



N° 12-002-XIF au catalogue

Le Bulletin technique et d'information des Centres de données de recherche

Automne 2004, vol.1 n° 2



Comment obtenir d'autres renseignements

Toute demande de renseignements au sujet du présent produit ou au sujet de statistiques ou de services connexes doit être adressée à : Le programme des centres de données de recherche, Statistique Canada, Ottawa, Ontario, K1A 0T6 (téléphone : 1 800 263-1136).

Pour obtenir des renseignements sur l'ensemble des données de Statistique Canada qui sont disponibles, veuillez composer l'un des numéros sans frais suivants. Vous pouvez également communiquer avec nous par courriel ou visiter notre site Web.

Service national de renseignements	1 800 263-1136
Service national d'appareils de télécommunications pour les malentendants	1 800 363-7629
Renseignements concernant le Programme des bibliothèques de dépôt	1 800 700-1033
Télécopieur pour le Programme des bibliothèques de dépôt	1 800 889-9734
Renseignements par courriel	infostats@statcan.ca
Site Web	www.statcan.ca

Renseignements sur les commandes et les abonnements

Le produit n° 12-002-XIF au catalogue est gratuit sur Internet. Les utilisateurs sont priés de se rendre à http://www.statcan.ca/cgi-bin/downpub/freepub_f.cgi

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois, et ce, dans la langue officielle de leur choix. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1 800 263-1136.



Statistique Canada
Le programme des centres de données de recherche

Le Bulletin technique et d'information des Centres de données de recherche

Automne 2004, vol.1 n° 2

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2004

Tous droits réservés. Il est interdit de reproduire ou de transmettre le contenu de la présente publication, sous quelque forme ou par quelque moyen que ce soit, enregistrement sur support magnétique, reproduction électronique, mécanique, photographique, ou autre, ou de l'emmagasiner dans un système de recouvrement, sans l'autorisation écrite préalable des Services de concession des droits de licence, Division du marketing, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Octobre 2004

N° 12-002-XIF au catalogue

Périodicité: semi-annuelle

ISSN : 1710-2200

Ottawa

This publication is available in English (Catalogue no. 12-002-XIE)

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

À propos du Bulletin technique et d'information

Le Bulletin technique et d'information des Centres de données de recherche est un forum permettant aux utilisateurs actuels et prospectifs des centres de partager de l'information et les techniques d'analyse des données disponibles dans les centres. Le bulletin paraît au printemps et à l'automne, et l'on publiera à l'occasion des numéros spéciaux sur des questions d'actualité.

Objectifs

Les objectifs principaux de ce bulletin sont les suivants :

- l'accroissement et la diffusion de la connaissance concernant les données de Statistique Canada;
- les échanges d'idées parmi les utilisateurs membres des Centres de données de recherche (CDR);
- l'aide aux nouveaux utilisateurs du programme CDR; et
- offrir des occasions supplémentaires permettant aux chercheurs dans les centres de communiquer avec les spécialistes et divisions spécialisées au sein de Statistique Canada.

Le contenu

Nous souhaitons publier des articles qui contribueront à accroître la qualité des travaux de recherche menés dans les Centres de données de recherche de Statistique Canada et qui fourniront des conseils méthodologiques aux chercheurs travaillant dans les CDR.

Les articles figurant dans le Bulletin technique et d'information portent principalement sur :

- l'analyse et la modélisation des données;
- la gestion des données;
- les pratiques statistiques, informatiques ou scientifiques éprouvées ou au contraire inefficaces;
- le contenu en données;
- les effets associés au libellé des questionnaires;
- la comparaison d'ensembles de données;
- l'examen des méthodes et de leur application;
- les particularités que présentent les données;
- les problèmes associés aux données et leurs solutions; et
- les outils innovateurs faisant appel aux enquêtes des CDR et aux logiciels pertinents.

Ceux et celles qui s'intéressent à soumettre un article au Bulletin technique et d'information sont priés de suivre les directives pour les auteurs.

Les rédacteurs et les auteurs tiennent à remercier les réviseurs de leurs commentaires précieux.

Rédacteur: James Chowhan

Rédacteurs adjoints: Heather Hobson, Leslie-Anne Keown, Darren Lauzon

Table des matières**Les articles**

Owen Phillips,	Comment utiliser les poids bootstrap avec Wes Var et SUDAAN	6
James Chowhan,	Fichiers de commandes Stat/Transfer et transfert efficace des fichiers de données	16

Note technique

Franck Larouche,	Note sur les identificateurs dans l'Enquête longitudinale nationale sur les enfants et les jeunes	23
Comité de révision,	Directives pour les auteurs	24

Comment utiliser les poids *bootstrap* avec Wes Var et SUDAAN

Par Owen Phillips

Résumé

Pour faciliter l'estimation de la variance fondée sur le plan de sondage, de plus en plus d'enquêtes de Statistique Canada fournissent des poids *bootstrap* avec leurs fichiers de microdonnées. Bien que les logiciels commerciaux comme SUDAAN et WesVar ne prennent pas explicitement en charge les poids *bootstrap*, on peut, en misant sur la similarité des techniques *bootstrap* et BRR (répliques répétées équilibrées), utiliser ces logiciels pour estimer les variances par *bootstrap*. Cet article explore les fondements théoriques qui justifient cette façon de procéder et montre, par un exemple, comment y arriver. La dernière section propose un bref aperçu des autres méthodes d'estimation de la variance fondée sur le plan de sondage, ainsi que les logiciels, les programmes et les procédures dans lesquels ces méthodes sont employées.

Introduction

On emploie de plus en plus la méthode du *bootstrap* pour l'estimation de la variance fondée sur le plan de sondage. Plusieurs enquêtes de Statistique Canada – l'Enquête sur la dynamique du travail et du revenu (EDTR – SLID), l'Enquête nationale sur la santé de la population (ENSP – NPHS), l'Enquête sociale générale (ESG – GSS), pour n'en nommer que quelques-unes – fournissent toutes avec leurs microdonnées des poids *bootstrap*, ou des variantes de ces poids, pour l'estimation de la variance fondée sur le plan de sondage.

Le *bootstrap* (enquête) appartient à une famille de méthodes d'estimation de la variance dite par « réplification ». On trouvera un traitement détaillé des méthodes de réplification dans les ouvrages de Lohr (1999), Rust et Rao (1996) ou Wolter (1985). Ces méthodes produisent des échantillons synthétiques appelés « répliques », à partir de l'échantillon original. Une autre méthode de réplification, la méthode des « répliques répétées équilibrées » (*Balanced Repeated Replication* ou BRR), a été implémentée dans certains logiciels commerciaux, comme SUDAAN et WesVar. Même si le *bootstrap* et le BRR ne construisent pas les répliques de la même façon, on peut utiliser les poids *bootstrap* pour calculer les variances avec les logiciels qui prennent en charge les poids BRR, particularité qui est rarement soulignée dans la documentation des logiciels.

Dans les sections suivantes, nous approfondissons les différences et les similitudes entre les méthodes *bootstrap* et BRR, et nous illustrerons par un exemple comment utiliser les poids *bootstrap* dans les logiciels SUDAAN et WesVar. L'ESG et l'Enquête sur le milieu de travail et les employés (EMTE) utilisent une variante du *bootstrap*, appelé *bootstrap* moyen, que l'on peut rapprocher de la variante du BRR due à Fay. Nous terminons cette note en présentant un bref aperçu des autres techniques d'estimation de la variance fondée sur le plan de sondage, et de quelques logiciels et programmes qui incorporent bon nombre des techniques décrites ci-dessous.

Afin d'assurer une plus grande lisibilité, nous limiterons la discussion au problème général de la pondération et nous laisserons de côté quelques-unes des complications connexes, comme les ajustements pour non-réponse ou la stratification *à posteriori*. Toutefois, nous

supposons que le lecteur est familier avec les concepts de la théorie de l'échantillonnage. Les lecteurs qui désirent en savoir plus à ce sujet pourront consulter avec profit les *Méthodes et pratiques d'enquête* (Statistique Canada 2003). Nous supposons également que le lecteur a une certaine habitude des logiciels SAS, SUDAAN ou WesVar.

II. Méthodes *bootstrap*

Les méthodes de rééchantillonnage par *bootstrap* sont utilisées dans un nombre grandissant d'enquêtes à Statistique Canada; mentionnons l'EDTR, l'ENSP, l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ), l'Enquête sur la santé dans les collectivités canadiennes (ESCC), l'Enquête sur la diversité ethnique (EDE), ainsi que l'Enquête auprès des jeunes en transition (EJT). Sans entrer dans les détails, on construit les répliques en choisissant avec remise un échantillon d'unités primaires d'échantillonnage (UPE) dans chaque strate, et en ajustant les poids d'échantillonnage originaux des unités dans les UPE sélectionnées pour refléter les probabilités du tirage dans le sous-échantillon. Si une unité ne figure pas dans la réplique de *bootstrap*, son poids *bootstrap* a la valeur zéro. Ce processus de tirage et de pondération est répété B fois afin d'obtenir B échantillons *bootstrap*, B poids *bootstrap*, et conséquemment B estimations par *bootstrap*.

La variance de l'estimateur $\hat{\theta}$ du paramètre de population finie θ – par exemple, un coefficient de régression, une moyenne, le quotient de deux totaux – est estimée par

$$\hat{V}_{BOOT}(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^B (\hat{\theta}_b - \hat{\theta})^2 \quad (1)$$

où $\hat{\theta}$ est l'estimateur utilisant les poids de sondage originaux et les estimateurs $\hat{\theta}_b$, $b = 1, \dots, B$ sont ceux obtenus exactement de la même manière en utilisant les variables de poids *bootstrap*^{1,2}.

Les enquêtes ESG et EMTE ont recours à une variante de *bootstrap* appelée ***bootstrap moyen***. Cette méthode a été mise au point à l'origine pour résoudre des problèmes de confidentialité découlant de la publication de poids *bootstrap* avec des fichiers publics de microdonnées (voir Yung 1997). Il s'agit en fait de calculer des poids *bootstrap* tel qu'indiqué ci-dessus et de prendre la moyenne de C échantillons *bootstrap*. Par exemple, pour certains cycles de l'ESG, on a produit 5 000 séries de poids *bootstrap*. On a ensuite fait la moyenne de groupes de taille $C = 25$ pour conserver $B = 200$ poids *bootstrap* moyens, ces derniers étant

1. Le poids de l'échantillonnage reflète la probabilité des tirages d'une unité dans l'échantillon complet : on peut le voir comme étant le nombre d'unités dans la population observée représentées par l'unité échantillonnée. Le poids d'échantillonnage est utilisé pour estimer le paramètre qui nous intéresse. Le poids *bootstrap* est utilisé afin de calculer l'erreur d'échantillonnage associée à ce paramètre. Tout comme le poids d'échantillonnage, un poids *bootstrap* peut être considéré comme le nombre d'individus dans la population observée représentés par une unité dans l'échantillon réduit (*bootstrap*).

2. Les recherches se poursuivent au sujet des méthodes visant à obtenir les estimateurs $\hat{\theta}_b$ à partir des échantillons *bootstrap*, en sus de l'approche utilisée pour estimer $\hat{\theta}$ à partir de l'échantillon complet. Ces approches pourraient permettre des calculs stables de la variance dans certaines situations (voir Roberts *et coll.*, 2003).

fournis avec les microdonnées. De la même façon, l'EMTE fournit 100 poids *bootstrap* moyens, chacun venant de groupes de taille $C = 50$.

L'estimateur de la variance par *bootstrap* moyen est donnée par :

$$\hat{V}_{MBOOT}(\hat{\theta}) = \frac{C}{B} \sum_{b=1}^B (\hat{\theta}_b - \hat{\theta})^2 \quad (2)$$

où $\hat{\theta}_b$ est obtenu au moyen du b^e poids *bootstrap* moyen.

III. Répliques répétées équilibrée (BRR)

Les répliques répétées équilibrées (BRR) sont utilisables quand deux et seulement deux UPE ont été choisie dans chaque strate. La plupart des enquêtes élaborées à Statistique Canada comptent plus de deux UPE par strate, et on ne peut donc pas utiliser la méthode BRR pour calculer la variance pour ces sondages. Toutefois, certaines caractéristiques de la méthode BRR nous permettent d'utiliser, pour estimer la variance par *bootstrap*, les logiciels conçus pour calculer la variance par la méthode BRR.

La méthode BRR consiste à construire des demi-échantillons ne conservant qu'une seule des deux UPE de chaque strate. En suivant une procédure prescrite³, un ensemble « équilibré » de G demi-échantillons est sélectionné. Dans chacun d'entre eux, les unités retenues voient leur poids doublé alors que les unités rejetées voient leur poids réduit à zéro. On assemble des estimations de demi-échantillons de la manière suivante :

$$\hat{V}_{BRR}(\hat{\theta}) = \frac{1}{G} \sum_{g=1}^G (\hat{\theta}_g - \hat{\theta})^2 \quad (3)$$

Une variante de la méthode BRR est la *méthode de Fay*. La méthode utilisée pour sélectionner les demi-échantillons est la même que pour la méthode BRR; toutefois, la pondération se fait différemment : les poids des unités dans les UPE sélectionnées sont multipliés par un facteur $(2-K)$; les poids des unités rejetées sont multipliés par K , où K est une constante choisie dans l'intervalle $[0,1)$. De cette façon, toutes les unités observées contribuent à chaque estimation $\hat{\theta}_g$, ce qui s'avère particulièrement utile quand on travaille sur de petits domaines. La variance est ensuite estimée par :

$$\hat{V}_{FAY}(\hat{\theta}) = \frac{1}{G(1-K)^2} \sum_{g=1}^G (\hat{\theta}_g - \hat{\theta})^2 \quad (4)$$

Le Programme international pour le suivi des acquis des élèves (PISA) utilise la méthode de Fay avec une constante $K = 0.5$.

3. La méthode de production des demi-échantillons équilibrés n'est pas nécessaire pour cette discussion. Le lecteur intéressé peut consulter la section 9.3.1 de Lohr (1999).

IV. Ce que la documentation ne vous dit pas

Comme on l'a vu dans les sections II et III, les façons de créer les poids *bootstrap* et *BRR* sont différentes. Ceci dit, la forme de l'estimateur de variance est la même pour les deux approches, c'est-à-dire les équations (1) et (3) sont équivalentes, pourvu que $G = B$. En d'autres mots, des logiciels comme SUDAAN ou WesVar qui permettent de calculer la variance par *BRR* peuvent calculer la variance par *bootstrap*, pour peu que l'utilisateur fournisse les poids *bootstrap*. De même, en posant $K = 1 - C^{-\frac{1}{2}}$ dans (4), on obtient l'équivalence des formules pour la variante de Fay (4) et le *bootstrap* moyen (2), et les logiciels qui acceptent la variante de Fay peuvent calculer la variance par *bootstrap* moyen.

Utilisation de poids *bootstrap* avec SUDAAN

On indique la méthode d'estimation de la variance dans l'appel de la procédure d'analyse. La démarche qui suit vaut pour toutes les procédures SUDAAN qui autorisent l'estimation par *BRR*⁴ :

- On utilise l'option DESIGN=BRR pour une estimation par *bootstrap*.
- L'énoncé REPWGT est utilisé pour indiquer le nom des variables contenant les poids *bootstrap*.
- L'énoncé WEIGHT n'est pas obligatoire, mais devrait être utilisé pour repérer la variable contenant le poids d'estimation (si disponible)⁵.
- Pour les enquêtes recourant au *bootstrap* moyen, on emploie l'option ADJFAY=C (on se rappellera que pour l'ESG, $C = 25$ et que pour l'EMTE, $C = 50$).

Utilisation des poids *bootstrap* avec WesVar

Dans WesVar, l'utilisateur spécifie la méthode d'estimation de la variance lors de la création du fichier de données WesVar. Le fichier résultant sert ensuite à la création des livrets dans lesquels les demandes de tableaux croisés et de modèles de régressions sont exécutées. Pour créer un fichier de données WesVar avec *bootstrap* ou *bootstrap* moyen, il faut:

- Déplacer les variables de poids *bootstrap* dans la case *Replicates*.
- Déplacer le poids final dans la case *Full sample*.
- Pour le *bootstrap*, cocher l'option BRR de la case *Method*.
- Pour le *bootstrap* moyen, cocher l'option Fay de la case *Method* et spécifier $Fay_K = 1 - C^{-\frac{1}{2}}$
- Déplacer les variables d'analyse dans la case *Variables*, indiquer l'identificateur unique dans la case *ID* (facultatif) et enregistrer le fichier.

4. Les versions 8.02 et antérieures de SUDAAN n'acceptent pas l'option DESIGN=BRR dans la procédure PROC SURVIVAL.

5. En l'absence de l'instruction de pondération, SUDAAN utilise $\hat{\theta}_{(b)} = \frac{1}{B} \sum_b \hat{\theta}_b$ à la place de $\hat{\theta}$.

V. Exemples d'utilisation de WesVar et de SUDAAN

Les exemples suivants sont des applications de l'information présentée à la section IV. Nous omettons ici les détails de la définition des nouvelles variables et de la manipulation des données dans un format utilisable avec le logiciel. L'objectif de ces exemples est de tout simplement montrer comment on peut utiliser la méthode *bootstrap* dans ces deux logiciels, et nous omettons donc l'interprétation des résultats. Le lecteur trouvera les détails de lecture des données et d'interprétation des résultats dans SUDAAN et WesVar dans les guides d'utilisation respectifs de ces logiciels (voir RTI, 2001 et Westat, 2002).

Bootstrap avec l'EDTR

Dans cet exemple, on s'intéresse aux individus longitudinaux ayant franchi le seuil de faible revenu entre 1996 et 1998 en utilisant le panel 1 de l'EDTR. Étant donné la période d'intérêt, l'un ou l'autre ou une combinaison des panels 1 et 2 aurait pu être utilisé. La variable *low96* prend la valeur 1 si, en 1996, le revenu de la personne longitudinale tiré de son emploi principal était sous un seuil donné, et 0 autrement. La variable *low98* est définie de façon similaire. Les données sont stockées dans un fichier intitulé *mobility*. Les poids *bootstrap bs1-bs1000* associés au poids longitudinal du panel 1 *ilgwt26* ont été ajoutés au fichier d'analyse.

(a) SUDAAN

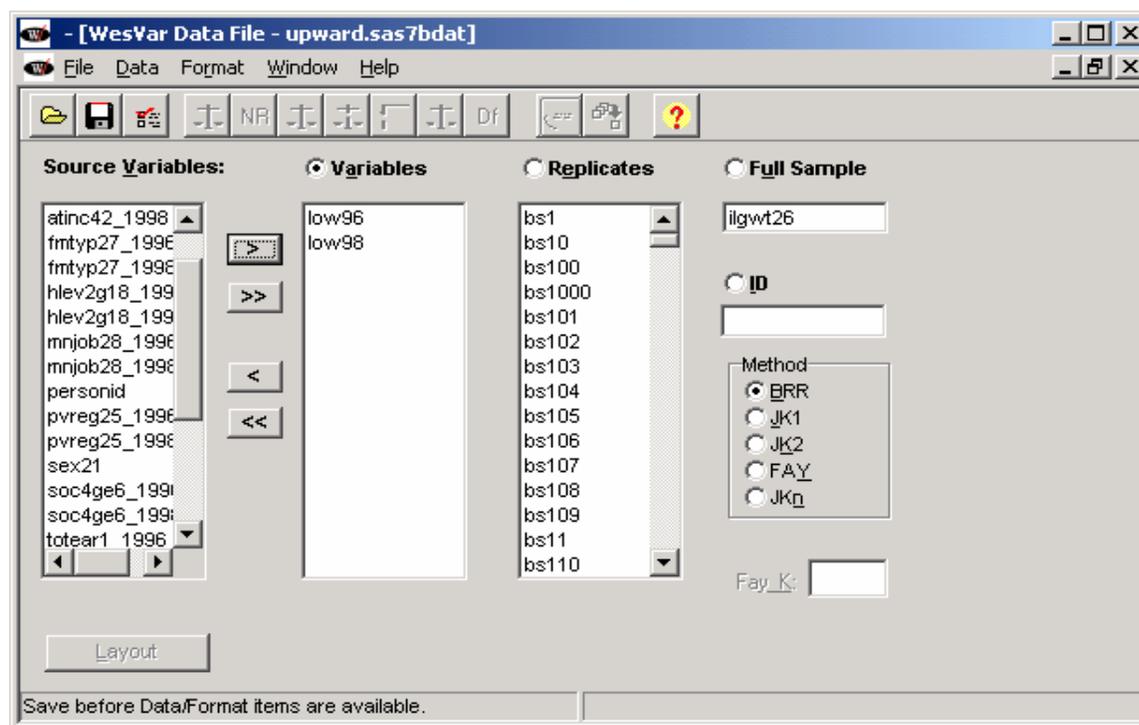
La procédure PROC CROSSTAB a été utilisée dans SUDAAN, appellable depuis SAS, pour produire les estimations du nombre de ceux qui sont passés dans la catégorie « faible revenu » entre les deux vagues d'enquête, ou qui en sont sortis. En suivant les étapes données dans la section IV, l'appel de la procédure PROC CROSSTAB serait formulé ainsi:

```
proc crosstab data = mobility design = BRR;
  weight ilgwt26;
  repwgt bs1-bs1000;
  recode low96 = (0 1) low98 = (0 1);
  subgroup low96 low98;
  levels 2 2;
  tables low96*low98;
run;
```

Pour de plus amples renseignements à propos de PROC CROSSTAB et de SUDAAN, veuillez consulter le guide de l'utilisateur (Research Triangle Institute, 2001).

(b) WesVar

Avec le même fichier SAS et en suivant les indications données à la section IV, un fichier WesVar serait construit comme dans la Figure 1, puis enregistré et rappelé lors de l'élaboration des tableaux croisés et des modèles de régression. Le guide de l'utilisateur de WesVar contient les instructions nécessaires (Westat 2002).

Figure 1: Exemple de *bootstrap* pour WesVar avec l'EDTR

Bootstrap moyen avec l'ESG

Dans cet exemple, on utilise le cycle 14 de l'ESG pour étudier la relation entre la propension à l'utilisation d'Internet et divers facteurs sociodémographiques au moyen d'une régression logistique. La variable dépendante est *netuse* (1 si le répondant utilise Internet, 0 sinon). Toutes les variables du modèle sont nominales. Le poids d'estimation sur le fichier SAS intitulé *internet* est *fwgt*, et les poids *bootstrap* moyens correspondants sont *bsw1-bsw200*. Rappelons que pour l'ESG, on utilise $C = 25$.

(a) SUDAAN

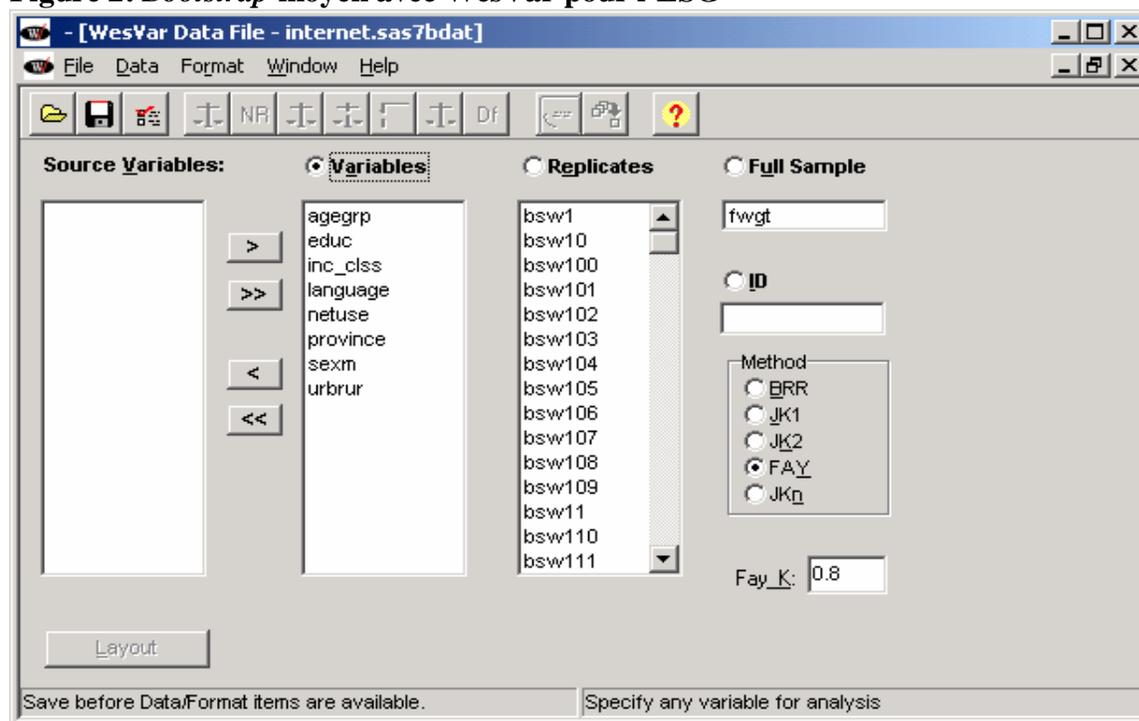
En suivant les étapes données dans la section IV, l'appel de la procédure PROC RLOGIST, dans SUDAAN appellable depuis SAS, serait formulé ainsi :

```
proc rlogist data = internet design = BRR;
  weight fwgt;
  repwgt bsw1-bsw200 / adjfay = 25;
  subpopn province = 59;
  subgroup sexm agegrp educ urbrur inc_class language;
  levels 2 3 4 5 5 3;
  model netuse = sexm agegrp educ urbrur inc_class language;
run;
```

(b) WesVar

On voit à la Figure 2 comment cette information est utilisée pour définir le fichier de données WesVar. Dans le cas de l'ESG, on pose $Fay_K = 1 - 25^{-\frac{1}{2}} = 0.8$. Le fichier ainsi obtenu serait enregistré et utilisé pour définir le modèle de régression dans un livret WesVar.

Figure 2: Bootstrap moyen avec WesVar pour l'ESG



VI. Autres méthodes d'estimation de la variance fondée sur le plan de sondage

Les méthodes *bootstrap* et *BRR* sont deux des méthodes possibles pour estimer la variance fondée sur le plan de sondage. Le *jackknife delete-1*, une autre méthode de rééchantillonnage, crée des répliques en retirant à tour de rôle chaque UPE de l'échantillon, et en repondérant les $n-1$ restantes. Comme les UPE sont sous-échantillonnées $n-1$ à la fois, il y a autant de répliques et autant de séries de poids *jackknife* qu'il y a d'UPE dans l'échantillon. Il y a d'autres variantes de la méthode *jackknife* qui ne sont pas décrites ici.

Les méthodes de linéarisation, ou approximations par séries de Taylor, forment une autre classe de méthodes d'estimation de la variance. C'est la méthode implémentée par SPSS et SAS dans leurs procédures spécialisées pour enquêtes à plan complexes. Les paramètres non linéaires (par exemple quotients ou coefficients de régression) sont alors exprimés comme une fonction lisse de statistiques simples (totaux ou moyennes) pour lesquelles on connaît les estimateurs de variance. Le paramètre d'intérêt est approché par un développement de Taylor de cette fonction au premier ordre autour de la valeur « vraie »; on approxime ensuite l'erreur d'échantillonnage.

Cette méthode ne convient pas aux statistiques que l'on ne peut pas bien approximer à l'aide d'une fonction linéaire, comme les statistiques chi carré, par exemple.

Pour utiliser la méthode de linéarisation ou le *jackknife* quand les poids *jackknife* ne sont pas fournis, il faut fournir les indicateurs de strate et d'UPE avec les microdonnées. De plus, les logiciels comme SUDAAN, WesVar et Stata ne savent pas prendre en compte la non-réponse, la stratification *a posteriori* et les autres ajustements quand on utilise la linéarisation de Taylor ou quand on crée des poids *jackknife*. Ceci dit, dans le cas des enquêtes qui ne fournissent pas les poids *bootstrap* ou si l'analyse souhaitée n'autorise pas l'utilisation de poids *bootstrap* (p ex., PROC SURVIVAL dans SUDAAN 8.02), il est recommandé d'utiliser l'une de ces méthodes.

Le chapitre 9 de Lohr (1999) contient un aperçu détaillé des méthodes de réplication et de linéarisation présentées dans la présente note.

Les logiciels commerciaux comme SUDAAN, WesVar, Stata et SAS offrent des commandes et des procédures analytiques pouvant effectuer des analyses sélectionnées fondées sur le plan de sondage. En outre, pour calculer la variance par la méthode *bootstrap*, on peut utiliser de manière récursive tout logiciel qui offre une commande ou une procédure analytique pouvant produire des estimations pondérées des paramètres d'intérêt et qui offre également la souplesse d'un langage de programmation. En se basant sur ce principe, les méthodologistes de Statistique Canada ont construit des macros SAS et SPSS, qui sont regroupées ensemble et fournies avec les microdonnées des enquêtes (*Bootvar*⁶ pour l'ENSP et l'ESCC, et le *Variance Estimation System (VES)* pour l'ELNEJ). Un programme similaire, pour le logiciel SAS et appelé *Bootmac* (écrit par un chercheur indépendant), est disponible des Centres de données de recherche (CRD). On peut également utiliser la commande *Bswreg*, définie par l'utilisateur dans Stata, pour calculer la variance par la méthode *bootstrap*, pour bon nombre des commandes de régression existantes dans Stata⁷. Les avantages de ce programme sont expliqués et exploités dans la dernière édition du présent bulletin (voir Piérard et coll. 2004).

Le tableau figurant en annexe compare les capacités de plusieurs logiciels et programmes mentionnés ci-dessus, pour l'estimation de la variance fondée sur le plan de sondage. Ce tableau indique les méthodes de calcul de la variance prises en charge par les logiciels, ainsi que les procédures analytiques offertes à l'utilisateur.

6. Une version générique du programme *Bootvar* est en train d'être produite comme outil requis d'estimation de la variance pour un certain nombre d'enquêtes qui fournissent des poids *bootstrap* avec leurs microdonnées. Cet outil devrait être largement disponible à l'automne 2004. Les fonctionnalités de ce programme seront probablement élargies pour permettre notamment l'estimation des erreurs d'échantillonnage pour les quantiles et des tests d'indépendance et d'homogénéité fondés sur le plan de sondage.

7. Le programme *Bswreg* ne prend pas en charge les commandes de régression de STATA qui comportent plus d'une ligne de codes pour mettre en œuvre la procédure de régression. Par exemple, ce programme ne prend pas en charge le modèle des risques proportionnels de Cox.

Référence

- Lohr, S. 1999. *Sampling: Design and Analysis*. Duxbury Press, USA.
- Piérard, E., Buckley, N., Chowhan, J. Pour une utilisation plus conviviale de la méthode *bootstrap* : fichier ADO dans Stata. *Le Bulletin technique et d'information des Centres de données de recherche* 1(1): 20-36.
- Research Triangle Institute. 2001. *SUDAAN User's Manual, Release 8.0*. Research Triangle Institute, Research Triangle Park, NC.
- Roberts, G., Binder, D., Kovacevic, M., Pantel, M., Phillips, O. 2003. Using an estimating function *bootstrap* approach for obtaining variance estimates when modelling complex health survey data. *Actes du Groupe des méthodes d'enquête*, Société statistique du Canada.
- Rust, K.F., Rao, J.N.K. 1996. Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research* 5: 283-310.
- Statistique Canada. 2003. *Méthodes et pratiques d'enquête*, 12-587-XPF.
- Westat. 2002. *WesVar 4.2 User's Guide*. Westat, USA.
- Wolter, K.M. 1985. *Introduction to Variance Estimation*. Springer-Verlag, New York.
- Yung, W. 1997. Variance estimation for public use microdata files. *Recueil du Symposium 97: Nouvelles orientations pour les enquêtes et les recensements*, Statistique Canada, 91-95.

Annexe : Outils d'analyse fondée sur le plan de sondage, disponibles dans quelques logiciels

Logiciel	SUDAAN 8.02	WesVar 4.2	Stata 8.0	SAS 8.2	SAS 9.1	Bootvar and NLSCY VES	Bootmac
Méthode d'estimation de la variance	BRR (Bootstrap) Jackknife Taylor Series	BRR (Bootstrap) Jackknife	Taylor Series	Taylor Series	Taylor Series	Bootstrap	Bootstrap
Modélisation							
régression linéaire	<i>proc regress</i>	oui	<i>svyreg</i>	<i>proc surveyreg</i>	<i>proc surveyreg</i>	oui	oui
régression par variables instrumentales	non	non	<i>svyireg</i>	non	non	non	non
régression par intervalles	non	non	<i>svyintrg</i>	non	non	non	oui
régression logistique	<i>proc logistic (rlogist)</i>	oui	<i>svylogit</i>	non	<i>proc surveylogistic</i>	oui	oui
régression probit	non	non	<i>svyprobt</i>	non	<i>proc surveylogistic</i>	non	oui
modèles logit généralisés	<i>proc multilog</i>	oui	<i>svymlog</i>	non	<i>proc surveylogistic</i>	non	non
modèles des odds proportionnels	<i>proc multilog</i>	non	<i>svyolog</i>	non	non	non	oui
régression probit ordonnée	non	non	<i>svyoprob</i>	non	<i>proc surveylogistic</i>	non	oui
régression de Poisson et log-linéaire	<i>proc loglink</i>	non	<i>svypoiss</i>	non	non	non	oui
modèles de Heckman	non	non	<i>svyheck</i>	non	non	non	non
modèles à risques proportionnels	<i>proc survival*</i>	non	non	non	non	non	oui
Statistique descriptive							
moyennes	<i>proc descript</i>	oui	<i>svymean</i>	<i>proc surveymeans</i>	<i>proc surveymeans</i>	oui	oui
totaux	<i>proc descript</i>	oui	<i>svytotal</i>	<i>proc surveymeans</i>	<i>proc surveymeans</i>	oui	oui
proportions	<i>proc descript</i>	oui	<i>svyprop</i>	non	non	oui	oui
quotients	<i>proc ratio</i>	oui	<i>svyratio</i>	non	non	oui	oui
tests d'indépendance	<i>proc crosstab</i>	oui	<i>svytab</i>	non	<i>proc surveyfreq</i>	non	oui
quantiles	<i>proc descript</i>	oui	non	non	non	non	oui
Valeurs plausibles	non	oui**	non	non	non	non	non

* Série de Taylor seulement.

** Pour statistiques descriptives et régression.

Fichiers de commandes Stat/Transfer et transfert efficace des fichiers de données

Par James Chowhan

Résumé

L'utilisation des fichiers de commandes dans Stat/Transfer permet d'accélérer le transfert de plusieurs ensembles de données, de manière efficace et reproductible. La présente note décrit une méthode simple, étape par étape, pour créer des fichiers de commandes et présente un exemple de code.

Introduction

Dans cette note, nous tentons de faciliter l'utilisation des fichiers de commandes Stat/Transfer par les chercheurs. La plupart de ceux-ci utilisent avec relativement d'aisance l'interface "pointer-cliquer" de Stat/Transfer. Toutefois, quand les chercheurs doivent transférer de nombreux ensembles de données, la méthode "pointer-cliquer" peut être répétitive et fastidieuse, car il faut lancer manuellement chaque transfert de fichier. L'utilisation des fichiers de commandes accélère le transfert de nombreux ensembles de données, d'une manière efficace et reproductible. Avec les fichiers de commandes Stat/Transfer, on peut exécuter les mêmes opérations qu'avec la méthode "pointer-cliquer", mais en plus le fichier de commandes est un enregistrement réutilisable dont l'exécution génère un journal des résultats.

La véritable puissance des fichiers de commandes de Stat/Transfer réside dans l'exécution séquentielle de nombreuses tâches de transfert ou d'analyses syntaxiques. Les chercheurs peuvent en effet lancer un fichier de commandes comportant de nombreuses tâches, et le laisser s'exécuter pendant qu'ils vaquent à d'autres tâches.

Tous les exemples présentés dans cette note ont été réalisés avec la version 7.0.02 de Stat/Transfer, mais la méthode devrait fonctionner avec toute version de ce logiciel¹. De plus, dans cette note, nous traitons uniquement des fichiers de commandes, mais non de l'utilisation interactive de Stat/Transfer dans un environnement DOS (commandes de type DOS), ni de la méthode "pointer-cliquer". Veuillez consulter Keown [2004] pour une présentation de l'interface "pointer-cliquer".

II. Utilisations de Stat/Transfer

Les ensembles de données des Centres de données de recherche (CDR) sont souvent très volumineux en longueur et en largeur. Le logiciel Stat/Transfer permet de réduire le nombre de variables et/ou le nombre d'observations dans un ensemble de données, ce qui le rend plus compact et plus facile à manipuler. Une autre utilisation, encore plus importante, est le transfert

1. Les exemples dans cette note ont été testés avec les versions 6 et 7 de Stat/Transfer.

d'un ensemble de données d'un format à un autre. Par exemple, on peut vouloir convertir un ensemble de données Stat/Transfer en un fichier SPSS ou un fichier Stata. L'utilisation de fichiers de type syntaxe ou commande est très utile pour conserver les fichiers journaux des tâches et des transformations passées et présentes, ou encore pour reproduire des tâches antérieures sur des ensembles de données nouveaux, identiques ou différents.

III. Création d'un fichier de commandes

Les fichiers de commandes Stat/Transfer sont très similaires à des documents texte (affichables dans Notepad) ou ASCII, car le texte peut être facilement manipulé (copier-coller). Toutefois, les fichiers Stat/Transfer et ASCII ne sont pas identiques, la principale différence étant que Stat/Transfer ne reconnaît que les fichiers de commandes portant une extension “.stc”. Nous prenons le temps ci-dessous d'expliquer l'importance, quand on crée un fichier de commandes (particulièrement l'étape 1), de lui donner l'extension “.stc”, que Stat/Transfer reconnaîtra. C'est pourquoi il faut utiliser un éditeur de texte pour créer ce fichier spécial.

Un fichier de commandes a habituellement la forme suivante :

```
log using "c:\path\log_name"
keep [variable list]
where [variable expression relational operator condition]
copy "infilename.ex1" "outfilename.ex2" /x1 /x2 ...
quit
```

Vous pouvez taper les commandes dans le fichier, et insérer des commentaires en les faisant précéder d'une double barre oblique, “//”. Tous les noms de lecteur, de chemin d'accès et de fichier doivent être indiqués entre guillemets anglais. Cela permet d'utiliser des espaces et autres symboles non alphanumériques (comme les parenthèses) dans le chemin spécifique. Les sauts de ligne sont indiqués par des “retours à la ligne forcés” et toutes les lignes de commande se terminent par un “retour à la ligne forcé”. Par conséquent, chaque commande occupe une seule ligne. Toutes les lignes de commande doivent avoir au plus 80 caractères de longueur.

L'ordre des commandes est important pour la bonne exécution du fichier. La commande KEEP vous permet de spécifier les variables qui demeureront dans le nouvel ensemble de données. À l'inverse, la commande DROP vous permet d'indiquer les variables qui seront retirées de l'ensemble de données final. La commande WHERE indique les conditions pour les observations qui seront transférées dans le nouvel ensemble de données. La commande COPY indique le fichier qui sera transféré ou analysé, et le nouveau fichier qui sera généré. De plus, cette commande est suivie d'une seule barre oblique et de toute option requise. Les options (/x1 /x2) peuvent désigner des variables ou des fichiers. Une brève discussion des options figure à la section IV.

Pour une même commande COPY, on ne peut pas utiliser plusieurs instructions KEEP, DROP ou WHERE (en effet, si on utilise plusieurs commandes, pour chacune d'entre elles, c'est la dernière qui sera acceptée). En outre, les instructions WHERE sont limitées à une ou deux

expressions/conditions, tout dépendant de la nature de l'instruction. Ainsi, chaque combinaison de commandes KEEP/DROP et WHERE requiert une commande COPY.

Pour créer un fichier de commandes, veuillez procéder comme suit² :

- 1) Ouvrez votre éditeur de texte préféré (Notepad ou l'éditeur MS-DOS, entre autres).
- 2) Quand l'éditeur est ouvert, tapez le texte du fichier de commandes (c'est-à-dire le code et les commandes).
- 3) Enregistrez ce fichier :
 - i. Sélectionnez "file ", puis "save as ".
 - ii. Sélectionnez le lecteur et le dossier d'accès (lecteur, dossier, sous-dossier).
 - iii. Entrez "nom_de_fichier.stc " (c'est l'étape la plus cruciale. L'extension ".stc " est requise, car sans celle-ci, Stat/Transfer ne pourra pas reconnaître le fichier, car il présume que tous les fichiers de commandes ont cette extension). Sélectionnez ensuite <ok>. Remarque : si vous utilisez un éditeur texte DOS, il se peut qu'il y ait des restrictions pour la longueur du nom du fichier³.
 - iv. Pour fermer le programme, sélectionnez "file ", puis "exit ".
- 4) Pour éditer le fichier, cliquez à droite sur le fichier et sélectionnez "edit "⁴. Vous pouvez alors éditer et entrer (en copiant et collant) le texte dans tout éditeur de textes. Conseil : Vous irez plus vite en copiant et collant le chemin d'accès à partir de la ligne d'adresse dans l'Explorateur de Windows. Après ces modifications, vous n'avez qu'à enregistrer le fichier et le fermer.
- 5) Pour exécuter directement le fichier à partir de l'environnement Windows, double-cliquez sur le fichier "nom_de_fichier.stc ".
- 6) Si l'étape 5 ne fonctionne pas, vous devrez peut-être associer les fichiers ".stc " à Stat/Transfer. Pour ce faire, cliquez à droite sur le fichier, sélectionnez "Open with ", puis choisissez "st.exe ". Le fichier "st.exe" est pour la version DOS à Stat/Transfer. Si l'option "st.exe " n'est pas affichée, sélectionnez "Browse " pour localiser cette option. Pour trouver cet exécutable, choisissez le lecteur/chemin d'accès/sous-dossier qui mène directement au répertoire Stat/Transfer, et sélectionnez dans ce répertoire le fichier "st.exe " pour associer les fichiers ".stc " à Stat/Transfer. Ainsi, vous pourrez exécuter tous les fichiers portant l'extension ".stc " en double-cliquant sur le nom du fichier.

IV. Exemple d'un fichier de commandes

Nous présentons ci-dessous un exemple de fichier de commandes utilisant un échantillon transversal tiré de l'Enquête sur la dynamique du travail et du revenu de 1996. Le nombre d'enregistrements dans ce fichier était de 76 055. Il y avait neuf variables que l'on pouvait choisir pour cet exemple (year, personid, ailbwt26, jobid, age26, sex21, clwkr1, totear1, and ttinc42)⁵.

2. En configurant votre fichier de la façon décrite, vous pourrez le lancer sous forme d'exécutable, c'est-à-dire en double-cliquant sur le nom du fichier. Les commandes seront exécutées une à la suite de l'autre et un fichier journal sera produit.

3. On peut aussi utiliser facilement la fonctionnalité de réaffectation des noms dans Windows pour donner l'extension ".stc " au fichier.

4. Il faut bien noter que si l'option "edit " ne s'affiche pas dans la liste des choix quand vous cliquez à droite sur le fichier, c'est parce que votre ordinateur ne reconnaît pas l'extension des fichiers de commandes de Stat/Transfer. Pour corriger ce problème et associer l'extension ".stc " à Stat/Transfer, reportez-vous à l'étape 6.

5. Les variables sont year — l'année de l'enquête; personid — un identificateur unique des personnes; ailbwt26 — le poids de la transaction; jobid — un identificateur unique d'une catégorie d'emplois; age26 — l'âge de la personne au 31 décembre; sex21 — le sexe du répondant (mâle 1 et femelle 2); clwkr1 — la classe de travailleurs

Cet exemple illustre deux tâches principales, soit la conversion du format des données vers un nouveau format (Stata) et une analyse syntaxique des données. De plus, dans cet exemple, les variables `year`, `jobid` et `ailbwt26` ne seront pas conservées, et seules les observations pour les femmes employées et âgées de plus de 25 ans et dont les salaires totaux étaient supérieurs à 80 % de leurs revenus totaux seront conservés dans l'ensemble de données final.

Veillez consulter Keown [2004] pour une discussion de la sélection des variables, particulièrement l'utilisation des variables de remplacement (*wildcards*) et du choix des variables avec des règles similaires d'affectation des noms. Toutes les techniques de sélection de variables qui s'appliquent dans l'environnement "pointer-cliquer" sont également utilisables avec la méthode du fichier de commandes.

Dans ce fichier de commandes, nous employons quelques options du processeur de commandes pendant la conversion de l'ensemble de données SAS vers un ensemble de données Stata, tant pour les variables que pour les fichiers. Les options de variables utilisées sont "o", qui désactive l'optimisation automatique des types de variables de sortie, et le double "d", afin d'empêcher les variables flottantes (ce qui préserve la longueur des variables). De plus, l'option d'écrasement de fichier "yes" ("/y") a été sélectionnée pour remplacer l'ensemble de données quand le nom dans le passage courant est identique au nom utilisé dans le passage précédent⁶.

```
Log using "H:\Projects\SLID_example\joint_log_abc"
```

```
//Log using "H:\Projects\SLID_example\joint_log_a"  
keep year personid age26 sex21 clwkr1 totear1 ttinc42  
where sex21 = 2 & clwkr1 = 01  
copy "H:\Projects\SLID_example\slid_1.sas7bdat" "H:\Projects\SLID_example\slid_a.dta" /od /y  
//quit
```

```
//Log using "H:\Projects\SLID_example\joint_log_b"  
drop year  
where age26 >= 25  
copy "H:\Projects\SLID_example\slid_a.dta" "H:\Projects\SLID_example\slid_b.dta" /od /y  
//quit
```

```
//Log using "H:\Projects\SLID_example\joint_log_c"  
keep personid age26 sex21 clwkr1 totear1 ttinc42  
where (totear1/ttinc42) > .8  
copy "H:\Projects\SLID_example\slid_b.dta" "H:\Projects\SLID_example\slid_c.dta" /od /y  
quit
```

(employé 01, travailleur familial non rémunéré 02, entreprise constituée en société avec aide rémunérée 03; entreprise constituée en société sans aide rémunérée 04; entreprise non constituée en société avec aide rémunérée 05; entreprise non constituée en société sans aide rémunérée 06; autres codes administratifs 96, 97, 98 et 99); totear1 — gains totaux provenant d'un travail rémunéré; et ttinc42 — revenu total (avant impôt).

6. Pour plus d'information sur les options, veuillez consulter les rubriques du fichier d'aide de Stat/Transfer traitant des commandes de copie et des options du processeur de commandes (pour les variables, les ensembles de données d'entrée et les messages).

Veillez noter que ces fichiers de commandes peuvent être exécutés séparément sous forme de trois fichiers distincts, ou bien en bloc, tout le texte figurant alors dans un même fichier de commandes. Cet exemple illustre comment ces programmes peuvent être exécutés séquentiellement dans un même fichier de commandes. C'est ce que l'on a fait ici en commentant la commande "quit" dans les deux premiers blocs de commandes et en gardant une seule commande "quit" à la fin du fichier. On a également commenté l'option des fichiers journaux qui auraient été générés si chaque ensemble de commandes était exécuté séparément.

La véritable puissance des fichiers de commandes de Stat/Transfer réside en ce que les chercheurs peuvent regrouper ensemble de nombreuses tâches de transfert ou d'analyses syntaxiques, car le fichier s'exécute automatiquement, permettant ainsi aux chercheurs de vaquer à d'autres occupations. Le fichier de commandes ci-dessus a généré les fichiers journaux suivants, qui documentent la conversion et la partition de l'ensemble de données. La figure 1 en donne une représentation graphique.

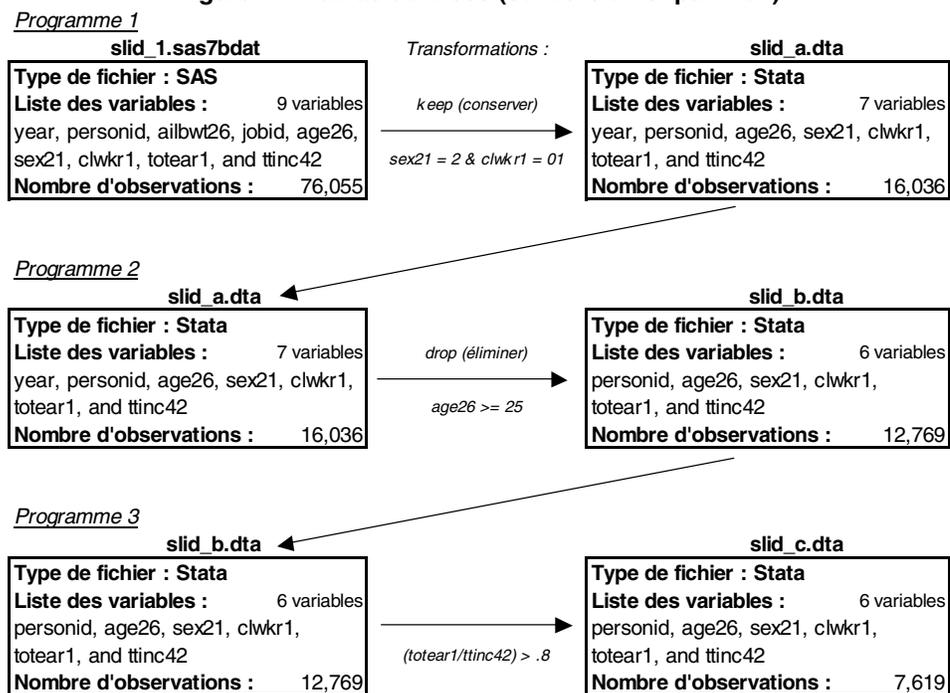
```
Transferring from SAS Data File- Versions 7/8: H:\Projects\SLID_example\slid_1.sas7bdat
Input file has 9 variables
Optimizing...
7 variables were kept
Transferring to Stata: H:\Projects\SLID_example\slid_a.dta
16036 cases were transferred (1.22 seconds)
```

```
Transferring from Stata: H:\Projects\SLID_example\slid_a.dta
Input file has 7 variables
Optimizing...
1 variable was dropped
Transferring to Stata: H:\Projects\SLID_example\slid_b.dta
12769 cases were transferred (0.58 seconds)
```

```
Transferring from Stata: H:\Projects\SLID_example\slid_b.dta
Input file has 6 variables
Optimizing...
6 variables were kept
Transferring to Stata: H:\Projects\SLID_example\slid_c.dta
7619 cases were transferred (0.46 seconds)
```

Les chercheurs qui choisissent de ne pas faire d'analyse syntaxique de leurs données mais de seulement convertir leurs fichiers courants n'ont qu'à omettre la commande WHERE. S'ils veulent conserver toutes les variables et tout simplement les transférer, ils n'ont qu'à omettre la commande KEEP (ou à les commenter). On peut également utiliser Stat/Transfer pour effectuer seulement une analyse syntaxique des données, sans convertir les données, ni les transférer dans un nouveau format. Pour ce faire, il suffit que les extensions de fichiers soient identiques dans les noms du fichier d'entrée et du fichier de sortie, bien que les noms puissent être différents. Ainsi, vous conservez votre ensemble de données de départ et vous générez un nouvel ensemble de données analysées. Il faut bien noter que les spécifications du fichier d'entrée et du fichier de sortie doivent être différentes. En d'autres mots, le nom ou le chemin d'accès des fichiers doit différer, afin qu'un ensemble de données ne puisse pas être écrasé au cours d'une même étape de la procédure.

Figure 1 : Flux de données (conversion et partition)



Enfin, pour vérifier que le fichier a été analysé et transféré correctement, on devrait calculer différentes statistiques exploratoires, comme les fréquences et les tabulations recoupées pour vérifier les totaux, les moyennes et les pourcentages afin de s'assurer que l'information a été correctement transférée. On peut utiliser le nouveau fichier ainsi généré pour effectuer les analyses voulues.

V. Conclusion

Dans la présente note, nous décrivons étape par étape une méthode simple pour créer les fichiers de commandes. Les chercheurs qui utilisent les fichiers de commandes Stat/Transfer pourront convertir ou analyser syntaxiquement les ensembles de données, d'une manière efficace et reproductible. La puissance des fichiers de commandes de Stat/Transfer réside dans la possibilité de regrouper en un seul fichier plusieurs tâches de conversion/transfert ou d'analyse syntaxique

Références

Keown, Leslie-Anne. 2004. "Production de fichiers de données "efficients" à l'aide du programme Stat/Transfer". Bulletin Technique et d'Information des Centres de Données de Recherche. 1(1):9-15. N° 12-002-XIDF au catalogue. Statistique Ottawa.

Statistique Canada, 2004, Enquête sur la dynamique du travail et du revenu. Années de données disponibles : 1992 à 2001. Ministère de l'Industrie, Ottawa.

Stat/Transfer. 2000. File Transfer Utility for Windows, Version 6.0. Circle Systems Inc., Seattle, Washington. (<http://www.stattransfer.com>)

Note technique

Note sur les identificateurs dans l'Enquête longitudinale nationale sur les enfants et les jeunes

Par Franck Larouche

Lors du traitement des données du cycle 4 de l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ), des révisions historiques ont été apportées au trois premiers cycles de l'enquête afin de corriger des erreurs et faire une mise à jour des données. La version originale des données a été remplacée par la nouvelle version au cours de l'automne 2003.

Au cours du traitement, une attention particulière a été portée à la variable PERSRUK (l'identificateur au niveau de la personne) et à la variable FIELDRUK (l'identificateur au niveau du ménage). D'abord, elles ont été normalisées de manière à ce qu'elles aient la même longueur dans tous les cycles de l'enquête ce qui était nécessaire pour fusionner les fichiers des différents cycles. Le FIELDRUK est maintenant une variable de 12 caractères de long et le PERSRUK, une variable de 14 caractères. Deuxièmement, les deux identificateurs ont été mis à jours. Quelques FIELDRUKs et PERSRUKs étaient erronés. Des corrections ont été effectuées et tous les FIELDRUKs et PERSRUKs devraient maintenant être valides.

L'identificateur unique pour l'enfant (PERSRUK) reste le même dans tous les cycles. Lorsqu'un enfant est ajouté à l'échantillon à un cycle donné, un PERSRUK unique est attribué à l'enfant qui restera le même dans les cycles subséquents. L'identificateur unique au niveau du ménage (FIELDRUK) est constant pour les enfants qui demeurent au sein de leur ménage d'origine. Au fil du temps, certains enfants quittent ce ménage. Cela peut arriver pour différentes raisons : les familles se séparent, les enfants quittent le foyer familial, etc. Aussi, l'identificateur au niveau du ménage (FIELDRUK) n'est pas nécessairement le même dans le temps, le numéro d'identification change si l'enfant change de ménage.

Le même niveau d'attention n'a pas été accordé aux autres identificateurs inclus dans la base de données, soit, la variable CHILDID (un identificateur au niveau de l'enfant) et la variable _IDHD01 (un identificateur au niveau du ménage). Ces identificateurs ont été créés pour les fichiers publics et ils se retrouvent par défaut dans les fichiers maîtres. Lorsque les fichiers maîtres sont utilisés, la variable PERSRUK devrait être utilisée pour lier les différents fichiers de données de l'enquête entre eux et la variable FIELDRUK pour déterminer le ménage.

Directives pour les auteurs

Les articles portant sur les questions méthodologiques et les sujets techniques reliés aux données qui se trouvent dans les CDR sont appropriés pour le Bulletin technique et d'information.

Langage du matériel soumis

Les manuscrits peuvent être soumis en français ou en anglais. Une fois accepté, les manuscrits seront traduits dans la deuxième langue officielle avant de les publier.

Longueur d'une soumission

Les articles ne doivent pas dépasser 20 pages à double interligne. Le Bulletin accepte également les notes et les commentaires brefs (idéalement, trois pages ou moins) traitant sur des solutions rapide aux problèmes analytiques soulevées antérieurement dans le Bulletin ou par les chercheurs collègues.

Le format électronique et la mise en page des manuscrits

Les manuscrits doivent être en format "Microsoft Word (.doc)". Les auteurs peuvent les soumettre par courrier ordinaire sur disquette ou disque compact. Ils peuvent également les envoyer comme attachement à un courriel.

Les noms des auteurs, le nom de l'établissement principal, et les coordonnées (numéro de téléphone, adresse postale et adresse électronique) du chercheur principal doivent paraître à la page couverture du manuscrit.

Les auteurs doivent se servir de la police Times New Roman de 12 points et des marges de 1 pouce (2,5 cm) en rédigeant leurs manuscrits.

Nous mettons la majuscule qu'au premier mot du titre (p.e. Pour une utilisation plus conviviale de la méthode bootstrap...).

Nous nous servons des caractères gras que pour les entêtes. Il ne faut pas souligner les mots ou les phrases ni faut il se servir des caractères en italiques pour les entêtes.

Les notes bas de page et les références doivent être à simple interligne. Les auteurs sont invités de consulter les lignes directrices relatives à la rédaction (voir ci-dessous).

Le format et mise en page des graphiques et tableaux

Les tableaux et graphiques doivent être soumis en format « Microsoft Excel (.xls) » ou en format séparation par virgule (.csv). Le nom des dossiers doit indiquer le contenu (p.e. tableau1, graphique6, etc.).

Les auteurs peuvent les soumettre par courrier ordinaire sur disquette ou disque compact. Ils peuvent également les envoyer comme attachement à un courriel. Les auteurs sont priés de suivre les directives du *Guide de rédaction* de Statistique Canada pour la rédaction des tableaux et graphiques.

Indiquez dans le texte l'emplacement des tableaux et graphiques plutôt que de les placer pas dans le texte. Servez vous du titre suivi par le nom du fichier entre parenthèses. p.e.

Graphique 6. La consommation du chocolat par les enfants au Canada, 2000 (graphique6)

Les expressions mathématiques

Toutes les expressions mathématiques doivent être dissociées du texte. Les équations doivent être numérotées, le numéro devant figurer à la droite de l'équation, aligné à la marge.

Guide de rédaction à l'intention des auteurs

Les auteurs sont priés de se servir du *Guide de rédaction* de Statistique Canada. Vous pouvez en procurer une copie du Comité de révision à l'adresse indiquée ci-dessous.

Où soumettre les manuscrits

Envoyez les manuscrits et toutes communications reliées au Bulletin au Comité de révision.

- Adresse électronique – rdc-cdr@statcan.ca

- Par la poste ordinaire :

Comité de révision, Bulletin technique et d'information du programme CDR

Centre de données pour la recherche McMaster, Statistique Canada

Mills Memorial Library

Library Room 217

1280 Main Street West

Hamilton (Ontario), L8S 4L6

Canada

Révision des soumissions

Le processus de révision initiale des articles relève du Comité de rédaction. Les rédacteurs peuvent inviter des auteurs ayant déjà publié des articles dans le BTI ou des spécialistes à participer au processus. Les articles soumis au Bulletin font l'objet d'une révision permettant d'en assurer l'exactitude, la cohérence et la qualité.

Au terme du processus de révision initiale par le Comité de rédaction, les articles sont soumis à un examen par les pairs et à un examen interne. L'examen par les pairs sera effectué conformément à la Politique concernant l'évaluation des produits d'information de Statistique Canada. En outre, des cadres supérieurs de Statistique Canada procéderont à des examens internes pour s'assurer que le matériel respecte les directives et les normes du Bureau et qu'il ne porte pas atteinte à la réputation d'impartialité politique, d'objectivité et de neutralité de Statistique Canada.

Veillez communiquer avec le comité de révision à l'adresse ci haut pour des plus amples renseignements.