

Enquête canadienne sur la nature 2012 – Fichier de microdonnées à grande diffusion

Guide de l'utilisateur des données

Nous vous remercions de l'intérêt que vous portez aux résultats de l'*Enquête canadienne sur la nature 2012*.

Nous vous encourageons à [lire le rapport](#) sur les résultats de l'*Enquête canadienne sur la nature 2012* qu'il est possible de télécharger gratuitement à partir de www.biodivcanada.ca.

Veillez lire particulièrement la section intitulée **Méthodes d'enquête** qui décrit en détail la méthode d'échantillonnage, le nombre de répondants selon la catégorie géographique et démographique, le processus de collecte des données, les facteurs de pondération appliqués aux différents échantillons ainsi que les conseils sur l'utilisation des données pour pousser l'analyse plus loin.

Le **questionnaire** de l'*Enquête sur la nature de 2012* est fourni dans un fichier distinct et reproduit dans le rapport ci-dessus à l'Annexe B : Instrument d'enquête.

Application de facteurs de pondération aux enregistrements contenus dans l'ensemble de données

Chaque enregistrement contenu dans l'ensemble de données contient **deux (2)** facteurs de pondération analytique : il **faut** en utiliser un dans l'analyse des données.

Les réponses à l'*Enquête canadienne sur la nature 2012* ont été pondérées afin de garantir que les estimations de l'enquête sont représentatives de la population du Canada. Deux ensembles de pondérations ont été calculés : une pondération « par adresse » pour l'analyse des données provenant de l'échantillon probabiliste par adresse et une seconde pondération d'analyse « combinée » pour permettre l'inclusion des réponses des participants sur le Web et des échantillons facultatifs. Tant la pondération de l'échantillon par adresse et que celle de l'échantillon combiné permettront d'estimer des totaux et des distributions de population semblables avec les variables utilisées dans les ajustements de la calibration. La différence entre la pondération de l'échantillon par adresse et celle de l'échantillon combiné est que la première s'appuie sur un échantillon probabiliste et qu'on peut s'y fier pour produire des estimations et des analyses visant à tirer des conclusions au sujet de la population canadienne avec une précision quantifiable à l'aide des procédures normalisées. La pondération combinée fournit des estimations représentatives avec une précision qui ne peut pas être estimée à l'aide des procédures normalisées. Par conséquent, on peut estimer les variances de l'échantillon par adresse, mais, compte tenu des méthodologies statistiques types, pas celle de l'échantillon combiné.

Les deux facteurs de pondération permettent d'effectuer deux types généraux d'analyse. Il est recommandé que le choix du facteur de pondération repose sur les buts de l'utilisateur :

- le facteur de pondération de l'échantillon fondé sur les adresses devrait servir aux analyses qui exigent des intervalles de confiance (p. ex., estimations de la population) ou un test statistique;
- le facteur de pondération de l'échantillon combiné pourrait servir aux analyses qui n'exigent pas d'intervalle de confiance ou de test statistique (qui exigent des estimations des écarts). Ces données fournissent un ensemble représentatif de réponses d'adultes canadiens dont il est impossible de quantifier la fiabilité statistique.

Pour en savoir plus, s'il vous plaît voir la section intitulée Méthodes d'enquête dans le rapport *Enquête canadienne sur la nature 2012 : connaissances, participation et dépenses liées aux activités récréatives, de conservation et de subsistance axées sur la nature*.

Pondération de l'échantillon par adresse

L'approche de pondération pour l'échantillon par adresse est composée de deux volets : une pondération d'échantillonnage, qui pondère pour des probabilités de sélection inégales dues à l'échantillonnage stratifié; et un ajustement de la calibration, qui pondère pour les personnes qui n'ont pas répondu à l'enquête et qui calibre les pondérations selon les totaux connus de

populations.¹ La composante « échantillonnage » de la pondération par adresse est l'inverse de la probabilité de choisir une résidence donnée à partir du fichier d'adresses de Postes Canada dans chaque strate de l'échantillonnage. La pondération d'échantillonnage a ensuite été calibrée (ou post-stratifiée) selon les totaux connus des populations. Cet ajustement a été effectué en fonction des dimensions suivantes :

- âge (de façon indépendante au sein de chaque province et territoire);
- sexe (de façon indépendante au sein de chaque province et territoire);
- milieu urbain ou rural (de façon indépendante au sein de chaque province et territoire);
- Canadien autochtone ou non autochtone.

Le résultat de cet ajustement est la pondération finale de l'échantillon par adresse.

Pondération de l'échantillon combiné (méthode des échantillons mixtes)

L'approche de pondération de l'échantillon combiné comporte deux composantes : une composante « échantillonnage » fondée sur une estimation de la probabilité de sélection de chaque répondant, et un ajustement de la calibration, qui ajuste la première composante selon les totaux connus des populations. La probabilité de sélection inconnue des participants sur le Web et des répondants des échantillons facultatifs qui répondent à l'enquête est estimée grâce à une modélisation en trois étapes de la propension à l'inclusion dans ces échantillons. Premièrement, la probabilité de sélection de l'adresse pour chaque cas a été estimée comme si elle était choisie à l'aide du mécanisme d'échantillonnage par adresse. C'est la même chose que la probabilité par adresse pour les répondants de l'échantillon par adresse.

Deuxièmement, la probabilité d'être sélectionné pour l'échantillon de participants sur le Web² a été estimée à l'aide de la probabilité modélisée d'être inclus dans les données du groupe. Ainsi, le fait de répondre dans le cadre de l'échantillon non probabiliste est considéré comme un cas d'inclusion dans cet échantillon. La modélisation s'appuie sur une régression logistique visant à prédire la propension d'inclusion selon l'emplacement géographique et la plupart des réponses à l'enquête. Dans les provinces où un échantillon par adresse et un échantillon de participants sur le Web ont été utilisés, la régression logistique a permis de combiner les répondants sur le Web et les répondants de l'échantillon par adresse qui ont répondu sur le Web. Les variables explicatives incluaient les données des réponses et les données démographiques obtenues dans les réponses au questionnaire de même que les variables de stratification. La régression logistique donne une probabilité d'être choisi dans le groupe pour chaque répondant (pas seulement ceux qui ont répondu à l'enquête sur le Web).

Enfin, la troisième étape tenait compte des deux probabilités de sélection désormais existantes (propension de l'échantillon par adresse et des participants sur le Web). Les deux probabilités ont été combinées à l'aide d'une procédure appelée accumulation de cas.^{3,4} Cette procédure calcule la probabilité qu'un répondant soit sélectionné pour l'échantillon par adresse,

¹ Les chiffres de la population par âge (pour les personnes de 18 ans et plus) et par sexe proviennent du recensement canadien de 2011. (Statistique Canada. 2012. Recensement du Canada de 2011 : Tableaux thématiques : Âge et sexe pour la population du Canada, provinces, territoires, divisions de recensement, subdivisions de recensement et aires de diffusion, Recensement de 2011. Ottawa (Ont.), Statistique Canada [consulté le 20 juin 2013]. Numéro au catalogue : 98-311-XCB2011018. Sur Internet : <http://www12.statcan.gc.ca/census-recensement/2011/dp-pd/tbt-tt/Ap-fra.cfm?LANG=F&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=0&GID=0&GK=0&GRP=1&PID=101998&PRID=0&PTYPE=101955&S=0&SHOWALL=0&SUB=0&Temporal=2011&THEME=88&VID=0&VNAMEE=&VNAMEF=>). Les chiffres des populations des régions urbaines et rurales ont été obtenus de Gordon Dewis, Statistique Canada (communication personnelle par courriel en juin 2013). Pour une analyse plus approfondie des centres de population et des régions rurales, voir Statistique Canada, « De régions urbaines à centres de population », 7 février 2011 (consulté le 21 juin 2013). Sur Internet : <http://www.statcan.gc.ca/subjects-sujets/standard-norme/sgc-cgt/notice-avis/sgc-cgt-06-fra.htm>. La population autochtone selon l'identité autochtone a été obtenue de Statistique Canada, Tableau : Population ayant une identité autochtone selon les groupes d'âge, l'âge médian et le sexe, chiffres de 2006 pour les deux sexes, pour le Canada, les provinces et les territoires, sur Internet : <http://www12.statcan.ca/census-recensement/2006/dp-pd/hlt/97-558/pages/page.cfm?Lang=F&Geo=PR&Code=01&Table=1&Data=Count&Sex=1&Age=1&StartRec=1&Sort=2&Display=Page>. Tous les totaux ont été ajustés selon les données de 2011.

² Les processus de sélection pour l'inclusion dans le groupe général de participants sur le Web et pour la participation à la présente enquête sont des processus propres au tiers fournisseur d'enquête sur le Web.

³ Pedlow, Steven, Kanru Xia et Michael Davern. 2010. « Dual-Frame Sample Sizes (RDD and Cell) for Future Minnesota Health Access Surveys », dans *Proceedings of the Survey Research Methods Section*, American Statistical Association, p. 2279-2288. Sur Internet : http://www.amstat.org/sections/srms/proceedings/v2010/Files/307156_57962.pdf.

l'échantillon de propension des participants sur le Web ou les deux. La composante « échantillonnage » de la pondération de l'échantillon combiné devient l'inverse de cette probabilité de sélection estimée. Pour l'enquête facultative des Territoires du Nord-Ouest et l'enquête facultative de communautés du Nunavut, on a présumé que la probabilité de sélection était constante. Pour produire la pondération de l'analyse combinée, la composante « échantillonnage » de la pondération de l'échantillon combiné a ensuite été calibrée (ou post-stratifiée) selon les totaux connus des populations, à l'aide de l'approche décrite ci-dessus pour la pondération de l'échantillon par adresse pour les enquêtes sur le Web.

Dictionnaire de données

Un dictionnaire de données est fourni dans un fichier distinct. Le dictionnaire de données définit les codes abrégés utilisés pour les zones et les valeurs qui figurent dans l'ensemble de données et, le cas échéant, indique le nombre de réponses aux différentes options contenues dans les questions. Il convient de signaler que les totaux sont basés sur la combinaison des échantillons selon les adresses et des échantillons de sites Web.

Comparaison avec des résultats publiés auparavant

L'analyse des données peut produire des résultats différents de ceux qui sont publiés dans le rapport de l'*Enquête canadienne sur la nature 2012 : connaissances, participation et dépenses* à cause de différences au niveau de la méthode ou de légères différences au niveau des données qui découlent de la stratégie de contrôle de la divulgation appliquée aux données (voir ci-dessous).

Contrôle de la divulgation

Le contrôle de la divulgation est un processus appliqué à un ensemble de données pour protéger la confidentialité des participants. Les méthodes suivantes de contrôle de la divulgation ont été appliquées à la production du fichier de microdonnées à grande diffusion de manière à assurer la confidentialité des participants à l'enquête :

- les codes postaux ont été supprimés de tous les enregistrements;
- l'année de naissance a été recodée en groupes d'âge de cinq ans pour tous les enregistrements;
- les enregistrements tirés de l'échantillon de la collectivité du Nunavut ont été supprimés (57 enregistrements). Veuillez consulter la partie du rapport intitulée **Méthodes d'enquête** qui présente de l'information sur l'échantillon de la collectivité du Nunavut.
- Les réponses texte aux questions ouvertes qui invitaient les participants à fournir d'autres renseignements s'ils ont choisi « Autre, veuillez préciser » ont été supprimées.
- En outre, 19 enregistrements des membres du forum électronique ont été supprimées en raison de réponses parasites.

⁴ O'Muircheartaigh, Colm, et Steven Pedlow. 2002. « Combining samples vs. cumulating cases: A comparison of two weighting strategies in NLSY97 », dans *Proceedings of the Survey Research Methods Section*, American Statistical Association, p. 2557-2562. Sur Internet : <http://www.amstat.org/sections/srms/proceedings/y2002/Files/JSM2002-001082.pdf>.