

Enquête nationale sur la santé de la population
Composante des ménages, 1998-1999
Fichiers fictifs pour le télé-accès aux fichiers maîtres

Février 2001

Avertissement

Les fichiers fictifs ne devraient pas servir à d'autres buts que le développement et la vérification des programmes informatiques devant être soumis par télé-accès. Les fichiers fictifs comprennent des données modifiées et ne doivent en aucun cas être utilisés 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, trois cycles de collecte ont été complétés : l'ENSP Cycle 1 (1994-1995), l'ENSP Cycle 2 (1996-1997) et l'ENSP Cycle 3 (1998-1999) et ce, pour chacune des trois composantes : la composante des ménages, la composante des établissements de santé et la composante du Nord.

Afin de permettre un accès élargi aux données, des fichiers de microdonnées à grande diffusion (FMGD) ont été produits pour la composante des ménages (cycles 1, 2 et 3), ainsi que pour la composante des établissements de santé (cycles 1, 2). Toutefois, 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. Les règles de confidentialité limitent également la création de FMGD de nature longitudinale.

En règle générale, les FMGD suffisent aux besoins de la plupart des initiatives de recherche. Cependant, certaines analyses nécessitent le recours aux fichiers maîtres, notamment les fichiers longitudinaux. Afin d'offrir au chercheur la possibilité d'avoir accès à ces fichiers, 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 (nphs@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,
- le titre du projet de recherche,
- un résumé du projet de recherche,
- les objectifs de la recherche,
- les données requises, incluant pourquoi la recherche nécessite le recours au(x) fichier(s) maître(s) plutôt qu'au(x) FMGD,
- les résultats projetés, et
- la date d'achèvement 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 contacter Colette Koeune par téléphone au (613) 951-1653 ou télécopieur au (613) 951-4198.

Lorsque l'approbation pour le télé-accès a été obtenue, le chercheur peut soumettre ses programmes informatiques à l'équipe de l'ENSP pour le traitement des données des fichiers maîtres. 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.

Les fichiers fictifs qui sont fournis sur ce CD-ROM permettent au chercheur de développer et de vérifier ses programmes informatiques avant de les soumettre. Bien que certaines variables à caractère administratif, qui ne représentent aucun intérêt analytique, ont été recodées à « blanc » ou « 9 », les fichiers fictifs simulent parfaitement les fichiers maîtres. Ils renferment les mêmes variables et ont le même cliché d'article. Cependant, les valeurs des variables des fichiers fictifs ont été modifiées afin de préserver la confidentialité des répondants.

Les fichiers fictifs sur ce CD-ROM portent sur les données recueillies auprès des ménages au cours du troisième cycle de l'ENSP (1998-1999). Au troisième cycle, environ 49 000 personnes ont répondu au questionnaire général. Ces répondants sont simulés sur le fichier fictif transversal général (« h35 »). Environ 17 000 personnes ont répondu au questionnaire détaillé sur la santé, et sont simulés sur le fichier fictif transversal santé (« h356 »). Les répondants sélectionnés pour la composante longitudinale de l'enquête sont simulés indépendamment sur le fichier fictif longitudinal, réponses complètes (« lngf »). Puisque la permutation des blocs est faite aléatoirement, les enregistrements fictifs pour les répondants sélectionnés n'auront pas les mêmes données d'un fichier à l'autre.

Le CD-ROM contient également un programme (SAS) pour le calcul de la variance des estimations. Des fichiers fictifs de poids bootstrap, associés à chacun des fichiers fictifs, 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 eux aussi les fichiers originaux, possèdent la même disposition et contiennent les mêmes variables que ceux-ci, mais les valeurs des poids sont modifiées par rapport aux fichiers maîtres. On retrouve également sur le CD-ROM des exemples de programmes (SAS) de calcul de variance, 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 des fichiers fictifs.

2. Création des fichiers fictifs

2.1 Catégories d'enregistrements

Les enregistrements de chaque fichier principal sont d'abord classés selon l'âge, le sexe et les caractéristiques géographiques. Dans le cas du fichier longitudinal, la classification est en outre fonction du 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 éléments de données cohérents.

Dans le cas du fichier longitudinal fictif, les catégories sont légèrement différentes. Puisque, pour ce fichier, les catégories se fondent sur l'âge au moment de l'interview en 1994-1995, en 1996-1997 et en 1998-1999, le nombre de catégories est très élevé. Afin d'assurer un nombre minimum 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 qu'à une étape ultérieure. À ce stade-là, 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 les données par la suite lorsqu'il y a incohérence entre certains éléments de données.

2.2 Création de 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 du questionnaire, puis on regroupe ces blocs élémentaires pour former les blocs de variables. Dans le cas des sections dont les questions ont été posées ou non selon les réponses aux questions d'une section antérieure, on intègre dans un même bloc les blocs élémentaires de variables correspondant à cette section, afin d'assurer la cohérence interne des enregistrements. Seules font exception à cette règle les variables utilisées pour former les catégories, puisque leurs effets sur le cheminement au long du questionnaire ont déjà été pris en compte.

Le principe directeur 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.

Pour créer les blocs, on distribue les variables entre les blocs élémentaires suivants (sauf un certain nombre de variables géographiques et d'établissement du plan de sondage, de caractéristiques des personnes et des ménages, ainsi que de variables relatives à l'administration de l'enquête) :

Blocs élémentaires

- | | |
|---|--|
| 1. Revenu du ménage | 22. VIH (1996 uniquement) |
| 2. Incapacité au cours des deux dernières semaines | 23. Renseignements sur la santé (Alberta, 1996 uniquement) |
| 3. Utilisation des soins de santé | 24. État de santé |
| 4. Limitation des activités | 25. Activités physiques |
| 5. Problèmes de santé chroniques | 26. Bronzage (Alberta, 1996 uniquement) |
| 6. Renseignements socio-démographiques | 27. Mouvements répétitifs (1996 et 1998 uniquement) |
| 7. Niveau de scolarité | 28. Blessures |
| 8. Population active | 29. Consommation de médicaments |
| 9. Insécurité alimentaire (1998 uniquement) | 30. Usage du tabac |
| 10. État de santé général | 30a. Usage du tabac (1996 uniquement) |
| 11. Taille/poids | 31. Variantes du tabagisme (1998 uniquement) |
| 11a. Taille/poids (1996 uniquement) | 32. Consommation d'alcool |
| 12. Mesures préventives de santé (1994 et 1998 uniquement) | 32a. Consommation d'alcool (Alberta, 1996 uniquement) |
| 13. Accès aux services C Tension artérielle (1996 uniquement) | 33. Dépendance à l'égard de l'alcool (1996 uniquement) |
| 14. Accès à - Test Pap (1996 uniquement) | 34. Stress actuel (1994 uniquement) |
| 15. Accès à - Mammographie (1996 uniquement) | 35. Stress récent (1994 uniquement) |
| 16. Accès à - Examen des seins (1996 uniquement) | 36. Traumatismes (1994 uniquement) |
| 17. Accès à - Examen général (1996 uniquement) | 37. Stress au travail (1994 uniquement) |
| 18. Accès à - Vaccination antigrippale (1996 uniquement) | 38. Estime de soi et contrôle (1994 uniquement) |
| 19. Accès à - Visites chez le dentiste (1996 uniquement) | 39. Sentiment de cohésion (1994 et 1998 uniquement) |
| 20. Accès à - Examen de la vue (1996 uniquement) | 40. Santé mentale |
| 21. Accès à - Services d'urgence (1996 uniquement) | 41. Soutien social |
| | 41a. Soutien social (Alberta, 1996 uniquement) |

- | | |
|--|--|
| 42. Opinions au sujet des parents (Alberta, 1996 uniquement) | 47a. Sens de la débrouillardise (Alberta, 1994) |
| 43. Services de santé (Alberta, 1996 uniquement) | 47b. Sens de la débrouillardise (Alberta, 1996) |
| 44. Santé sexuelle (1996 uniquement) | 48. Rapport avec les gens (Manitoba, 1994 uniquement) |
| 44a. Santé sexuelle (Alberta, 1996 uniquement) | 49. Soins personnels (1998 uniquement) |
| 45. Sécurité routière (1996 uniquement) | 50. Assurance (1998 uniquement) |
| 46. Violence/sécurité personnelle (Alberta, 1996 uniquement) | 51. Antécédents médicaux de la famille (1998 uniquement) |
| | 52. Nutrition (1998 uniquement) |

On notera que les blocs élémentaires n° 23, 26, 32a, 41a, 42, 43, 44a, 46, 47a, 47b et 48 renferment le contenu spécial pour le Manitoba et l'Alberta. Ce contenu a été rajouté à une étape subséquente et, en règle générale, les catégories sont fondées sur la province (Manitoba, Alberta et le reste du Canada), l'âge, le sexe et le profil de réponse longitudinale.

2.3 Variables recodées

Pour fins de confidentialité, les variables énumérées ci-après ont été recodées à « blanc » ou « 9 » lors de la création des fichiers fictifs.

Fichier fictif général (h35) :

AM58_RNO, AM38_LP, AM38_LNG, AM58_LNG, AM58_SHA, AM38_SRC, AM38_CAS, AM58_CAS, AM58_BMM, AM58_BDD, AM58_BYYY, AM58_EMM, AM58_EDD, AM58_EYY, SP38_CPS, SP38_CPA, SP38_STA, AM68_STA, SP38_PAR, SP38DPC, CYCLE, SUBCYCLE, MEMCYCLE, STRATUM, REPLICAT.

Fichier fictif santé (h356) :

Toutes les variables énumérées sous « Fichier général », sauf AM58_RNO, auxquelles on a ajouté les variables suivantes :

AM68_RNO, AM68_LNG, AM68_SHA, AM68_SRC, AM68_CAS, AM68_BMM, AM68_BDD, AM68_BYYY, AM68_EMM, AM68_EDD, AM68_EYY, AM68_LNK.

Fichier fictif longitudinal, réponses complètes (lngf) :

DOD, SP34_CPS, SP34_MET, CYCLE, SUBCYCLE, SP34DPC, AM34_SRC, AM34_LNG, SP34_CPA, AM54_BMM, AM54_BDD, AM54_BYYY, AM54_SRC, AM54_LNG, AM54_EMM, AM54_EDD, AM54_EYY, AM64_STA, AM64_BMM, AM64_BDD, AM64_BYYY, AM64_SRC, AM64FR, AM64_LNG, AM64_EMM, AM64_EDD, AM64_EYY, SP36FOUT, SP36_CPA, SP36DPC, AM36_SRC, AM36_LNG, SP36_STA, AM56_STA, AM56_BMM, AM56_BDD, AM56_BYYY, AM56_SRC, AM56_LNG, AM56_EMM, AM56_EDD, AM56_EYY, AM66LDUR, AM66_STA, AM66_BMM, AM66_BDD, AM66_BYYY, AM66_SRC, AM66FR, AM66_5, AM66_LNG, AM66_EMM, AM66_EDD, AM66_EYY, AM68LDUR, SP38FOUT, SP38DPC, SP38_CPA, AM38_LP, AM38_SRC,

AM38_LNG, SP38_STA, AM58_STA, AM68_STA, AM58_BDD, AM58_BMM, AM58_BYY, AM58_LNG, AM58_EMM, AM58_EDD, AM58_EYY, AM68_BMM, AM68_BDD, AM68_BYY, AM68_SRC, AM68FR, AM68_LNK, AM68_LNG, AM68_EMM, AM68_EDD, AM68_EYY.

3. Contenu du CD_ROM

Lisez_moi.pdf	Documentation sur le télé-accès, la création des fichiers fictifs, le contenu du CD-ROM (en français)
Read_me.pdf	Documentation sur le télé-accès, la création des fichiers fictifs, le contenu du CD-ROM (en anglais)

Fichiers fictifs :

DATA\ dumyh35.txt	Fichier fictif général (h35), 49 046 enregistrements
dumyh356.txt	Fichier fictif santé (h356), 17 244 enregistrements
dumylngf.txt	Fichier fictif longitudinal, réponses complètes (lngf), 14 619 enregistrements

Clichés d'enregistrement, énoncés :

LAYOUT\ h35_fmt.sas	h35 SAS FORMAT
h35_i.sas	h35 SAS INFILE et INPUT
h35_lbf.sas	h35 SAS LABEL en français
h35_lbe.sas	h35 SAS LABEL en anglais
h35_pff.sas	h35 SAS PROC FORMAT en français
h35_pfe.sas	h35 SAS PROC FORMAT en anglais
h35_i.sps	h35 SPSS FILE HANDLE et DATA LIST
h35miss.sps	h35 SPSS MISSING VALUES
h35valf.sps	h35 SPSS VALUE LABELS en français
h35vale.sps	h35 SPSS VALUE LABELS en anglais
h35varf.sps	h35 SPSS VARIABLE LABELS en français
h35vare.sps	h35 SPSS VARIABLE LABELS en anglais
 h356_fmt.sas	 h356 SAS FORMAT
h356_i.sas	h356 SAS INFILE et INPUT
h356_lbf.sas	h356 SAS LABEL en français
h356_lbe.sas	h356 SAS LABEL en anglais
h356_pff.sas	h356 SAS PROC FORMAT en français
h356_pfe.sas	h356 SAS PROC FORMAT en anglais
h356_i.sps	h356 SPSS FILE HANDLE et DATA LIST

h356miss.sps
h356valf.sps
h356vale.sps
h356varf.sps
h356vare.sps

h356 SPSS MISSING VALUES
h356 SPSS VALUE LABELS en français
h356 SPSS VALUE LABELS en anglais
h356 SPSS VARIABLE LABELS en français
h356 SPSS VARIABLE LABELS en anglais

lngf_fmt.sas
lngf_i.sas
lngf_lbf.sas
lngf_lbe.sas
lngf_pff.sas
lngf_pfe.sas
lngf_i.sps
lngfmiss.sps
lngfvalf.sps
lngfvale.sps
lngfvarf.sps
lngfvare.sps

lngf SAS FORMAT
lngf SAS INFILE et INPUT
lngf SAS LABEL en français
lngf SAS LABEL en anglais
lngf SAS PROC FORMAT en français
lngf SAS PROC FORMAT en anglais
lngf SPSS FILE HANDLE et DATA LIST
lngf SPSS MISSING VALUES
lngf SPSS VALUE LABELS en français
lngf SPSS VALUE LABELS en anglais
lngf SPSS VARIABLE LABELS en français
lngf SPSS VARIABLE LABELS en anglais

Documentation :

DOC\

PDF_E\

h35_dd.pdf

Dictionnaire des données, fichier maître général (h35) en anglais

h35_ind.pdf

Index alphabétique, fichier maître général (h35) en anglais

h35_lay.pdf

Cliché d'enregistrement, fichier maître général (h35) en anglais

h35_top.pdf

Index par sujet, fichier maître général (h35) en anglais

h356_dd.pdf

Dictionnaire des données, fichier maître santé (h356) en anglais

h356_ind.pdf

Index alphabétique, fichier maître santé (h356) en anglais

h356_lay.pdf

Cliché d'enregistrement, fichier maître santé (h356) en anglais

h356_top.pdf

Index par sujet, fichier maître santé (h356) en anglais

lngf_dd.pdf

Dictionnaire des données, fichier maître longitudinal, réponses complètes (lngf) en anglais

lngf_ind.pdf

Index alphabétique, fichier maître longitudinal, réponses complètes (lngf) en anglais

lngf_lay.pdf

Cliché d'enregistrement, fichier maître longitudinal, réponses complètes (lngf) en anglais

lngf_top.pdf

Index par sujet, fichier maître longitudinal, réponses complètes (lngf) en anglais

DOC\ PDF_F\ h35_dd.pdf	Dictionnaire des données, fichier maître général (h35) en français
h35_ind.pdf	Index alphabétique, fichier maître général (h35) en français
h35_lay.pdf	Cliché d'enregistrement, fichier maître général (h35) en français
h35_top.pdf	Index par sujet, fichier maître général (h35) en français
h356_dd.pdf	Dictionnaire des données, fichier maître santé (h356) en français
h356_ind.pdf	Index alphabétique, fichier maître santé (h356) en français
h356_lay.pdf	Cliché d'enregistrement, fichier maître santé (h356) en français
h356_top.pdf	Index par sujet, fichier maître santé (h356) en français
lngf_dd.pdf	Dictionnaire des données, fichier maître longitudinal, réponses complètes (lngf) en français
lngf_ind.pdf	Index alphabétique, fichier maître longitudinal, réponses complètes (lngf) en français
lngf_lay.pdf	Cliché d'enregistrement, fichier maître longitudinal, réponses complètes (lngf) en français
lngf_top.pdf	Index par sujet, fichier maître longitudinal, réponses complètes (lngf) en français

Fichiers pour l'estimation de la variance par BOOTSTRAP :

Poids Bootstrap :

BOOTSTRP\ bd5h35.txt	Poids bootstrap fictifs pour le fichier fictif général (h35)
bd5h356.txt	Poids bootstrap fictifs pour le fichier fictif santé (h356)
bd5lngf.txt	Poids bootstrap fictifs pour le fichier fictif longitudinal, réponses complètes (lngf)

Documentation :

BOOTSTRP\ DOC\ btdoceng.pdf	Documentation du programme bootvar.sas en anglais
btdocfr.pdf	Documentation du programme bootvarf.sas en français
lisezmoi.txt	Description des fichiers pour l'estimation de la variance par bootstrap (en français)
readme.txt	Description des fichiers pour l'estimation de la variance par bootstrap (en anglais)

Programme de calcul de variance et exemples :

BOOTSTRP\
PGM_E\
bootvar.sas

Programme de calcul de variance par bootstrap avec commentaires en anglais

ex1a_eng.sas

Exemple 1a (Commentaires en anglais)

ex1b_eng.sas

Exemple 1b (Commentaires en anglais)

ex2a_eng.sas

Exemple 2a (Commentaires en anglais)

ex2b_eng.sas

Exemple 2b (Commentaires en anglais)

BOOTSTRP\
PGM_F\
bootvarf.sas

Programme de calcul de variance par bootstrap avec commentaires en français

ex1a_fr.sas

Exemple 1a (Commentaires en français)

ex1b_fr.sas

Exemple 1b (Commentaires en français)

ex2a_fr.sas

Exemple 2a (Commentaires en français)

ex2b_fr.sas

Exemple 2b (Commentaires en français)

Clichés d'enregistrement, :

BOOTSTRP\
LAYOUT\
b35_i.sas

bd5h35 SAS INFILE et INPUT

b356_i.sas

bd5h356 SAS INFILE et INPUT

blngf_i.sas

bd5lngf SAS INFILE et INPUT