

Enquête nationale sur la santé de la population
Volet ménages
Cycle 1 à 7 (1994-1995 à 2006-2007)
Fichier fictif pour le télé-accès au fichier maître

Décembre 2008

Avertissement

Le fichier fictif ne devrait pas servir à d'autres buts que le développement et la vérification des programmes informatiques devant être soumis par télé-accès. Le fichier fictif comprend des données modifiées et ne doit en aucun cas être utilisé à des fins d'analyse.

1. Introduction et contexte

L'Enquête nationale sur la santé de la population (ENSP), menée par la Division de la statistique de la santé à Statistique Canada, recueille de l'information sur la santé et les caractéristiques socio-économiques de la population canadienne. À ce jour, sept cycles de collecte ont été complétés : l'ENSP Cycle 1 (1994-1995), l'ENSP Cycle 2 (1996-1997), l'ENSP Cycle 3 (1998-1999), l'ENSP Cycle 4 (2000-2001), l'ENSP Cycle 5 (2002-2003), l'ENSP Cycle 6 (2004-2005) et l'ENSP Cycle 7 (2006-2007). Les sept cycles comportent un volet ménages alors que seuls les cycles 1 à 5 comportent un volet établissements de soins de santé. Les cycles 1, 2 et 3 ont aussi un volet Nord.

Les cycles précédents ont produit des fichiers de microdonnées à grande diffusion (FMGD) pour le volet ménages (cycles 1, 2 et 3), ainsi que pour le volet établissements de soins de santé (cycles 1 et 2) afin de permettre un accès élargi aux données transversales. Un seul fichier a été créé pour le volet ménages aux cycles 4, 5, 6 et 7. Il s'agit du fichier longitudinal carré qui contient tous les 17 276 membres du panel. La création d'un FMGD nécessite l'application de procédures rigoureuses afin d'assurer la confidentialité des données. Ainsi, certaines des variables de l'enquête doivent être soit agrégées, soit assujetties à des valeurs maximales ou minimales, soit simplement supprimées. Puisque les règles de confidentialité limitent la création de FMGD de nature longitudinale et que l'ENSP est purement longitudinale depuis le cycle 4, un FMGD n'a pas été créé pour le cycle 7.

Afin d'offrir aux chercheurs la possibilité d'avoir accès au fichier maître, un service de télé-accès a été instauré. Ce service permet aux chercheurs de soumettre des programmes informatiques par courriel à une adresse électronique dédiée (nphs-ensp@statcan.gc.ca) et d'en recevoir les résultats par retour de courriel. Pour se prévaloir du service de télé-accès, les chercheurs doivent nécessairement obtenir l'approbation préalable de la Division de la statistique de la santé. Ils doivent, à cette fin, soumettre une demande d'accès à l'adresse électronique susmentionnée dans laquelle sont clairement indiqués les points suivants :

- l'affiliation du chercheur,
- les noms de tous les chercheurs impliqués dans le projet de recherche,
- le titre du projet de recherche,
- un résumé du projet de recherche,
- les objectifs de la recherche,
- les données requises (enquête, cycle),
- pourquoi la recherche nécessite le recours aux fichiers maîtres plutôt que les FMGD (pour les cycles où un FMGD existe),
- pourquoi le service de télé-accès est choisi au lieu de l'accès sur place dans un centre de données de recherche (CDR),

- les résultats projetés, et
- la date de complétion prévue du projet.

Pour obtenir plus de renseignements sur le processus, veuillez communiquer avec l'équipe de l'ENSP à l'adresse électronique mentionnée précédemment ou par téléphone au (613) 951-1653. Lorsque l'approbation pour le téléaccès a été obtenue, le chercheur peut envoyer ses programmes informatiques à l'équipe de l'ENSP pour leur exécution avec les données du (des) fichier(s) maître(s). L'équipe vérifie les sorties des programmes par rapport aux normes de confidentialité et les transmet au chercheur. Cependant, le bon fonctionnement et l'exactitude de chacun des programmes demeurent en tout temps l'entière responsabilité du chercheur.

Le fichier fictif qui est fourni sur ce CD-ROM permet au chercheur de développer et de vérifier ses programmes informatiques avant de les soumettre. Le fichier longitudinal fictif du cycle 7 porte sur les données recueillies aux cycles 1 à 7 auprès des 17 276 membres du panel longitudinal. Bien que certaines variables à caractère administratif, ne représentant aucun intérêt analytique, aient été recodées à « blanc » ou « 9 », le fichier fictif simule parfaitement le fichier maître. Il renferme les mêmes variables et a le même cliché d'article. Cependant, les valeurs des variables du fichier fictif ont été modifiées afin de préserver la confidentialité des répondants.

Le CD-ROM fournit également les liens du site Web de Statistique Canada vers la documentation et les programmes Bootvar en formats SAS et SPSS en pour le calcul de la variance des estimations. Deux fichiers fictifs de poids bootstrap associés au fichier fictif sont aussi inclus afin de permettre au chercheur de développer et vérifier ses programmes de calcul de variance. Ces fichiers fictifs de poids bootstrap simulent le fichier original, utilisent les mêmes clichés d'article et contiennent les mêmes variables que celui-ci, mais les valeurs des poids sont modifiées par rapport au fichier maître. On retrouve également sur le CD-ROM, des fichiers périphériques d'énoncés SAS et SPSS pour la lecture des fichiers, ainsi que la documentation technique et méthodologique accompagnant normalement les fichiers maîtres. Veuillez consulter la section 3 du présent document pour la liste complète des fichiers de ce CD-ROM.

La prochaine section de ce document décrit de façon plus détaillée les étapes menant à la création du fichier fictif.

2. Création du fichier fictif

2.1 Catégories d'enregistrements

Les enregistrements du fichier maître sont d'abord classés selon l'âge, le sexe et le profil de réponse longitudinale. Un des objectifs consiste à former des catégories regroupant des enregistrements pour lesquels le cheminement à travers le questionnaire est le même, de sorte que, après permutation aléatoire des données à l'intérieur des catégories, chaque enregistrement artificiel résultant contienne des données cohérentes.

Les catégories utilisées pour créer le fichier longitudinal fictif du cycle 7 sont en fonction de l'âge au moment de l'interview en 1994-1995, en 1996-1997, en 1998-1999, en 2000-2001, en 2002-2003, en 2004-2005 et en 2006-2007. Afin d'avoir un nombre suffisant d'enregistrements dans chaque catégorie, on ne tient pas compte des caractéristiques géographiques pour la classification des enregistrements.

Le contenu des questions supplémentaires, posées en Alberta et au Manitoba en 1994-1995, et en Alberta en 1996-1997, est rajouté aux enregistrements du fichier fictif à une étape ultérieure. À ce stade, le problème des catégories trop petites est moindre, puisque le nombre de catégories selon l'âge établi pour le contenu spécial est assez faible. Lorsque le problème se pose malgré tout, on y remédie en utilisant des classes de taille suffisamment large et en corrigeant toute incohérence ayant pu être introduite.

2.2 Blocs de variables

La prochaine étape consiste à créer des blocs de variables. On produit les enregistrements artificiels pour le fichier fictif en permutant de façon aléatoire des blocs de variables entre les enregistrements d'une même catégorie.

Le principe est de former des blocs ayant une signification analytique tout en étant suffisamment petits pour être conformes aux normes de confidentialité des données. Les variables qui, regroupées, permettent d'identifier des personnes sont donc placées dans des blocs différents. De plus, plusieurs variables sont modifiées dans chaque bloc. Toutes les modalités de chaque variable ne se trouvent pas nécessairement sur le fichier fictif. Pour obtenir de l'information sur toutes les modalités possibles pour une variable, les utilisateurs sont priés de se référer à la documentation du fichier maître. Les autres détails relatifs à la création des blocs ne sont pas rapportés car ils pourraient fournir trop de renseignements sur la méthode utilisée pour créer les données fictives et mettraient la confidentialité en péril. Bien que les efforts soient faits afin d'assurer la cohérence entre les blocs, cette cohérence n'est pas nécessairement toujours présente.

2.3 Variables recodées à « blanc » ou « 9 »

Pour fins de confidentialité, les variables énumérées ci-après ont été recodées à « blanc » ou « 9 » lors de la création du fichier fictif :

DOD, SP34_CPS, SP34_MET, STRATUM, REPLICAT, CYCLE, SUBCYCLE, AM68_SHA, AM34_SRC, AM34_LNG, SP34_CPA, AM54_BMM, AM54_BDD, AM54_BYE, AM54_SRC, AM54_LNG, AM54_EMM, AM54_EDD, AM54_EYY, AM64_STA, AM64_BMM, AM64_BDD, AM64_BYE, AM64_SRC, AM64FR, AM64_LNG, AM64_EMM, AM64_EDD, AM64_EYY, AM64_SHA, SP34_STA, AM54_STA, AM64_LNK, AM64_TEL, AM54_PXY, AM64_PXY, AM54_TEL, AM64_ALO, AM64_AFF, GE34DPOP, SP36FOUT, SP36_CPA, AM36_SRC, AM36_LNG, SP36_STA, AM56_STA, AM56_BMM, AM56_BDD, AM56_BYE, AM56_SRC, AM56_LNG, AM56_EMM, AM56_EDD, AM56_EYY, AM66LDUR, AM66_STA, AM66_BMM, AM66_BDD, AM66_BYE, AM66_SRC, AM66FR, AM66_LNG, AM66_EMM, AM66_EDD, AM66_EYY, AM66_SHA, AM66_LNK, AM36_TEL, AM66_TEL, AM56_PXY, AM66_PXY, AM56_TEL, AM66_ALO, AM66_AFF, GE36DPOP, AM68LDUR, SP38FOUT, SP38_CPA, AM38_LP, AM38_SRC, AM38_LNG, SP38_STA, AM58_STA, AM68_STA, AM58_BDD, AM58_BMM, AM58_BYE, AM58_LNG, AM58_EMM, AM58_EDD, AM58_EYY, AM68_BMM, AM68_BDD, AM68_BYE, AM68_SRC, AM68FR, AM68_LNK, AM68_LNG, AM68_EMM, AM68_EDD, AM68_EYY, AM58_SRC, AM38_TEL, AM68_TEL, AM58_PXY, AM68_PXY, AM58_TEL, AM68_ALO, AM68_AFF, GE38DPOP, AM60LDUR, SP30FOUT, SP30_CPA, AM30_PL, SP30_STA, AM30_TEL, AM60_STA, AM60_BMM, AM60_BDD, AM60_BYE, AM60_SRC, AM60FR, AM60_LNK, AM60_SHA, AM60_TEL, AM60_LNG, AM60_PXY, GE30DPOP, AM62LDUR, SP32FOUT, SP32_CPA, AM32_PL, SP32_STA, AM32_TEL, AM62_STA, AM62_BMM, AM62_BDD, AM62_BYE, AM62_SRC, AM62FR, AM62_LNK, AM62_SHA, AM62_TEL, AM62_LNG, AM62_PXY, GE32DPOP, AM6ALDUR, SP3AFOUT, SP3A_CPA, AM3A_PL, SP3A_STA, AM3A_TEL, AM6A_STA, AM6A_BMM, AM6A_BDD, AM6A_BYE, AM6A_SRC, AM6AFR, AM6A_LNK, AM6A_SHA, AM6A_TEL, AM6A_LNG, AM6A_PXY, GE3ADPOP, AM6BLDUR, SP3BFOUT, SP3B_CPA, AM3B_PL, SP3B_STA, AM3B_TEL, AM6B_STA, AM6B_BMM, AM6B_BDD, AM6B_BYE, AM6B_SRC, AM6BFR, AM6B_LNK, AM6B_SHA, AM6B_TEL, AM6B_LNG, AM6B_PXY, GE3BDPOP.

2.4 Sous-échantillonnage

Finalement, un sous-échantillonnage du fichier a été fait afin de préserver la confidentialité. Le sous-échantillonnage consiste à enlever une proportion des 17 276 enregistrements, et de doubler une partie des enregistrements qui restent afin de garder le nombre d'enregistrements égal à 17 276. Le nombre de répondants dans le sous-ensemble complet a aussi été respecté dans ce sous-échantillonnage.

Les poids d'échantillonnage et les poids bootstrap ont ensuite été ajustés conformément aux estimations de la population de 1994 par province et groupe d'âge-sexe. Cependant, il est à noter que ces poids demeurent des poids artificiels.