

Estimation

Dans une enquête par sondage, chaque répondant est représentatif de lui-même, mais aussi des autres personnes qui n'ont pas été échantillonnées. Par conséquent, un poids est lié à chaque répondant, afin d'indiquer le nombre de personnes que ce répondant représente. Ce poids doit être utilisé pour toutes les estimations.

Pondération (Phase I – Dans les communautés et hors des communautés)

Les poids ont été calculés au moyen d'un processus en trois étapes. La PREMIÈRE étape a consisté à attribuer un poids initial fondé sur le plan d'échantillonnage. Celui-ci correspondait simplement à l'inverse de la probabilité d'inclusion (probabilité d'appartenir à l'échantillon).

Dans le cas de la partie de la phase I concernant les personnes hors des communautés, le poids initial était le produit de deux composantes : l'inverse de la fraction d'échantillonnage de l'unité primaire d'échantillonnage (appelée poids de l'UPE) et le poids du recensement. Trois bases indépendantes pour les Indiens de l'Amérique du Nord, les Métis et les Inuits ont été élaborées pour sélectionner l'échantillon. Les UPE ont été constituées de façon indépendante pour chaque base de sondage, et trois échantillons indépendants ont été sélectionnés. Cela signifie que certaines personnes qui avaient des identités autochtones multiples figuraient dans plus d'une base de sondage. Elles avaient donc plus d'une chance d'être sélectionnées. Étant donné que l'on avait besoin d'un poids d'enquête unique, on a rajusté les probabilités de sélection, afin de tenir compte des probabilités d'inclusions multiples.

Après ce calcul, les personnes sélectionnées par erreur¹ et celles manquées au moment de la sélection de l'échantillon ont été prises en compte, et des rajustements de poids appropriés ont été appliqués au poids initial.

Pour l'échantillon de la phase I des personnes vivant dans les communautés, étant donné que la détermination de la taille de l'échantillon et que la sélection de l'échantillon comportaient des opérations manuelles,

¹ Étant donné la nécessité de l'établissement d'une liste manuelle pour sélectionner l'échantillon, des erreurs se sont produites. Parfois, des personnes qui n'auraient pas dû être incluses figuraient dans la liste, et d'autres qui auraient dû y figurer, en étaient absentes. Des méthodes de sélection d'échantillon ont été mises à l'essai, à partir des questionnaires de l'essai national du Recensement de 1998, afin de déterminer les difficultés qui pourraient se poser et de réduire les erreurs.

par opposition à des calculs informatiques, les poids initiaux ont été calculés en divisant le nombre pondéré d'adultes et d'enfants dans chaque communauté (calculé à partir du recensement) par le nombre de personnes de l'échantillon.

La DEUXIÈME étape du processus de pondération a pris la forme d'un ajustement pour la non-réponse. Deux ajustements ont été effectués pour tenir compte du fait que les non-répondants pouvaient être classés en deux catégories très différentes : les personnes non contactées et les personnes contactées qui n'ont pas répondu. Les poids ont été rajustés tout d'abord pour les cas de non-contact, puis pour la non-réponse. L'ajustement pour la non-réponse a été effectué à partir de catégories, de façon à ce que les enregistrements de chaque catégorie comportent des probabilités de réponse similaires. Les probabilités de réponse estimatives ont été obtenues au moyen d'un modèle de régression logistique, afin de prévoir la probabilité de réponse à partir de variables explicatives.

De nombreuses variables explicatives ont pu être utilisées, du fait que toutes les données du formulaire détaillé du recensement étaient disponibles pour chaque répondant et non-répondant. On a eu recours à des modèles distincts pour les enfants et les adultes. Environ dix catégories de la même taille environ ont été obtenues pour chaque modèle de régression logistique. L'inverse du taux de réponse pondéré dans une catégorie a servi comme facteur de correction pour cette catégorie, et les poids initiaux des répondants à l'intérieur de la catégorie ont été rajustés en conséquence.

La TROISIÈME étape du rajustement de pondération a pris la forme d'une stratification a posteriori et a permis de faire en sorte que la somme des poids finaux pour les répondants soit égale aux chiffres de population du recensement. L'ajustement a été effectué pour des groupes (appelés post-strates) définis sur la base d'une combinaison de plusieurs variables.

Dans le cas de l'EAPA, on a procédé à deux stratifications a posteriori consécutives. La première a reposé (notamment) sur les réponses de chacun des répondants aux questions 17, 18, 20 et 21 du recensement (questions filtres du recensement), pour créer des post-strates. Les poids, qui ont été corrigés pour la non-réponse, ont été rajustés à partir du ratio des chiffres du recensement et des chiffres de l'échantillon pour chaque post-strate. Cette première stratification a posteriori visait à faire en sorte que l'échantillon n'entraîne pas une sous-représentation ou une sur-représentation des groupes autochtones du recensement. Étant donné que les réponses aux questions filtres peuvent différer entre l'EAPA et le recensement, une deuxième stratification a posteriori a été effectuée, afin de garantir que l'ensemble de la population autochtone, estimée à partir des questions filtres de l'EAPA, corresponde à celle estimée à partir des

questions filtres du recensement. On n'a pas effectué de rajustement selon le groupe autochtone, mais plutôt pour le total de la population autochtone (d'identité ou d'origine).

Pondération et correction pour la non-réponse (Phase II – Hors des communautés)

Le processus de pondération utilisé à la phase II était identique à celui de la phase I, sauf pour un point. Étant donné que la sélection a été effectuée directement à partir de la base du recensement (c.-à-d. qu'une liste manuelle n'était pas nécessaire), aucun rajustement n'était requis pour les personnes sélectionnées par erreur ou pour les personnes qui avaient été manquées.

En résumé, la première étape a consisté à attribuer un poids initial fondé sur le plan d'échantillonnage. Le poids initial correspondait simplement au produit du poids du recensement et du poids du sous-échantillon. La deuxième étape a pris la forme d'un rajustement tenant compte de la non-réponse. La méthode utilisée était identique à celle ayant servi à la phase I et a consisté à établir les groupes de pondération à partir de modèles de régression logistique. Des modèles ont été élaborés pour les cas de non-contact et les refus, séparément pour les adultes et les enfants, selon les variables explicatives disponibles.

STRATIFICATION A POSTERIORI

Première stratification a posteriori (Dans les communautés)

Dans le cas de l'échantillon des personnes vivant dans les communautés, outre les questions filtres du recensement sur l'identité, trois autres variables ont été utilisées pour créer les post-strates : les communautés autochtones, les groupes d'âge et le sexe. Dans le cas des communautés où le groupe autochtone dominant représentait 90 % de la population et plus, le rajustement a été effectué au niveau de la communauté grâce à un croisement des variables du groupe d'âge et du sexe. Pour toutes les autres communautés, les rajustements ont été effectués pour le groupe dominant et pour le reste de la communauté. Pour ces deux portions, les post-strates ont aussi été définies grâce au croisement des variables du groupe d'âge et du sexe. Lorsque les communautés ont été rajustées globalement et comprenaient 200 répondants ou plus, cinq groupes d'âge ont été utilisés (0-4, 5-14, 15-24, 15-39, 40 et plus). Autrement, on a utilisé trois groupes d'âge (0-14, 15-39, 40 et plus).

Première stratification a posteriori (Hors des communautés, Phase I et Phase II combinées)

Pour la composante des personnes vivant hors des communautés, les post-strates ont été définies à partir des niveaux géographiques infraprovinciaux (qui correspondent étroitement aux strates des personnes vivant hors des communautés – Phase I), de la variable enfant/adulte, ainsi que de l'identité et de l'ascendance autochtone selon le recensement. Les stratifications a posteriori de « l'identité » et de « l'ascendance seulement » ont été effectuées indépendamment. Tout d'abord, on a utilisé le même niveau géographique pour les deux composantes de la définition de la post-strate. Au niveau de l'identité ou de l'ascendance autochtone, les groupes d'IAN, les Métis, les Inuits, les IAN multiples et les Métis, ainsi que les autres multiples (tous les types multiples combinés, à l'exclusion des IAN et des Métis) ont servi à définir les post-strates. Dans certains cas, des agrégations ont été nécessaires. Par exemple, à Terre-Neuve-et-Labrador, la stratification a posteriori pour le rajustement de l'identité a été effectuée pour l'ensemble de la province plutôt que pour la région infraprovinciale (rurale/urbaine). Par ailleurs, dans certains domaines, les enfants et les adultes ont été combinés au moment de la constitution des post-strates.

Deuxième stratification a posteriori

Une deuxième stratification a posteriori a été effectuée pour rajuster les chiffres obtenus à partir des questions filtres de l'EAPA en fonction des chiffres du recensement. Cette stratification a posteriori a été effectuée selon le nombre total d'Autochtones, et non pas selon chaque groupe autochtone, afin de ne pas cacher les transitions observées entre l'enquête et le recensement. En fait, ces transitions montrent que le statut d'autochtone, comme concept, peut être affecté par le contexte de l'enquête, la méthode de collecte des données et les réponses obtenues par personne interposée.

Cette stratification a posteriori a été effectuée pour tous les répondants de la Phase I et de la Phase II combinées. Ce rajustement a été effectué pour chaque domaine géographique (infraprovincial : urbain, rural ou RMR) et pour l'âge (enfant ou adulte), pour la composante de l'enquête qui touche les personnes vivant hors des communautés. Dans le cas de la composante des personnes vivant dans les communautés, le rajustement a été effectué séparément pour les adultes et les enfants, ainsi que pour chaque communauté pour lesquelles on prévoit publier un profil. Les autres communautés pour lesquelles on ne prévoit pas publier de profil ont été regroupées pour cette stratification a posteriori.

Estimation de la variance

Les estimations qui peuvent être calculés à partir de cette enquête sont fondées sur un échantillon de personnes. Des estimations assez différentes peuvent être obtenues si l'on a procédé à un recensement complet au moyen des mêmes questionnaires, intervieweurs, superviseurs, méthodes de traitement, etc. que ceux réellement utilisés. La différence entre les estimations obtenues à partir de l'échantillon et celles découlant d'un dénombrement complet effectué dans des conditions similaires est appelée erreur d'échantillonnage des estimations.

Afin de fournir des estimations de l'erreur d'échantillonnage pour les statistiques produites dans le cadre de l'EAPA, on a eu recours à la méthode d'auto-amorçage (bootstrap). Celle-ci, qui est une méthode de rééchantillonnage, consiste à sélectionner M sous-échantillons à partir de l'échantillon principal et à produire des estimations pour chaque sous-échantillon.

L'estimation de la variance d'auto-amorçage (la variance constitue une mesure particulière de l'erreur d'échantillonnage) a été calculée comme correspondant à la variance empirique des M estimations. Pour chaque sous-échantillon, les poids initiaux ont d'abord dû être rajustés pour le sous-échantillonnage d'auto-amorçage, avec pour résultat ce que l'on appelle des « poids initiaux d'auto-amorçage ». Ces derniers ont par la suite été rajustés selon la méthode décrite précédemment, en vue d'obtenir les poids définitifs pour chaque sous-échantillon. Autrement dit, on a procédé à des rajustements pour tenir compte des personnes manquées ou échantillonnées par erreur, de la non-réponse et de la stratification a posteriori, pour chaque sous-échantillon, presque exactement de la même façon que pour l'échantillon principal. Dans le cas de l'EAPA, on a sélectionné 500 échantillons d'auto-amorçage.

Chaque sous-échantillon a été sélectionné en calquant le plan d'échantillonnage de l'échantillon principal. Toutefois, les unités ont été sélectionnées *avec remise*, par opposition à *sans remise*. Pour la partie de l'enquête touchant les personnes vivant hors des communautés, chaque strate a été subdivisée en deux composantes selon que les unités primaires d'échantillonnage (UPE) avaient été sélectionnées avec certitude (UPE à tirage complet) ou non (UPE à tirage partiel). Dans la composante à tirage partiel de chaque strate, un échantillon aléatoire simple *avec remise* de $n-1$ UPE parmi n UPE a été sélectionné pour chaque échantillon d'auto-amorçage. Dans ce cas, n correspond au nombre d'UPE à tirage partiel de l'échantillon original pour cette strate. Ce choix particulier de $n-1$ a comme avantage qu'il simplifie la formule pour les poids d'auto-amorçage.

Dans la composante à tirage complet de chaque strate, ainsi que dans celle des personnes vivant dans les communautés, un échantillon aléatoire simple *avec remise* de $m-1$ personnes parmi m personnes a été sélectionné. Dans ce cas, m correspond au nombre de personnes de l'échantillon initial comprises dans la composante à tirage complet de la strate, ou à la taille de l'échantillon de chaque strate de personnes vivant dans les communautés. Cela rend compte du fait que, pour la composante à tirage complet de la strate, l'échantillon initial peut être perçu comme un échantillon à un degré de personnes, par opposition à un échantillon à deux degrés d'unités primaires d'échantillonnage (UPE).

Dans le cas de la portion de l'enquête comprenant les personnes vivant hors des communautés, trois bases indépendantes d'Indiens de l'Amérique du Nord, de Métis et d'Inuits ont été élaborées pour sélectionner l'échantillon principal. Les UPE ont été constituées indépendamment dans chaque base, et trois échantillons indépendants ont été sélectionnés. Pour la méthode d'auto-amorçage, on a eu recours au même plan d'échantillonnage. Par conséquent, 500 échantillons d'auto-amorçage ont été sélectionnés dans chacune des trois bases de sondage. Comme pour l'échantillon de l'enquête, les poids d'auto-amorçage ont été rajustés, afin que les personnes ayant des identités/origines multiples aient un poids d'auto-amorçage unique dans chacune des 500 répétitions.