

Enquête sur la santé dans les collectivités canadiennes
Cycle 1.1, 2000-2001
Fichier fictif pour le télé-accès au fichier maître intégré

Avril 2003

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 sur la santé dans les collectivités canadiennes (ESCC) est une nouvelle enquête menée par Statistique Canada en collaboration avec ses partenaires en vue de recueillir des informations sur la santé, l'état de santé et l'utilisation du système de santé de la population canadienne à un niveau géographique infraprovincial. Les données de l'ESCC sont recueillies sur un cycle biennal composé de deux enquêtes distinctes : une enquête régionale la première année et une provinciale la deuxième année. Jusqu'à maintenant, le premier cycle de collecte a été complété.

Afin de permettre un accès élargi aux données, un fichier de microdonnées à grande diffusion (FMGD) a été produit pour le cycle 1.1. La création d'un FMGD nécessite l'application de procédures rigoureuses afin d'assurer la confidentialité des données. Ainsi, certaines variables de l'enquête doivent être soit agrégées, soit assujetties à des valeurs maximales ou minimales, soit simplement supprimées.

Dans les cas où le FMGD ne peut suffire à ces besoins, un chercheur a comme alternative d'exécuter ses analyses sur le fichier maître intégré de l'ESCC. Afin justement d'offrir aux chercheurs la possibilité d'avoir accès à ce fichier maître, un service de télé-accès a été instauré. Ce service permet au chercheur de soumettre des programmes informatiques par courriel à une adresse électronique dédiée (cchs-escc@statcan.ca) et d'en recevoir les résultats par retour de courriel. Pour se prévaloir du service de télé-accès, le chercheur doit nécessairement obtenir l'approbation préalable de la Division de la statistique de la santé. Il doit, 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 au fichier maître plutôt que le FMGD,
- 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'ESCC à l'adresse électronique mentionnée précédemment ou par téléphone au (613) 951-1653. Une fois l'approbation au télé-accès obtenue, le chercheur peut soumettre ses programmes informatiques à l'équipe de l'ESCC pour le traitement des données du fichier maître intégré. 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 fictif du cycle 1.1 est basé sur les données recueillies auprès de 131 535 répondants. Bien que certaines variables à caractère administratif ne représentant aucun intérêt analytique aient été recodées à « blanc » ou à « 9 », le fichier fictif présente la même structure que le fichier maître intégré; il renferme les mêmes variables et a le même cliché d'article. Cependant, les valeurs des variables et la taille du fichier fictif ont été modifiées afin

de préserver la confidentialité des répondants. Le fichier fictif du cycle 1.1 de l'ESCC contient en fait 65 101 enregistrements, soit environ 50 % de la taille du fichier maître. Il est à noter que la région socio-sanitaire, *Nord du Québec* a été exclue du FMGD pour des raisons de confidentialité alors qu'elle est présente sur le fichier fictif.

Le CD-ROM contient également des programmes SAS et SPSS pour le calcul de la variance des estimations. Quatre fichiers fictifs de poids bootstrap, chacun étant associé à un des quatre poids du fichier maître intégré et étant présent sur le 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 les fichiers originaux, utilisent les mêmes clichés d'article et contiennent les mêmes variables que celui-ci, mais les valeurs des poids sont modifiées par rapport à celles présentes dans les fichiers originaux. Étant donné que le nombre d'enregistrements dans le fichier de données fictives a été réduit, le nombre d'enregistrements dans les fichiers bootstrap fictifs a aussi été réduit en conséquence. 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.

Les prochaines sections de ce document décrivent de façon plus détaillée les étapes menant à la création du fichier fictif.

2. Création du fichier fictif

2.1 Catégories de variables

Les enregistrements du fichier maître intégré sont d'abord classés selon l'âge, le sexe et la région sociosanitaire (pour les modules optionnels seulement). 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 respectant une certaine cohérence interne.

2.2 Blocs de variables

La deuxième 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. On regroupe d'abord les variables en blocs élémentaires correspondant aux diverses sections (modules) du questionnaire. Dans le cas où le contenu d'un bloc élémentaire dépend des réponses obtenues à des questions d'un bloc précédent, les blocs élémentaires respectifs ont été intégrés ensemble afin d'assurer la cohérence interne des enregistrements.

2.3 Échantillonnage et restriction de valeurs des variables

Comme mesure additionnelle, le fichier fictif a été créé en sous-échantillonnant environ 50 % des enregistrements du fichier maître intégré (une fois celui-ci perturbé). De plus, toutes les variables quantitatives ont été tronquées par le haut. Finalement, certaines variables jugées sensibles à la confidentialité ou n'ayant aucune valeur analytique ont quant à elles été tout simplement recodées à « blanc » ou « 9 » (voir sous-section 2.4). Toutes ces modifications font en sorte qu'on ne retrouve pas nécessairement toutes les modalités possibles pour chaque variable sur le fichier fictif. Pour obtenir la liste complète des modalités pour chaque variable, les utilisateurs sont priés de se référer à la documentation du fichier maître intégré. Finalement, les poids d'échantillonnage et les poids bootstrap ont été repoststratifiées selon les mêmes méthodes utilisées lors de la poststratification des poids du fichier maître intégré. Cependant, il est à noter que ces poids demeurent des poids artificiels. Les autres détails relatifs à la création de ce fichier 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.

2.4 Variables recodées à « blanc » ou « 9 »

Les variables énumérées ci-après ont été recodées à « blanc » ou « 9 » lors de la création du fichier fictif :

ADMA_RNO, PERSONID, SAMA_TYP, SAMA_CP, ADMA_PRX, ADMA_ENT, ADMA_IMP, ADMA_N09, ADMA_N10, ADMA_N11, ADMA_N12, ADMA_DAT, ADMA_STA, ADMA_LHH, DHHA_DB, SAMADNUM, ADMAFFE, ADMA_PH, INCAFDN et FINAFDN.

3. Contenu du CD-ROM

LISEZ_MOI.PDF	Documentation en français sur le télé-accès, la création du fichier fictif, le contenu du CD-ROM.
READ_ME.PDF	Documentation en anglais sur le télé-accès, la création du fichier fictif, le contenu du CD-ROM.

Fichiers relatifs à l'utilisation des poids bootstrap pour le calcul de la variance :

BOOTSTRP\	
LISEZ_MOI.TXT	Documentation en français des fichiers relatifs à l'utilisation du programme de variance bootstrap.
READ_ME.TXT	Documentation en anglais des fichiers relatifs à l'utilisation du programme de variance bootstrap.
BOOTSTRP\DATA\	
B5.TXT	Fichier de poids bootstrap en format ASCII.
B5_PEI.TXT	Fichier de poids bootstrap en format ASCII pour le poids spécial de l'Île-du-Prince-Édouard.
B5_Q4.TXT	Fichier de poids bootstrap en format ASCII pour le poids spécial du quatrième trimestre.
B5_BC.TXT	Fichier de poids bootstrap en format ASCII où la Colombie Britannique est subdivisée en 16 régions sociaux-sanitaires.
BOOTSTRP\Layout\	
B5_I.SAS	Cliché d'article SAS pour le fichier de poids bootstrap B5.TXT
B5_I.SPS	Cliché d'article SPSS pour le fichier de poids bootstrap B5.TXT
B5_PEI_I.SAS	Cliché d'article SAS pour le fichier de poids bootstrap B5_PEI.TXT
B5_PEI_I.SPS	Cliché d'article SPSS pour le fichier de poids bootstrap B5_PEI.TXT
B5_Q4_I.SAS	Cliché d'article SAS pour le fichier de poids bootstrap B5_Q4.TXT
B5_Q4_I.SPS	Cliché d'article SPSS pour le fichier de poids bootstrap B5_Q4.TXT
B5_BC_I.SAS	Cliché d'article SAS pour le fichier de poids bootstrap B5_BC.TXT
B5_BC_I.SPS	Cliché d'article SPSS pour le fichier de poids bootstrap B5_BC.TXT
BOOTSTRP\SAS\DOC\	
SASBOOTDOC_F.PDF	Documentation en français du programme BOOTVARF_V20.SAS.
SASBOOTDOC_E.PDF	Documentation en anglais du programme BOOTVARE_V20.SAS.
BOOTSTRP\SAS\PGM\	
BOOTVARF_V20.SAS	Programme SAS de calcul de la variance utilisant le bootstrap, avec commentaires en français.
BOOTVARE_V20.SAS	Programme SAS de calcul de la variance utilisant le bootstrap, avec commentaires en anglais.
MACROF_V20.SAS	Programme de macros SAS avec commentaires en français.
MACROE_V20.SAS	Programme de macros SAS avec commentaires en anglais.
ETAPE1.SAS	Exemple, avec commentaires en français, du programme SAS créant le fichier d'analyse.
STEP1.SAS	Exemple, avec commentaires en anglais, du programme SAS créant le fichier d'analyse.
BOOTSTRP\SPSS\DOC\	
SPSSBOOTDOC_F.PDF	Documentation en français du programme BOOTVARF_V21.SPS.
SPSSBOOTDOC_E.PDF	Documentation en anglais du programme BOOTVARE_V21.SPS.
BOOTSTRP\SPSS\PGM\	
BOOTVARF_V21.SPS	Programme SPSS de calcul de la variance utilisant le bootstrap, avec commentaires en français.
BOOTVARE_V21.SPS	Programme SPSS de calcul de la variance utilisant le bootstrap, avec commentaires en anglais.
MACROF_V21.SPS	Programme de macros SPSS avec commentaires en français.
MACROE_V21.SPS	Programme de macros SPSS avec commentaires en anglais.
ETAPE1.SPS	Exemple, avec commentaires en français, du programme SPSS créant

STEP1.SPS	le fichier d'analyse. Exemple, avec commentaires en anglais, du programme SPSS créant le fichier d'analyse.
-----------	--

Fichier fictif :

DATA\ DUMY_HSI.TXT	Fichier fictif du cycle 1.1, 65 101 enregistrements
-----------------------	---

Documentation:

DOC\PDF_E\ DD_E.PDF	Dictionnaire de données en anglais.
INDEX_A_E.PDF	Index alphabétique des variables en anglais.
INDEX_T_E.PDF	Index des variables par sujet en anglais.
LAYOUT_E.PDF	Cliché d'enregistrement en anglais.
QUEST_E.PDF	Questionnaire en anglais.
DVDOC_E.PDF	Documentation sur les variables dérivées en anglais.

DOC\PDF_F\ DD_F.PDF	Dictionnaire de données en français.
INDEX_A_F.PDF	Index alphabétique des variables en français.
INDEX_T_F.PDF	Index des variables par sujet en français.
LAYOUT_F.PDF	Cliché d'enregistrement en français.
QUEST_F.PDF	Questionnaire en français.
DVDOC_F.PDF	Documentation sur les variables dérivées en français.

Clichés d'enregistrement, énoncés :

LAYOUT\ HSI_FMT.SAS	Énoncé SAS Format.
HSI_I.SAS	Énoncés SAS Infile et Input.
HSI_LBF.SAS	Énoncé SAS Label en français.
HSI_LBE.SAS	Énoncé SAS Label en anglais.
HSI_PFF.SAS	Énoncé SAS Proc Format en français.
HSI_PFE.SAS	Énoncé SAS Proc Format en anglais.
 HSI_I.SPS	 Énoncé SPSS Infile.
HSIMISS.SPS	Énoncé SPSS Missing Values.
HSIVALF.SPS	Énoncé SPSS Value Labels en français.
HSIVALE.SPS	Énoncé SPSS Value Labels en anglais.
HSIVARF.SPS	Énoncé SPSS Variable Labels en français.
HSIVARE.SPS	Énoncé SPSS Variable Labels en anglais.

