête

La population cible est constituée de la population civile non institutionnalisée âgée de 18 ans et plus des dix provinces canadiennes. Sont exclues de façon particulière de l'enquête : les personnes vivant dans des réserves indiennes et les personnes vivant dans les territoires. Ensemble, ces groupes représentent une exclusion de moins de 3 % de la population canadienne âgée de 18 ans et plus.

3.2 Plan d'échantillonnage

L'Enquête nationale sur les voyages est une enquête transversale par échantillon. Un plan d'échantillonnage à trois degrés est utilisé : 1) Logement; 2) Personne; 3) Voyage.

Pour le premier degré,

- L'enquête se base sur le fichier de l'univers des logements (FUL) ainsi que sur le fichier de données socio-économiques (FSE). De cette base sont exclus les résidents du Yukon, des Territoires du Nord-Ouest et du Nunavut, les personnes vivant dans les réserves indiennes ainsi que les parties d'un logement collectif n'étant pas des LAAC (logements associés avec un collectif).
- La base est ensuite stratifiée selon la province, le groupe de revenus et le groupe de passeport. Les données des passeports sont utilisées pour stratifier la base de sondage et ainsi avoir plus

de voyageurs internationaux. Pour certains mois (Janvier, février, octobre, novembre et décembre), une variable additionnelle a été utilisée pour stratifier le groupe – « TC regions ». Cette variable réfère à des RMR spécifique ou des groupes de RMR. Dans chaque strate un échantillon aléatoire simple est tiré. L'échantillon global comporte environ 39 000 ménages mensuellement. Des ménages additionnels pourraient s'ajouter pour d'autres partenaires.

Pour le deuxième degré, un adulte par ménage sélectionné sera choisi aléatoirement selon un plan d'échantillonnage aléatoire simple. La méthode de sélection selon l'âge est utilisée avec l'aide d'une lettre postale. En utilisant cette méthode basée sur les lettres, la sélection aléatoire d'une personne sera faite avant l'accès en ligne, donc seulement la personne sélectionnée ira en ligne pour compléter le questionnaire électronique.

Pour le troisième degré, le questionnaire électronique demande au répondant une courte description de tous ses voyages se terminant dans le mois de référence. L'application sélectionne ensuite un maximum de trois voyages selon un plan d'échantillonnage séquentiel de Poisson.

4.0 Collecte des données

La période de référence pour la collecte des données de l'année 2020 était du 1^{er} janvier 2020 au 10 janvier 2021. Il s'agit d'une enquête à participation volontaire et les données sont obtenues directement auprès des répondants.

Les ménages sélectionnés reçoivent une lettre d'invitation par la poste. Cette lettre explique qui, du ménage, est sélectionné pour répondre à l'enquête en utilisant la méthode de sélection selon l'âge. Un ménage peut avoir jusqu'à deux rappels par la poste. Le code d'accès sur les lettres permet au répondant d'accéder au questionnaire électronique. Le questionnaire électronique est disponible dans les deux langues officielles, soit le français et l'anglais. Le répondant doit donner l'information de base sur tous ses voyages (domestiques et internationaux) se terminant dans le mois de référence. Ensuite il donne plus de détails sur les voyages sélectionnés. Le délai moyen requis pour répondre à l'enquête est de 15 minutes.

Les projections démographiques produites par la division de la démographie à Statistique Canada sont utilisées à l'étape de pondération de l'ENV. Les poids au niveau personne sont calés sur ces projections.

Les estimations provenant du programme de dénombrement à la frontière de Statistique Canada sont également utilisées à l'étape de pondération de l'ENV. Concrètement, les poids voyage-personne sont calés sur ces estimations de façon à ce que les deux sources de données soient cohérentes. De plus, ce calage a pour effet d'améliorer la qualité de des estimations de l'ENV en stabilisant l'estimateur du nombre de voyages internationaux faits par les Canadiens.

5.0 Traitement des données

Le traitement transforme les réponses du questionnaire obtenues pendant la collecte pour qu'elles conviennent à la totalisation et à l'analyse des données. Il comprend toutes les activités de traitement des données, automatisées et manuelles, après la collecte et avant l'estimation.

5.1 Saisie des données

Pour les questionnaires électroniques, les répondants saisissent directement les réponses aux questions de l'enquête. L'emploi d'un questionnaire électronique réduit les délais de traitement des données et les coûts associés à la saisie des données, aux erreurs de transcription et à la transmission des données. Les réponses sont sécurisées via des protocoles de cryptage standards de l'industrie, des pare-feu et des couches de clés d'encodages.

Un contrôle se fait au moment même que le questionnaire électronique est rempli par le répondant. 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, le répondant était invité, vis des messages apparaissant à l'écran de l'ordinateur, de vérifier les renseignements.

Cependant, pour certaines questions, le répondant a la possibilité de passer outre aux contrôles et de sauter des questions si il ou elle 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. Lorsque les données du questionnaire électronique sont reçues, elles sont converties en fichiers textes.

5.2 Vérification

Des vérifications peuvent être faites à plusieurs étapes pendant le processus de l'enquête; cela va des simples vérifications préliminaires à l'intérieur du questionnaire électronique fait pendant l'interview aux vérifications automatisées plus complexes exécutées par un programme informatique après la saisie des données. Les règles de la vérification sont généralement formulées d'après ce qui peut être logique ou valide, compte tenu :

- des connaissances de l'expert en la matière;
- d'autres enquêtes ou données connexes;
- de la structure du questionnaire et de ses questions;
- d'une théorie statistique.

Il y a trois principales catégories de vérification : les contrôles de la validité, de la cohérence et de la distribution. Les vérifications de validité ciblent la syntaxe des réponses et comprennent la vérification des caractères non numériques entrés dans les champs numériques et le repérage des valeurs manquantes. Celles-ci peuvent aussi déterminer si les données codées s'inscrivent dans l'étendue permise des valeurs. Une vérification de l'étendue peut être faite, par exemple, pour l'âge déclaré d'un répondant, afin de vérifier s'il se situe entre 0 et 125 ans.

Les vérifications de cohérence déterminent si les liens entre les questions sont respectés. Elles peuvent utiliser des liens logiques, juridiques, comptables ou structurels entre les questions ou entre les volets d'une question.

Les contrôles de la distribution se font par l'examen des données de l'ensemble des questionnaires. Ces contrôles tentent de détecter les enregistrements qui contiennent des valeurs aberrantes relativement à la distribution de données. Les contrôles de la distribution sont parfois nommés vérification statistique ou détection des valeurs aberrantes.

Un exemple de vérification de cohérence serait que les enregistrements dont le total des dépenses pondérées du voyage est extrême sont identifiés et traités dans le système de traitement des valeurs aberrantes.

Vérification générée par ordinateur

Tel que mentionné dans une section précédente, un contrôle se fait au moment même que le questionnaire électronique est rempli par le répondant. 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, le répondant est invité, via des messages apparaissant à l'écran de l'ordinateur, à vérifier les renseignements. Cependant, pour certaines questions, le répondant peut ignorer ces contrôles et sauter des questions si il ou elle ne connaît pas la réponse ou refuse de répondre. Pour cette raison, les données des réponses sont soumises à d'autres processus de vérification et d'imputation après réception au bureau central.

Vérifications d'intervalles

Les vérifications d'intervalles ont été intégrées au QE pour les questions où il faut indiquer des valeurs numériques. Si les chiffres indiqués ne s'inscrivent pas dans l'intervalle, une fenêtre apparaît instantanément pour indiquer l'erreur et demander au répondant QE de corriger la

réponse erronée. Par exemple, si la valeur saisie dans l'ordinateur pour le nombre de nuitées passées à l'extérieur au cours du voyage est élevée, un message contextuel apparaîtra demandant au répondant de confirmer la réponse.

Vérifications du cheminement de questions

Tous les cheminements de questions sont en mode intégration automatique au QE. Par exemple, si le voyage a été défini comme un voyage d'une nuit ou plus, l'application affichera une série de questions au sujet de l'hébergement. Dans le cas contraire, il saute automatiquement ces questions.

Les données sont transmises à Statistique Canada, où l'on procède à des étapes de traitement de l'information pour préparer les données en vue de la diffusion. On vérifie les données afin d'y déceler les incohérences. Les enregistrements de voyage sont validés pour s'assurer que les valeurs entrées dans les champs essentiels sont acceptables. Pour certaines variables, un intervalle de valeurs acceptables est utilisé. Par exemple, on s'assure que le nombre de nuits ne dépasse pas l'intervalle logique, que le type de voyage est valide, etc.

Pour la majorité des enregistrements de voyage, le codage de la région géographique se fait automatiquement. Pour un petit nombre d'enregistrements, le codage de la région géographique est fait manuellement, au bureau central de Statistique Canada.

Vérifications de la cohérence générale

On a prévu un certain nombre de vérifications de la cohérence dans le système de collecte QE pour permettre au répondant de revenir à des questions déjà posées et de rectifier les incohérences. Les répondants reçoivent des instructions à l'écran pour traiter ou régler des problèmes de réponse incomplète ou erronée. Le système génère une fenêtre contextuelle qui signale l'erreur et demande au répondant de revenir à la question pour vérifier les données et apporter les corrections nécessaires.

Plusieurs contrôles de la cohérence sont effectués sur les données pour vérifier la relation entre deux variables ou plus. Par exemple, le nombre d'adultes d'un ménage qui ont pris part à un voyage ne peut dépasser le nombre total d'adultes dans le ménage. Si une ville ou autre lieu géographique précis ne correspond pas à la province ou autre lieu géographique plus global, alors un seul lieu sera gardé, selon la question. Pour les variables de dépenses, plusieurs règles de vérification sont appliquées afin que borner ces valeurs. Si la valeur n'est pas dans l'intervalle acceptable prédéterminée, alors cette dernière sera imputée ultérieurement.

Vérifications préliminaires

Pour tous les enregistrements où il manque des valeurs (blancs) provenant de la collecte, on insère la valeur « 9 », « 99 », « 999 » etc. pour indiquer qu'aucune information n'a été recueillie. Les entrées « Ne sait pas » retournées par les deux applications de collecte comme « 9 » sont modifiées à « 7 » à l'étape de la vérification préliminaire. De même, on a modifié les questions pour lesquelles on peut lire « Inscrivez toutes les réponses qui s'appliquent » afin de transformer leurs valeurs en réponses « Oui » (« 1 ») ou « Non » (« 2 »). Enfin, toutes les réponses comprenant du texte sont enlevées du fichier de traitement et sont mises de côté afin d'être traitées séparément.

Vérifications de l'enchaînement

Cette vérification consiste à reproduire le cheminement des questions incluses dans le questionnaire. Les variables qui manquent, en suivant l'enchaînement des questions, sont changées de « Non déclaré » à « Saut valide » (« 6 », « 96 », « 996 » etc.). Pour les sauts fondés sur l'âge ou selon la réponse à certaines questions, les questions sautées sont établies à « Saut valide ». Dans le cas de sauts fondés sur les réponses « Ne sait pas » et les « Refus », celles-ci sont établies à « Non déclaré ».

Vérification de la cohérence

Après la vérification de l'enchaînement, on procède à la vérification de la cohérence qui consiste à vérifier les liens entre deux variables ou plus. On utilise des tables de décision pour préciser les vérifications de la cohérence. Le logiciel LogiPlus sert à saisir l'information dans les tables de décision et à générer le code SAS. Pour les vérifications plus complexes, la programmation SAS est utilisée. On produit un rapport comprenant les comptes « avant » et « après » des variables, de même qu'un rapport précisant le nombre de décisions prises pour chaque table de décision.

5.3 Codage des questions ouvertes

Les répondants du QE enregistrent, sur le questionnaire de l'ENV, quelques éléments d'information sous forme de réponses à des questions ouvertes. Par exemple, si le croisiériste de la croisière ne figure pas sur la liste, le répondant doit le saisir manuellement. Ce texte est codé selon un système interne.

5.4 Création de variables dérivées

Pour faciliter l'analyse des données, on a calculé un certain nombre de données élémentaires figurant dans le fichier de microdonnées en combinant des réponses sur le questionnaire. À titre d'exemple, MREASON définit la raison principale du voyage. Plusieurs questions sont combinées pour créer cette nouvelle variable.

5.5 Imputation

L'imputation est un processus utilisé pour déterminer et attribuer des valeurs de remplacement, afin de résoudre les problèmes que suscitent les données manquantes, invalides ou incohérentes.

5.5.1 Imputation pour le nombre de personnes qui accompagnait

Le nombre de personnes qui résident dans le même ménage que le répondant et qui l'ont accompagné lors du voyage est requis pour l'analyse et la calibration. Il est également utilisé pour l'imputation des dépenses. En l'absence de ces valeurs (ou si elles ne correspondent pas à la taille déclarée du ménage), elles doivent être imputées.

La stratégie d'imputation par donneur est utilisée dans le logiciel BANFF, qui dépend de la méthode du plus proche voisin.

5.5.2 Imputation des dépenses

Le système d'imputation de dépenses de voyage veille à ce que tous les voyages choisis pour être détaillés comptent une valeur de dépense valide pour l'ensemble des catégories de dépenses. Toutes les autres valeurs manquantes ou erronées sont soit corrigées ou soit remplacées par un code « non déclaré ».

La composante des dépenses est imputée pour certains voyages quand la probabilité que le répondant ait entré une valeur erronée est élevée, ou quand aucune valeur n'a été entrée.

La première étape du système d'imputation des dépenses est d'utiliser des indicateurs pour déterminer les valeurs à imputer. Ces indicateurs sont d'abord créés au cours de la procédure de contrôle du système de traitement des données.

La méthode d'imputation par hot deck aléatoire pour une classe est utilisée quand les variables des dépenses de voyage devant faire l'objet d'une imputation ont été repérées. Plutôt que d'imputer directement le donneur de dépenses, celui-ci est divisé par un dénominateur (p. ex. le nombre de nuitées) et multiplié à nouveau par le dénominateur du receveur.

Après qu'une valeur ait été repérée en vue d'une imputation, le système choisit aléatoirement un enregistrement donneur dans la même classe d'imputation que le receveur. Le rapport (p. ex. dépenses par jour ou dépenses par personne par jour) du donneur est imputé au receveur. Ensuite, ce rapport est converti en une valeur des dépenses par multiplication avec le dénominateur. Par exemple, si les dépenses par jour sont imputées à un enregistrement donné, cette valeur est multipliée par le nombre de jours. Le tableau ci-dessous présente les variables de dénominateur pour créer le rapport pour chaque catégorie de dépenses.

Pour éviter l'imputation de valeurs extrêmes, les valeurs supérieures à 95 % et inférieures à 5 % du bassin de donneurs sont exclues et ne peuvent servir à imputer des enregistrements. Les valeurs aberrantes sont également exclues du bassin de donneurs.

5.5.3 Imputation des coûts des forfaits de voyages

Si un répondant a déclaré avoir acheté un voyage à forfait, mais que le coût de celui-ci est 0 \$ ou non déclaré, le coût total du forfait est imputé au moyen d'un donneur. Dans ce cas, la valeur des dépenses d'un donneur (et non un rapport) est imputée au receveur. La méthode d'imputation utilisée est celle du donneur aléatoire à l'intérieur des classes d'imputation. Des classes hiérarchiques sont créées à partir de huit caractéristiques. Par exemple, s'il est impossible de trouver un donneur dont les valeurs correspondent aux huit caractéristiques pour un voyage intérieur, alors la variable ayant la plus faible corrélation avec les dépenses liées au forfait est éliminée. Le système cherche alors un donneur pour lequel toutes les autres variables correspondent. En l'absence d'un tel donneur, la variable se classant au deuxième rang de corrélation la plus faible est éliminée et ainsi de suite jusqu'à ce qu'un donneur potentiel soit trouvé.

5.6 Contrôle de la divulgation

6.0 Qualité des données

Diverses sources expliquent les erreurs d'enquête. Elles peuvent être classées en deux principales catégories : erreur d'échantillonnage et erreurs non dues à l'échantillonnage.

6.1 Erreurs non dues à l'échantillonnage

Les erreurs non dues à l'échantillonnage peuvent être définies comme des erreurs pouvant survenir au cours d'à peu près toutes les activités d'une enquête, mis à part l'échantillonnage. Ces erreurs se retrouvent dans l'enquête-échantillon et le recensement (contrairement à l'erreur d'échantillonnage qui est présente seulement dans l'enquête-échantillon). Les principales sources d'erreurs non dues à l'échantillonnage sont les suivantes : non-réponse, couverture, mesure et traitement.

6.1.1 Non-réponse

La non-réponse résulte de l'incapacité de recueillir l'information nécessaire pour toutes les unités sélectionnées dans l'échantillon.

La non-réponse peut occasionner deux types d'erreurs dans l'enquête. D'abord, les non-répondants ont souvent des caractéristiques différentes de celles des répondants, et les estimations de l'enquête seront biaisées si les non-réponses ne sont pas corrigées. Ensuite, la non-réponse diminue la taille réelle de l'échantillon parce qu'il était prévu que davantage

d'unités répondraient à l'enquête. La variance d'échantillonnage augmente donc et la précision des estimations diminue.

Le tableau suivant résume les taux de réponse.

Province	Ménage sélectionnés	Ménages hors champ	Ménages répondants	Taux de réponse
Terre-Neuve et Labrador	24 685	1 241	4 150	17,7%
Île-du-Prince-Édouard	16 428	1 537	2 881	19,3%
Nouvelle-Écosse	28 444	1 430	5 938	22,0%
Nouveau-Brunswick	26 138	1 151	4 845	19,4%
Québec	82 199	1 827	19 850	24,7%
Ontario	106 950	2 218	28 582	27,3%
Manitoba	38 138	686	9 645	25,8%
Saskatchewan	40 991	870	10 479	26,1%
Alberta	54 706	1 214	12 879	24,1%
Colombie-Britannique	57 517	1 338	14 803	26,3%
Canada	476 196	13 512	114 052	24,7%

6.1.2 Erreurs de couverture

Les erreurs de couverture sont des omissions, des ajouts erronés, des répétitions et des erreurs de classification d'unités dans la base de sondage. Elles ont des répercussions sur chaque estimation de l'enquête et sont donc l'un des plus importants types d'erreur. Elles peuvent même être la principale source d'erreur d'un recensement. Les erreurs de couverture peuvent susciter des estimations biaisées, et les répercussions peuvent varier pour différents sous-groupes de la population.

6.1.3 Erreurs de mesure

L'erreur de mesure (ou erreur de réponse) est la différence entre la réponse inscrite à une question et la « vraie » valeur. Le répondant, l'intervieweur, le questionnaire, la méthode de collecte des données et l'outil de mesure peuvent susciter ce genre d'erreur. Si les erreurs de mesure reflètent

systématiquement certaines valeurs ou catégories, un biais se glissera et les estimations de l'enquête seront trompeuses.

Le questionnaire électronique de l'ENV a des caractéristiques incorporées afin de maximiser la qualité de la collecte des données. Des vérifications sont intégrées dans le QE pour repérer des valeurs inhabituelles dans les entrées de données et discerner des incohérences de logique. Le QE contrôle la séquence ou l'enchaînement des questions en fonction des réponses aux questions précédentes. D'autres vérifications sont aussi effectuées au cours du questionnaire pour réduire le nombre d'erreurs attribuables aux fautes de frappe ou aux malentendus. Par exemple, si le nombre de nuitées passées dans les divers types d'hébergement ne correspond pas au nombre total de nuitées passées à l'extérieur du domicile, un message de contrôle apparaîtra à l'écran. Le répondant peut alors corriger l'erreur, ce qui réduit les vérifications devant être effectuées au bureau central de Statistique Canada. De plus, pour certaine question, le répondant a la possibilité de choisir « Ne sait pas » ou « Refus » comme réponse valide s'il ou elle ne peut répondre à la question.

6.1.4 Erreurs de traitement

Une erreur de traitement est une erreur associée avec les activités menées une fois les réponses de l'enquête reçues. Cela comprend toutes les activités de manutention des données après la collecte et avant l'estimation. Comme toutes les autres erreurs, elles peuvent être aléatoires et accroître ainsi la variance des estimations de l'enquête, ou elles peuvent être systématiques et ajouter un biais. Il est difficile d'obtenir des mesures directes des erreurs de traitement ainsi que leur incidence sur la qualité des données, en particulier parce qu'elles sont mélangées à d'autres types d'erreur (non-réponse, mesure et couverture).

Le traitement des données de l'ENV a été fait par étape, comprenant la vérification, le codage, le contrôle, l'imputation, etc. Des mesures sont prises à chaque étape pour veiller à produire des données de qualité et des rapports sont créés pour vérification.

6.2 Erreurs d'échantillonnage

L'erreur d'échantillonnage est définie comme étant l'erreur résultant de l'estimation d'une caractéristique de la population en mesurant une partie au lieu de toute la population. Des méthodes de calcul de l'erreur d'échantillonnage s'appliquent à l'enquête-échantillon probabiliste. Ces méthodes découlent directement du plan d'échantillonnage et de la méthode d'estimation de l'enquête.

La mesure appliquée le plus souvent pour quantifier l'erreur d'échantillonnage est la variance d'échantillonnage. La variance d'échantillonnage détermine à quel point l'estimation d'une caractéristique de divers échantillons possibles de même taille et de même conception est différente l'une de l'autre. Dans le cas des plans d'échantillonnage qui utilisent l'échantillonnage probabiliste, l'ampleur de la variance d'échantillonnage d'une estimation peut être déterminée. Le point principal est l'ampleur de la variance d'échantillonnage estimée d'une estimation relativement à la taille de l'estimation de l'enquête : si la variance est relativement grande, la précision de l'estimation est donc médiocre et n'est pas fiable.

Les éléments qui ont des répercussions sur l'ampleur de la variance d'échantillonnage comprennent :

1. La variabilité de la caractéristique d'intérêt dans la population : plus la caractéristique dans la population est variable, plus la variance d'échantillonnage est grande.
2. La taille de la population : en général, la taille de la population a des répercussions sur la variance d'échantillonnage seulement pour les populations de petite taille ou de taille moyenne.

3. Le taux de réponse : la variance d'échantillonnage augmente dans la mesure où la taille de l'échantillon diminue. Étant donné que les non-répondants diminuent en fait la taille de l'échantillon, les non-réponses augmentent la variance d'échantillonnage.
4. Le plan d'échantillonnage et les méthodes d'estimation : certains plans d'échantillonnage sont plus efficaces que d'autres parce que, pour la même taille d'échantillon et la même méthode d'estimation, un plan peut donner une variance d'échantillonnage moindre que l'autre.

L'erreur-type d'un estimateur est la racine carrée de sa variance d'échantillonnage. Cette mesure est plus facile à interpréter parce qu'elle donne une indication de l'erreur d'échantillonnage à l'aide de la même échelle que l'estimation, tandis que la variance est basée sur les différences quadratiques.

Même l'erreur-type peut cependant être difficile à interpréter lorsqu'on pose la question « Quelle ampleur d'erreur-type est acceptable? » C'est l'importance de l'estimation qui détermine la largeur. Une erreur-type de 100, par exemple, serait considérée grande pour mesurer la moyenne du poids des gens, mais pas pour estimer le revenu annuel moyen.

Il est plus utile dans de nombreuses situations d'évaluer la taille de l'erreur-type par rapport à l'estimation de la caractéristique mesurée. Le coefficient de variation donne ce genre de mesure. C'est le rapport entre l'erreur-type de l'estimation de l'enquête et la valeur moyenne de l'estimation elle-même, pour tous les échantillons possibles. Le coefficient de variation est habituellement calculé comme l'estimation de l'erreur-type de l'estimation de l'enquête par rapport à l'estimation elle-même. Cette mesure relative de l'erreur d'échantillonnage est habituellement exprimée en pourcentage (10 % au lieu de 0,1). Elle est très utile pour comparer la précision des estimations de l'échantillon lorsque leurs tailles ou échelles sont différentes l'une de l'autre.

6.2.1 Résultats des estimations principales

Par conséquent, des tableaux de CV approximatifs ont été fournis aux utilisateurs pour leur permettre d'estimer le CV des caractéristiques qui les intéressent. L'annexe B explique comment utiliser les tableaux de CV dans diverses situations et donne des exemples.

7.0 Pondération

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

L'étape de la pondération est celle qui consiste à calculer 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.

La présente section renferme des détails au sujet de la méthode utilisée pour calculer les poids pour l'ENV.

7.1 Procédure de pondération pour l'ENV

Quatre ensembles de poids sont créés et fournis pour l'ENV : les poids des personnes, les poids des voyages-personnes et les poids des voyages.

Poids des personnes

Le poids initial de l'ENV correspond à l'inverse de la probabilité de sélection des ménages sélectionnés. Les ménages dont tous les membres sont âgés de moins de 18 ans sont hors du champ d'observation de l'ENV et sont donc retirés. Dans les ménages admissibles aux fins de l'ENV, une personne de 18 ans et plus est choisie au hasard.

Le poids initial de l'ENV est ensuite ajusté en fonction du membre du ménage choisi selon les facteurs suivants :

- a) Ajustement de la non-réponse (ménage) : un facteur est appliqué aux ménages répondants à l'intérieur de classes où la propension de réponse est similaire.
- b) Un calage est effectué par province et taille de ménage avec des totaux de contrôle connus.
- c) Nombre de personnes admissibles (18 ans et plus) dans le ménage : Ce facteur d'ajustement est plafonné à 6 pour contrôler les poids des répondants dans les rares ménages comptant de nombreux membres admissibles âgés de 18 ans et plus.
- d) Ajustement de la non-réponse (personne) : pour compenser la non-réponse, l'ENV crée des classes de non-réponse pour chaque province. Elle crée ces classes en modélisant les probabilités de réponse au moyen des variables démographiques et en utilisant un algorithme de groupement pour créer des classes ayant des probabilités de réponse semblable.
- e) Un calage est effectué par province, âge, sexe et RMR (région métropolitaine de recensement) avec des totaux de contrôle connus.

Poids des voyages-personnes

L'unité d'échantillonnage de l'ENV est le ménage (personne âgée de 18 ans et plus), mais c'est le voyage qui représente la principale unité d'analyse. Une unité d'échantillonnage donnée peut avoir effectué un ou plusieurs voyages, ou même n'en avoir fait aucun. En outre, la personne peut avoir fait un voyage en compagnie d'autres voyageurs. Afin de créer un enregistrement pour le voyage assorti du poids approprié, on définit tout d'abord un voyage-personne, terme désignant tout répondant ayant déclaré avoir fait un voyage (quel que soit le nombre de personnes ayant pris part à ce voyage).

- a) Le poids de départ pour le poids des voyages-personnes est le poids des personnes.
- b) Ajustement des voyages ayant des données essentielles manquantes: Si le répondant a déclaré à la fois des voyages valides appartenant au champ de l'enquête et des voyages appartenant au champ de l'enquête dont des données essentielles étaient manquantes, le poids des voyages dont les données essentielles étaient manquantes sera redistribué parmi les voyages valides..

Poids des voyages

Le poids des voyages-personnes est ensuite ajusté pour tenir compte du nombre de personnes du même ménage âgées de 18 ans et plus et ayant pris part au voyage (y compris le répondant), afin de déterminer la probabilité de sélection du voyage.

8.0 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 des chiffres semblables à ceux produits par Statistique Canada, tout en étant en mesure d'obtenir des chiffres actuellement inédits de façon conforme à ces lignes directrices établies. Dans certains cas les chiffres peuvent être légèrement différents car ces chiffres produits par Statistique Canada le sont en utilisant les fichiers maîtres et non le FMGD.

8.1 Lignes directrices pour l'arrondissement d'estimations

Afin que les estimations qui sont destinées à la publication ou à toute autre forme de diffusion et 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. Veuillez prendre note que les données publiées dans les tableaux de l'ENV ont été arrondies au millier près.
- 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 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. Les proportions et rapports doivent d'abord être calculés à partir de composantes non arrondies puis être arrondis à leur tour à la troisième décimale à l'aide de la technique d'arrondissement normale.
- d) Les sommes et les différences d'agrégats 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. Les sommes et les différences de pourcentages (ou rapports) doivent être calculées à partir de leurs composantes non arrondies correspondantes, puis être arrondies à leur tour à la décimale près (ou troisième décimale) à l'aide de la technique d'arrondissement normale.
- e) Dans les cas, où, en raison de limitations d'ordre techniques ou de toutes autres limites, une technique d'arrondissement autre que la technique normale est utilisée produisant des estimations à être publiées ou autrement diffusées différentes des estimations correspondantes publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces différences dans le ou les documents à publier ou à diffuser.
- f) En aucun cas, les utilisateurs ne doivent publier ou autrement diffuser des estimations non arrondies. Des estimations non arrondies laissent entendre qu'elles sont plus précises qu'elles le sont en réalité.

Annexe A – Estimation de la variance pour les fichiers maîtres

Pour juger de la qualité des estimations et en calculer le CV, on doit d'abord établir l'écart-type. On a aussi besoin de l'écart-type de l'estimation pour les intervalles de confiance. Dans l'ENV, on utilise des plans d'échantillonnage et de calibration à plusieurs degrés et, par conséquent, il n'y a pas de formule simple d'estimation de variance qu'on puisse appliquer. Il faut donc procéder par approximation, la méthode bootstrap. Par l'utilisation des poids bootstrap et du programme BOOTVAR, dont il est question à la prochaine section, on peut calculer avec exactitude les coefficients de variation et d'autres estimations de la variance.

Méthode bootstrap pour estimer la variance

On prélève indépendamment sur chaque strate un échantillon aléatoire simple de $(n-1)$ des n unités primaires d'échantillonnage (UPE) de l'échantillon avec remise, ainsi que tous les ménages de l'échantillon appartenant à ces UPE. À noter que, comme la sélection s'opère avec remise, une unité peut être choisie plus d'une fois. On calcule un poids bootstrap initial pour chaque unité échantillonnée de la strate. Toute l'opération (prélèvement d'échantillons aléatoires simples et pondération des unités pour chaque strate) a lieu à B reprises, B étant une valeur élevée, ce qui donne B valeurs initiales de poids bootstrap.

Le paragraphe précédent donne une description de la méthode bootstrap générale. Dans le cas de l'ENV, la plupart des strates comprennent une seule UPE. Les strates sont donc groupées en super-strates, et $n-1$ strates sont sélectionnées au hasard dans chaque super-strate pour chacun des B rééchantillonnages bootstrap. On doit également procéder à la coordination des échantillons bootstrap pour tenir compte dans le calcul de la variance de la dépendance entre les échantillons au cours de deux cycles d'enquête. Pour l'ENV, la coordination est faite parce qu'il y a toujours un chevauchement des échantillons entre deux mois de référence consécutifs (chaque interview menée contribuant aux estimations de deux mois de référence consécutifs). En vue de tenir compte de ce chevauchement des échantillons, le même ensemble de B sélections des $n-1$ strates par super-strate est reporté de mois en mois.

Une fois que l'on a sélectionné les échantillons bootstrap et créé les poids bootstrap initiaux, on ajuste ces poids selon la même procédure de pondération que pour les poids ordinaires : ajustement pour tenir compte de la non-réponse, calage, etc. On obtient B poids bootstrap finaux pour chaque unité de l'échantillon. La variabilité entre les B estimations basées sur les B poids bootstrap est associée à la variance de l'estimateur basé sur les poids réguliers et peut être utilisée pour estimer cette dernière.

Les estimations de la variance diffusées par Statistique Canada sont calculées à partir des poids bootstrap.

Progiciels statistiques pour estimer la variance

Bootvar

On signale aux utilisateurs que des poids bootstrap sont fournis et devraient être utilisés pour l'estimation de la variance. BOOTVAR est un programme de macro qui peut être utilisé pour calculer la variance au moyen de poids bootstrap. Le programme Bootvar est disponible en format SAS. Il est formé de macro instructions d'estimation de variance pour les totaux, les rapports et les différences entre rapports, ainsi que de régression linéaire et logistique.

Le Bootvar peut être téléchargé à partir du site Internet des Centres de données de recherche (CDR) de Statistique Canada. Il faut accepter la licence d'adhésion automatique Bootvar avant de pouvoir lire les fichiers. Il y a un document sur le site expliquant comment adapter le système aux besoins des utilisateurs.

SAS : <https://www150.statcan.gc.ca/n1/pub/12-002-x/2014001/article/11901-fra.htm#a8>

Autres progiciels

Outre Bootvar, il existe des progiciels commerciaux différents qui peuvent effectuer une certaine analyse fondée sur le plan d'enquête pour estimer la variance; Stata 9 ou plus, SUDAAN et WesVar. Des descriptions de certains progiciels se trouvent ici:

<http://www.statcan.gc.ca/pub/12-002-x/2014001/article/11901-fra.htm#a4>

Annexe B – Estimation de la variance pour les fichiers de microdonnées à grande diffusion

Tables de variabilité d'échantillonnage approximative

Afin de fournir des coefficients de variation (CV) qui pourraient s'appliquer à une vaste gamme d'estimations produites à partir de ce fichier de microdonnées et auxquels il serait facilement possible pour l'utilisateur d'avoir accès, deux ensembles de tables de variabilité d'échantillonnage approximative ont été produits. 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. Chacun de ces deux ensembles de tables devrait être utilisé pour un type particulier d'estimations. L'un des ensembles de tables ne doit servir que pour les estimations catégoriques, comme les estimations des pourcentages d'agrégats, tandis que l'autre ne doit servir que pour les estimations de variables numériques, par exemple, les dépenses, les nuitées-personnes, les voyages-personnes ou les voyages-ménages.

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 ratios et des différences entre de telles estimations. Il convient de noter que ces tables ne peuvent être utilisées que pour les estimations mensuelles.

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

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

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

Le coefficient de variation d'une proportion estimée ou d'un pourcentage estimé dépend à la fois de la taille de la proportion ou du pourcentage et de la taille du total sur lequel la proportion ou le pourcentage repose. Les proportions estimées ou les pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque la proportion ou que le pourcentage repose sur un sous-groupe de la population. Par exemple, pour une période et une région géographique données, la proportion de femmes ayant fait un voyage est plus fiable que le nombre estimé de femmes ayant voyagé. (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 ratios

Dans le cas où le numérateur est un sous-ensemble du dénominateur, le ratio 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 ayant fait un voyage d'une nuit ou plus.

Dans le cas où le numérateur n'est pas un sous-ensemble du dénominateur, comme dans l'exemple du ratio du nombre de voyageurs ayant fait un voyage du même jour comparativement au nombre de voyageurs ayant fait un voyage d'une nuit ou plus, l'erreur-type du ratio 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 ratio $\left(\hat{R} = \hat{X}_1 / \hat{X}_2\right)$ 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 ratios

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

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

Les exemples ci-dessous sont fondés sur des données fictives d'une enquête sur les voyages intérieurs menées en 2009 dans un pays imaginaire appelé « Pangea ». Ces exemples sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter. Veuillez noter que les données utilisées dans ces exemples diffèrent des véritables résultats d'enquête et tiennent seulement lieu de guide.

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

Supposons qu'un utilisateur estime que 5 414 335 personnes ont fait un voyage au Pangea en avril 2009. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Reportez-vous à la table des coefficients de variation des personnes pour le Pangea.

Tableaux de la Variabilité d'échantillonnage Approximative - Pondération de personne														
Enquête sur les voyages intérieurs en Pangea - Mensuel - 2009														
Pangea														
Numérateur du Pourcentage		Pourcentage Estimé												
(°000)	0.10%	1.00%	2.00%	5.00%	10.00%	15.00%	20.00%	25.00%	30.00%	35.00%	40.00%	50.00%	70.00%	90.00%
1	213.1	212.1	211.1	207.8	202.3	196.6	190.7	184.6	178.4	171.9	165.1	150.8	116.8	67.4
2	150.7	150.0	149.2	146.9	143.0	139.0	134.8	130.6	126.1	121.5	116.8	106.6	82.6	47.7
3	123.0	122.5	121.9	120.0	116.8	113.5	110.1	106.6	103.0	99.2	95.3	87.0	67.4	38.9
4	106.5	106.1	105.5	103.9	101.1	98.3	95.3	92.3	89.2	85.9	82.6	75.4	58.4	33.7
5	95.3	94.9	94.4	92.9	90.5	87.9	85.3	82.6	79.8	76.9	73.9	67.4	52.2	30.2
.
.
.
.
.
75	*****	24.5	24.4	24.0	23.4	22.7	22.0	21.3	20.6	19.8	19.1	17.4	13.5	7.8
80	*****	23.7	23.6	23.2	22.6	22.0	21.3	20.6	19.9	19.2	18.5	16.9	13.1	7.5
85	*****	23.0	22.9	22.5	21.9	21.3	20.7	20.0	19.3	18.6	17.9	16.4	12.7	7.3
90	*****	22.4	22.2	21.9	21.3	20.7	20.1	19.5	18.8	18.1	17.4	15.9	12.3	7.1
95	*****	21.8	21.7	21.3	20.8	20.2	19.6	18.9	18.3	17.6	16.9	15.5	12.0	6.9
100	*****	21.2	21.1	20.8	20.2	19.7	19.1	18.5	17.8	17.2	16.5	15.1	11.7	6.7
125	*****	19.0	18.9	18.6	18.1	17.6	17.1	16.5	16.0	15.4	14.8	13.5	10.4	6.0
150	*****	17.3	17.2	17.0	16.5	16.0	15.6	15.1	14.6	14.0	13.5	12.3	9.5	5.5
200	*****	15.0	14.9	14.7	14.3	13.9	13.5	13.1	12.6	12.2	11.7	10.7	8.3	4.8
250	*****	*****	13.3	13.1	12.8	12.4	12.1	11.7	11.3	10.9	10.4	9.5	7.4	4.3
300	*****	*****	12.2	12.0	11.7	11.3	11.0	10.7	10.3	9.9	9.5	8.7	6.7	3.9
350	*****	*****	11.3	11.1	10.8	10.5	10.2	9.9	9.5	9.2	8.8	8.1	6.2	3.6
400	*****	*****	10.6	10.4	10.1	9.8	9.5	9.2	8.9	8.6	8.3	7.5	5.8	3.4
450	*****	*****	9.9	9.8	9.5	9.3	9.0	8.7	8.4	8.1	7.8	7.1	5.5	3.2
500	*****	*****	*****	9.3	9.0	8.8	8.5	8.3	8.0	7.7	7.4	6.7	5.2	3.0
750	*****	*****	*****	7.6	7.4	7.2	7.0	6.7	6.5	6.3	6.0	5.5	4.3	2.5
1,000	*****	*****	*****	*****	6.4	6.2	6.0	5.8	5.6	5.4	5.2	4.8	3.7	2.1
1,500	*****	*****	*****	*****	5.2	5.1	4.9	4.8	4.6	4.4	4.3	3.9	3.0	1.7
2,000	*****	*****	*****	*****	*****	4.4	4.3	4.1	4.0	3.8	3.7	3.4	2.6	1.5
3,000	*****	*****	*****	*****	*****	*****	3.4	3.3	3.3	3.1	3.0	2.8	2.1	1.2
4,000	*****	*****	*****	*****	*****	*****	3.0	2.9	2.8	2.7	2.6	2.4	1.8	1.1
5,000	*****	*****	*****	*****	*****	*****	*****	2.6	2.5	2.4	2.3	2.1	1.7	1.0
6,000	*****	*****	*****	*****	*****	*****	*****	*****	2.3	2.2	2.1	1.9	1.5	0.9
7,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	2.1	2.0	1.8	1.4	0.8
8,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.8	1.7	1.3	0.8
9,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.7	1.6	1.2	0.7
10,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.5	1.2	0.7
12,500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.5	1.0	0.6
15,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.0	0.6
20,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	0.5
NOTE: POUR BIEN UTILISER CES TABLES. VEUILLEZ CONSULTER LA DOCUMENTATION SUR LES MICRODONNÉES														

- 2) L'agrégat estimé (5 414 335) ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, c'est-à-dire 5 000 000.
- 3) On trouve le coefficient de variation pour un agrégat estimé en se reportant à la première entrée autre que des astérisques sur cette rangée, c'est-à-dire 2,6 %.
- 4) Le coefficient de variation approximatif de l'estimation est donc 2,6 %. Le résultat selon lequel il y avait 5 414 335 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 8.1) voyageurs durant la période de référence peut être publié sans réserve.

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

Supposons qu'un utilisateur estime à $516\,365 / 2\,865\,929 = 18\%$ la proportion d'hommes qui ont fait une randonnée au cours de leur voyage au Pangea en janvier 2009. 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 Pangea du fichier des personnes (voir ci-dessus). La table au niveau du Pangea devrait être utilisée parce qu'il s'agit de la plus petite table qui contient le domaine de l'estimation, tous les hommes qui ont fait un voyage au Pangea en janvier 2009. Si, par exemple, le pourcentage se rapporte aux hommes qui ont fait un voyage dans une province donnée du Pangea, il faudra utiliser la table de cette province.
- 2) Parce que l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c'est-à-dire les hommes qui ont fait un voyage au Pangea en janvier 2009), il faut utiliser à la fois le pourcentage (18 %) et la portion numérateur du pourcentage (516 365) pour déterminer le coefficient de variation.
- 3) Le numérateur, 516 365, 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 500 000. 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 20,0 %.
- 4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 8,5 %, est le coefficient de variation à employer.
- 5) Le coefficient de variation approximatif de l'estimation est donc 8,5 %. Le résultat selon lequel 18 % des hommes qui ont voyagé au Pangea en janvier 2009 ont fait une randonnée durant leur voyage peut être publié sans réserve.

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

Supposons qu'un utilisateur estime à $2\,548\,406 / 11\,814\,359 = 21,6\%$ la proportion de femmes qui ont fait un voyage au Pangea en janvier 2009 et à $2\,865\,929 / 11\,436\,728 = 25,1\%$ la proportion correspondante pour les hommes. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

- 1) L'utilisation de la table des coefficients de variation pour le Pangea du fichier des personnes (voir ci-dessus) de la même façon que celle décrite dans l'exemple 2 donne un CV de l'estimation pour les femmes de 3,4 % et un CV de l'estimation pour les hommes de 3,3 %.
- 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.

- 3) C'est-à-dire que l'erreur-type de la différence $\hat{d} = 0,251 - 0,216 = 0,035$ est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0,251)(0,033)]^2 + [(0,216)(0,034)]^2} \\ &= \sqrt{0,000069 + 0,000054} \\ &= 0,011\end{aligned}$$

Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,011 / 0,035 = 0,314$.

Le coefficient de variation approximatif de la différence entre les estimations est donc 31,4 %. La différence entre les estimations est considérée comme médiocre, et Statistique Canada recommande de faire preuve de prudence lors de la diffusion de cette estimation et de mettre en garde les utilisateurs quant à l'exactitude de l'estimation.

Exemple 4 : Estimations de ratios

Supposons qu'un utilisateur estime à 807 261 le nombre de femmes qui ont fait au moins un voyage d'une nuit ou plus au Pangea en novembre 2009 et à 968 511 le nombre d'hommes qui ont fait au moins un voyage d'une nuit ou plus au Pangea au cours de la même période. L'utilisateur veut comparer l'estimation des femmes à celle des hommes sous la forme d'un ratio. Comment détermine-t-il le coefficient de variation de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation d'un ratio, où le numérateur de l'estimation (\hat{X}_1) est le nombre de femmes ayant fait au moins un voyage d'une nuit ou plus au Pangea en novembre 2009. Le dénominateur de l'estimation (\hat{X}_2) est le nombre d'hommes ayant fait au moins un voyage d'une nuit ou plus au Pangea au cours de la même période.
- 2) Reportez-vous à la table des coefficients de variation pour le Pangea du fichier des personnes.
- 3) Le numérateur de cette estimation de ratio est 807 261. Le chiffre qui s'en rapproche le plus est 750 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 7,6 % (voir ci-dessus).
- 4) Le dénominateur de cette estimation de ratio est 968 511. 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 6,4 %.
- 5) Le coefficient de variation approximatif de l'estimation du ratio 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 :

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0,076)^2 + (0,064)^2} \\ &= \sqrt{0,0058 + 0,0041} \\ &= 0,099\end{aligned}$$

- 6) Le ratio obtenu entre les femmes et les hommes qui ont fait au moins un voyage d'une nuit ou plus au Pangea en novembre 2009 est 807 261/ 968 511, c'est-à-dire 0,83 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 8.1). Le coefficient de variation de cette estimation est 9,9 %, ce qui fait qu'on peut la diffuser sans réserve.

Comment utiliser les tables de coefficients de variation pour obtenir des intervalles 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, alors dans 95 % des échantillons, l'intervalle englobera la vraie valeur de la population.

En utilisant l'erreur-type d'une estimation, des intervalles de confiance pour des estimations peuvent être obtenus 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 vraie valeur 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 on désire un intervalle de confiance de 68 %;
- $t = 1,6$ si on désire un intervalle de confiance de 90 %;
- $t = 2$ si on désire un intervalle de confiance de 95 %;
- $t = 2,6$ si 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.

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

Un intervalle de confiance de 95 % pour la proportion estimée des hommes qui ont fait une randonnée lors de leur voyage au Pangea en janvier 2009 (d'après l'exemple 2) serait calculé comme suit :

$$\hat{X} = 18 \% \text{ (ou exprimé sous forme de 0,18)}$$

$$t = 2$$

$\alpha_{\hat{x}} = 8,5 \%$ (0,085 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.

$$IC_{\hat{x}} = \{0,18 - (2) (0,18) (0,085), 0,18 + (2) (0,18) (0,085)\}$$

$$IC_{\hat{x}} = (0,18 - 0,031, 0,18 + 0,031)$$

$$IC_{\hat{x}} = (0,149, 0,211)$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 14,9 % et 21,1 % des hommes qui ont voyagé au Pangea en janvier 2009 ont fait une randonnée durant leur voyage.

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

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

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

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 ratio 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.

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 femmes qui ont fait un voyage au Pangea en janvier 2009 et la proportion d'hommes qui ont fait un voyage au pays au cours de la même période. D'après l'exemple 3, l'erreur-type de la différence entre ces deux estimations s'est établie à 0,011. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0,251 - 0,216}{0,011} = \frac{0,035}{0,011} = 3,18$$

Puisque $t = 3,18$ ne se trouve pas entre -2 et 2, il faut en conclure qu'il existe une différence significative entre les deux estimations au niveau de signification de 0,05.

Comment utiliser les tables de coefficients de variation pour des estimations de variables numériques

On peut obtenir les coefficients de variation approximatifs à partir des tables de variabilité d'échantillonnage approximative pour les estimations du nombre de voyages-ménages, de voyages-personnes, de nuitées-personnes, et pour les dépenses (c'est-à-dire pour les variables numériques) en utilisant un autre ensemble de tables qui diffèrent légèrement des tables présentées pour les agrégats et les pourcentages. En ce qui concerne le calcul des coefficients de variation des ratios ou des différences entre les estimations de variables numériques, les règles 3 à 5 présentées ci-dessus s'appliquent également.

Estimations de variables numériques

Comme dans les cas discutés plus haut, 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 et la période appropriées, repérez tout d'abord la colonne qui correspond à la variable qui vous intéresse (voyages-ménages, voyages-personnes, nuitées-personnes ou dépenses). L'intersection de cette colonne et de la taille correspondante de l'estimation elle-même devrait donner le CV que vous cherchez. Il convient de noter qu'il y a deux colonnes distinctes présentant des échelles différentes pour les estimations. La dernière colonne à droite doit être utilisée pour les estimations des dépenses seulement, tandis que la première colonne à gauche doit servir pour les estimations des autres variables numériques.

Estimations des différences pour des variables numériques

Reportez-vous à la règle 3 plus haut pour les estimations de différences entre des agrégats ou des pourcentages.

Estimations de ratios

Reportez-vous à la règle 4 plus haut pour les estimations de ratios.

Estimations de différences entre des ratios

Reportez-vous à la règle 5 plus haut pour les estimations de différences entre des ratios.

Exemples d'utilisation des tables de coefficients de variation pour des estimations de variables numériques

Une série d'exemples fondés sur les mêmes données fictives de l'Enquête sur les voyages intérieurs au Pangea de 2009 est présentée ci-dessous pour aider l'utilisateur à se servir des tables visant les estimations de variables numériques. Dans la mesure où les règles 3 à 5 s'appliquant aux estimations d'agrégats s'appliquent également aux estimations de variables numériques, on ne donne pas dans le présent manuel d'exemples des différences ou des ratios entre des estimations de variables numériques. On ne présente pas non plus d'exemples de calcul des intervalles de confiance ou de test t pour les mêmes raisons. Veuillez noter que les données utilisées dans ces exemples diffèrent des véritables résultats d'enquête.

Exemple 5 : Estimations des dépenses

Supposons qu'un utilisateur estime que les résidents du Pangea aient dépensé 786 millions de dollars au titre des voyages intérieurs en mai 2009. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Reportez-vous à la table des coefficients de variation des personnes pour le Pangea – Mai 2009 ci-dessous. Notez que, contrairement aux tables se rapportant aux estimations catégoriques, des tables distinctes sont produites pour chaque mois. Il existe en outre des tables pour les trimestres et pour l'ensemble de l'année.
- 2) Puisqu'il s'agit d'une estimation des dépenses, la valeur de l'estimation se trouve à la dernière colonne à la droite de la table.
- 3) La valeur de l'estimation elle-même (786 millions de dollars) n'apparaît pas dans la table, de sorte qu'elle doit être arrondie à 800 millions de dollars.
- 4) La colonne « Dépenses » indique que le coefficient de variation approximatif est 10,3 %. Le résultat selon lequel la population du Pangea a dépensé 786 millions de dollars (à être arrondi selon les

lignes directrices pour l'arrondissement figurant à la section 8.1) au titre des voyages intérieurs peut être publié sans réserve.

Tableaux de la variabilité d'échantillonnage approximative					
Enquête sur les voyages intérieurs en Pangea : mai 2009					
Pangea					
Estimation ('000)	Coefficient de variation pour:			Dépenses	Estimation ('000,000)
	Voyages- Ménages	Voyages- Personnes	Nuitées- Personnes		
5	70.2	80.2	142.3	43.7	5
10	57.4	64.4	109.8	35.8	10
20	47.0	51.8	84.7	29.4	20
30	41.7	45.6	72.7	26.2	30
40	38.4	41.6	65.3	24.1	40
50	36.0	38.8	60.1	22.6	50
60	34.1	36.6	56.1	21.5	60
70	32.6	34.9	52.9	20.6	70
80	31.4	33.4	50.4	19.8	80
90	30.3	32.2	48.2	19.1	90
100	29.4	31.2	46.3	18.6	100
200	24.1	25.0	35.7	15.2	200
300	21.4	22.0	30.7	13.6	300
400	19.7	20.1	27.5	12.5	400
500	18.4	18.7	25.3	11.7	500
600	17.5	17.7	23.7	11.1	600
700	16.7	16.9	22.3	10.7	700
800	16.1	16.2	21.2	10.3	800
900	15.5	15.6	20.3	9.9	900
1,000	15.1	15.1	19.5	9.6	1,000
2,000	12.3	12.1	15.1	7.9	2,000
3,000	11.0	10.6	12.9	7.0	3,000
4,000	10.1	9.7	11.6		4,000
5,000	9.4	9.1	10.7		5,000
6,000	9.0	8.6	10.0		6,000
7,000	8.6	8.1	9.4		7,000
8,000	8.2	7.8	9.0		8,000
9,000	8.0	7.5	8.6		9,000
10,000	7.7	7.3	8.2		10,000
15,000	6.9	6.4	7.1		15,000
20,000		5.8			20,000
25,000					25,000
NOTE: POUR BIEN UTILISER CES TABLES. VEUILLEZ CONSULTER LA DOCUMENTATION					
SUR LES MICRODONNÉES					

Exemple 6 : Estimations des voyages-personnes

Supposons qu'un utilisateur estime que les résidents du Pangea âgés de 18 à 24 ans aient fait 63 226 voyages-personnes de même jour en mai 2009. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) À partir de la table utilisée à l'exemple 5, repérez la valeur de l'estimation dans la première colonne à la gauche puisqu'elle vise les voyages-personnes. Il en irait de même pour les estimations des nuitées-personnes et des voyages-ménages.
- 2) Comme dans l'exemple précédent, la valeur de l'estimation elle-même (63 226) n'apparaît pas dans la table, de sorte qu'elle doit être arrondie à 60 000.
- 3) La colonne « Voyages-personnes » indique que le coefficient de variation approximatif de l'estimation est 36,6 %. Cette estimation est considérée comme inacceptable et Statistique Canada recommande de ne pas la diffuser.

Annexe C – Différences entre l'Enquête sur les voyages des résidents du Canada et l'Enquête nationale sur les voyages**Introduction**

En 2015, une étude a été menée dans le but d'évaluer la faisabilité d'une nouvelle enquête-ménage sur les voyages touristiques pour recueillir des renseignements sur les voyages nationaux et internationaux des résidents canadiens. Cette nouvelle enquête, l'Enquête nationale sur les voyages (ENV), devait répondre aux exigences du Système de comptabilité nationale, être autonome sur le plan du financement, avoir la souplesse nécessaire pour l'ajout de nouveau contenu et d'échantillons supplémentaires par suite d'ententes de recouvrement des coûts, et améliorer la qualité des données sur les voyages à l'étranger des Canadiens qui participent à l'Enquête sur les voyages internationaux. Les résultats de l'étude de faisabilité appuient clairement la création de la nouvelle enquête.

La nouvelle ENV remplace l'Enquête sur les voyages des résidents du Canada (voyages intérieurs) et la composante des résidents canadiens (voyages à l'étranger) de l'Enquête sur les voyages internationaux. Elle a été mise en œuvre en février 2018 pour l'année de référence 2018.

Lors d'un remaniement d'enquête, les ruptures dans les séries chronologiques et les biais potentiels résultant d'un mode de collecte unique sont les principaux enjeux pour les intervenants et les utilisateurs, c'est-à-dire le Système de comptabilité nationale et les partenaires extérieurs.

Les estimations tirées des résultats de l'ENV de 2018 aident à mieux comprendre la rupture dans les séries chronologiques et l'ampleur du biais lorsqu'un seul mode de collecte électronique est utilisé.

Le présent document traite principalement des voyages intérieurs. L'objectif est de donner un aperçu de la méthodologie des deux enquêtes et des différences entre les estimations de 2018 produites au moyen l'ENV et celles de 2017 obtenues de l'Enquête sur les voyages des résidents du Canada (EVRC). Les différences observées dans les estimations peuvent être dues à un changement réel dans la population entre 2017 et 2018. Elles peuvent également s'expliquer par les différences de méthodologie entre les deux enquêtes, ou découler d'une combinaison de ces deux possibilités.

Méthodologie de l'ENV et de l'EVRC

Les sections suivantes décrivent différents aspects de la méthodologie employée pour les deux enquêtes.

Plan d'échantillonnage

Les deux enquêtes sont des enquêtes-ménages utilisant la même base de sondage de logements canadiens. L'EVRC est un supplément à l'Enquête sur la population active (EPA), tandis que l'ENV est une enquête indépendante. Lorsque l'ENV a été élaborée, le plan d'échantillonnage de l'enquête a été fondé en grande partie sur celui de l'EPA, puisqu'il s'agit d'un plan éprouvé et efficace pour une enquête-ménage. En raison de la similarité des plans d'échantillonnage de l'EVRC et de l'ENV, on ne s'attend pas à ce que le plan d'échantillonnage soit un facteur contribuant aux différences entre les deux enquêtes.

Population cible et couverture

En matière de couverture, il n'y a aucune différence entre les populations cibles des deux enquêtes. La population cible est constituée des résidents des dix provinces canadiennes qui sont âgés de 18 ans et plus. Sont exclus les personnes qui vivent dans les réserves et autres établissements autochtones dans les provinces, les membres à temps plein des Forces canadiennes et les personnes qui vivent en établissement.

L'EVRC est un supplément l'EPA, et cette dernière couvre la population des résidents canadiens âgés de 15 ans et plus. Toutefois, le groupe d'âge sélectionné pour l'EVRC se limite aux personnes de 18 ans et plus. Les deux populations cibles sont donc essentiellement les mêmes.

Mode de collecte et taux de réponse

Pour ce qui est de la non-réponse, il existe une différence importante entre les deux enquêtes qui pourrait expliquer l'écart entre les taux de réponse (plus de détails sur les taux sont fournis au paragraphe suivant). Dans le cas de l'EVRC, il s'agit d'un supplément à l'EPA. Même l'EVRC n'est pas obligatoire, il est possible que l'enquête ait tiré avantage de la nature obligatoire de l'EPA, ce qui aurait donné lieu à un taux de réponse plus élevé. Comme les répondants ont déjà participé à l'EPA, les chances sont plus élevées qu'ils participent également à l'EVRC. De plus, le taux de réponse de l'EVRC a probablement bénéficié du processus de collecte de données de l'EPA, qui comprend des interviews téléphoniques assistées par ordinateur (ITAO) et un questionnaire électronique. D'autre part, l'ENV n'est pas une enquête obligatoire et ne bénéficie pas d'un premier contact en personne comme l'EPA. Le premier contact se fait par lettre, et il n'y a pas d'ITAO de suivi (en raison de contraintes budgétaires), ce qui entraîne un taux de réponse plus faible.

En tant que supplément à l'EPA, l'EVRC a obtenu un taux de réponse annuel d'environ 75 %, c'est-à-dire qu'environ 105 000 ménages ont été sélectionnés pour l'EVRC, et qu'environ 80 000 ménages ont répondu au cours de l'année. Par contre, le taux de réponse annuel à l'ENV est de 28 %. Avec un échantillon d'environ 30 000 ménages par mois, la taille de l'échantillon annuel est d'environ 360 000 ménages, dont environ 108 000 sont des ménages répondants.

Pour les deux enquêtes, un facteur d'ajustement pour les non-répondants est prévu. Les ajustements font partie du processus de pondération. Pour traiter la non-réponse, des catégories de non-réponse dans chaque province sont créées pour l'EVRC. Ces catégories sont réalisées en modélisant les probabilités de réponse à l'aide de variables démographiques et en utilisant un algorithme de regroupement pour créer des catégories dont les probabilités de réponse sont semblables. Pour traiter la non-réponse à l'ENV, un processus semblable est utilisé. Les trois variables démographiques pour l'ENV (groupe d'âge, sexe, revenu du ménage) font partie de la modélisation des probabilités de réponse.

L'ENV n'inclut pas d'ITAO de suivi. Il serait possible d'explorer plus en détail les différences entre les répondants et les non-répondants en ajoutant de telles interviews avec les non-répondants, et possiblement confirmer ou infirmer nos hypothèses sur les non-répondants. De plus, si d'autres facteurs ont une incidence sur la non-réponse, le suivi pourrait être adapté à des unités particulières qui pourraient avoir une incidence plus grande sur les estimations.

Traitement des données

Puisque le système de traitement de l'ENV est largement fondé sur celui de l'EVRC, rien n'indique que les erreurs de traitement seraient à l'origine de divergences importantes entre l'EVRC et l'ENV.

Enfin, comme les deux enquêtes utilisent un questionnaire électronique, les erreurs de mesure initiales devraient être semblables entre les répondants à l'EVRC et à l'ENV. Le recours aux ITAO de suivi pour certains des répondants à l'EVRC pourrait réduire les erreurs de mesure, mais cela ne devrait pas être un facteur important. Encore une fois, la tenue d'ITAO de suivis spécifiques et ciblées pourrait nous permettre de savoir si nos concepts causent des problèmes aux répondants.

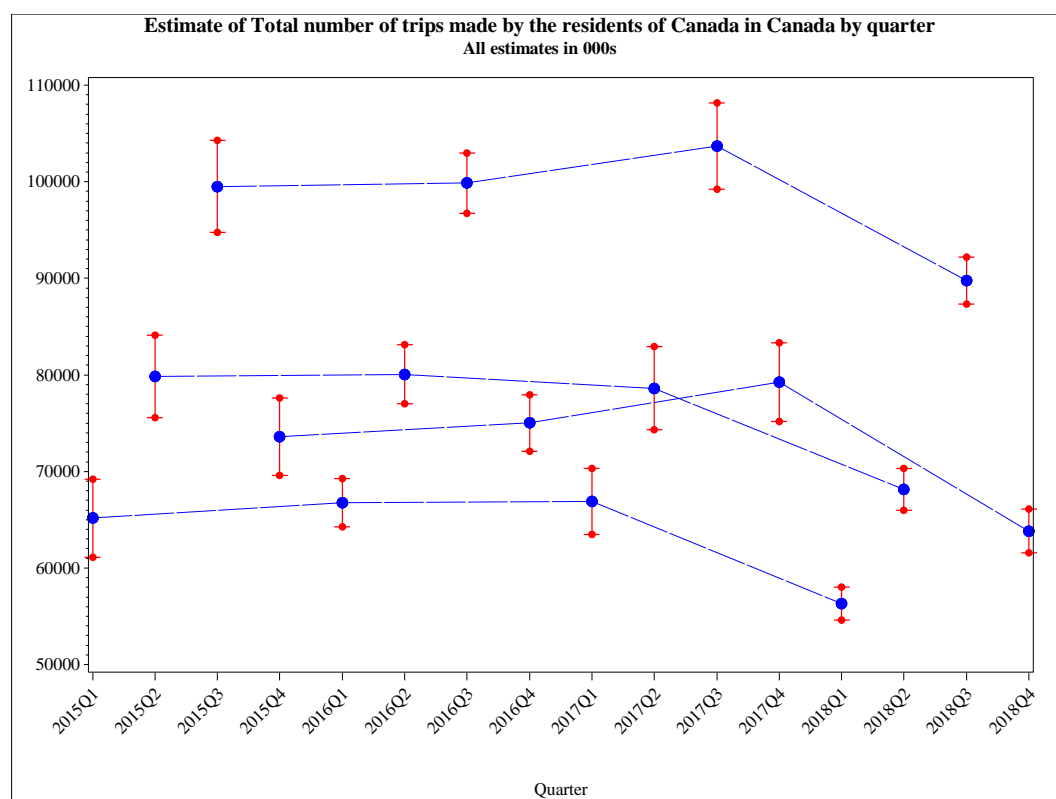
Comparaison détaillée des estimations

Une comparaison détaillée entre l'ENV et l'EVRC a été effectuée. L'analyse compare trois variables principales : le nombre total de nuitées, les dépenses totales et le nombre total de voyages. L'analyse a été effectuée en tenant compte des coefficients de variation et des intervalles de confiance construits autour des estimations.

Pour les estimations du nombre total de voyages, les estimations de l'ENV de 2018 sont inférieures à celles de l'EVRC de 2017, et très différentes de celles-ci. Elles sont également inférieures à celles de l'EVRC de 2015 et de 2016.

Résultats des voyages intérieurs : Nombre total de voyages

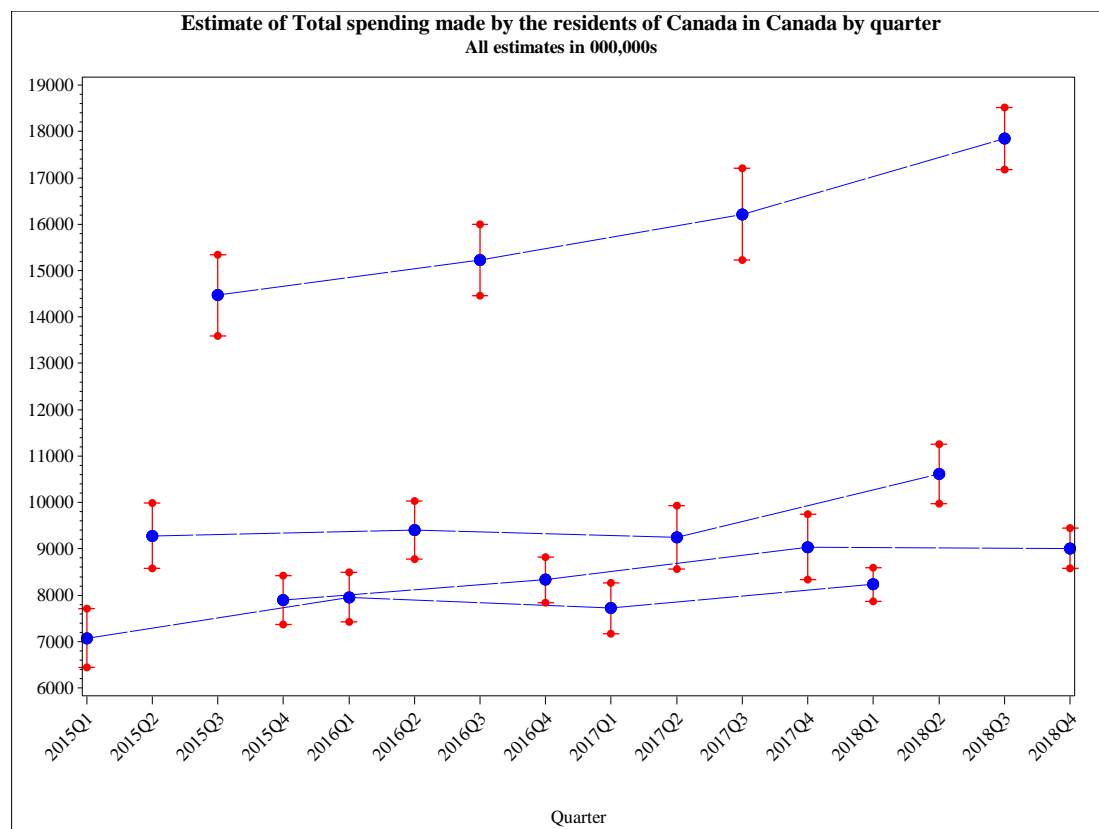
*Nota : Les résultats du 1^{er} trimestre de 2015 au 4^e trimestre de 2017 sont obtenus de l'EVRC.
Les résultats du 1^{er} au 4^e trimestre de 2018 sont obtenus de l'ENV.*



Pour les estimations des dépenses totales, il n'y a pas de différence statistique entre les estimations de l'ENV de 2018 et celles de l'EVRC de 2017. Les estimations montrent une augmentation qui n'est pas statistiquement différente, étant donné que les intervalles de confiance se chevauchent.

Résultats des voyages intérieurs : Dépenses totales

*Note : Les résultats du 1^{er} trimestre de 2015 au 4^e trimestre de 2017 sont obtenus de l'EVRC.
Les résultats du 1^{er} au 4^e trimestre de 2018 sont obtenus de l'ENV.*

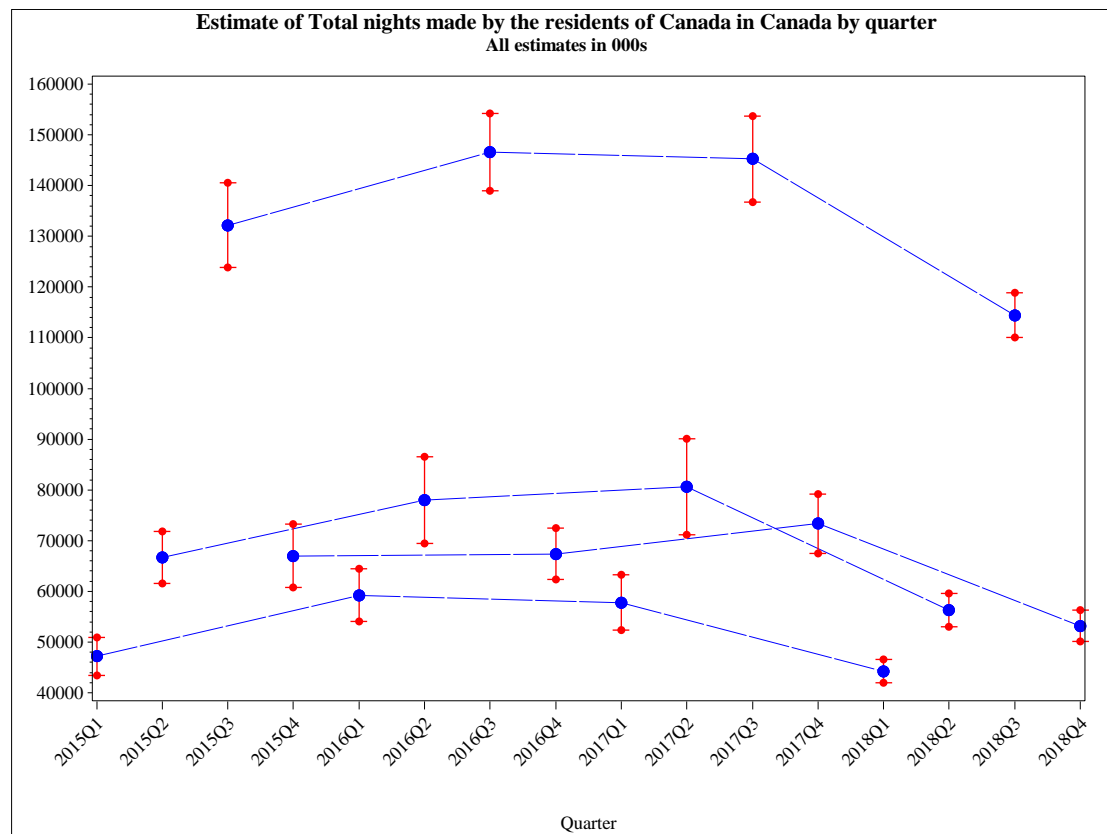


Pour les estimations du nombre total de nuitées, les estimations de l'ENV de 2018 pour chaque trimestre sont inférieures à celles de l'EVRC de 2017, et très différentes de celles-ci. Cependant, elles semblent correspondre à celles de l'EVRC de 2015.

Résultats des voyages intérieurs : Nombre total de nuitées

Nota : Les résultats du 1^{er} trimestre de 2015 au 4^e trimestre de 2017 sont obtenus de l'EVRC.

Les résultats du 1^{er} au 4^e trimestre de 2018 sont obtenus de l'ENV.



Une analyse plus poussée des variables démographiques de la population de répondants pour l'EVRC et l'ENV sera effectuée dans un proche avenir.

Conclusion

L'Enquête sur les voyages des résidents du Canada et l'Enquête nationale sur les voyages sont deux enquêtes distinctes, comme en atteste leur description. Bien qu'elles ciblent la même population, elles donnent des résultats différents, dont certains sont très différents (le nombre total de voyages, par exemple) tandis que d'autres ne le sont pas (les dépenses totales). Il est impossible, à l'heure actuelle, de déterminer si ces différences sont le résultat de dissimilitudes entre les deux enquêtes ou si elles reflètent des changements réels dans la population entre 2017 et 2018. Les utilisateurs doivent faire preuve de prudence lorsqu'ils comparent les résultats de l'EVRC et de l'ENV en raison des changements apportés à la méthodologie entre les deux enquêtes. Statistique Canada ne recommande pas de comparaisons entre les deux enquêtes.