

N° 12-002-X au catalogue

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



Printemps 2009, vol. 4, n° 1



Statistique
Canada

Statistics
Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca. Vous pouvez également communiquer avec nous par courriel à infostats@statcan.gc.ca ou par téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

Centre de contact national de Statistique Canada

Numéros sans frais (Canada et États-Unis) :

Service de renseignements	1-800-263-1136
Service national d'appareils de télécommunications pour les malentendants	1-800-363-7629
Télécopieur	1-877-287-4369

Appels locaux ou internationaux :

Service de renseignements	1-613-951-8116
Télécopieur	1-613-951-0581

Programme des services de dépôt

Service de renseignements	1-800-635-7943
Télécopieur	1-800-565-7757

Comment accéder à ce produit

Le produit n° 12-002-X au catalogue est disponible gratuitement sous format électronique. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.gc.ca et de choisir la rubrique « Publications ».

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « À propos de nous » > « Offrir des services aux Canadiens ».

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

Printemps 2009, vol. 4, n° 1

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2009

Tous droits réservés. Le contenu de la présente publication électronique peut être reproduit en tout ou en partie, et par quelque moyen que ce soit, sans autre permission de Statistique Canada, sous réserve que la reproduction soit effectuée uniquement à des fins d'étude privée, de recherche, de critique, de compte rendu ou en vue d'en préparer un résumé destiné aux journaux et/ou à des fins non commerciales. Statistique Canada doit être cité comme suit : Source (ou « Adapté de », s'il y a lieu) : Statistique Canada, année de publication, nom du produit, numéro au catalogue, volume et numéro, période de référence et page(s). Autrement, il est interdit de reproduire le contenu de la présente publication, ou de l'emmagasiner dans un système d'extraction, ou de le transmettre sous quelque forme ou par quelque moyen que ce soit, reproduction électronique, mécanique, photographique, pour quelque fin que ce soit, sans l'autorisation écrite préalable des Services d'octroi de licences, Division des services à la clientèle, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Avril 2009

N° 12-002-X, vol 4, n° 1 au catalogue
ISSN 1710-2200

Périodicité : irrégulier

Ottawa

This publication is available in English upon request (catalogue no. 12-002-X).

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 et aux logiciels pertinents des CDR.

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: Denis Gonthier, Heather Hobson, Logan McLeod

Table des matières

Les articles

Bradley A. Corbett, and Michael A. Busseri		
	Suivi des variables mesurées à maintes reprises dans l'Enquête longitudinale nationale sur les enfants et les jeunes : illustration fondée sur le bénévolat à l'adolescence	6
Kailing Shen		
	Construction personnalisée des données de durée : Un exemple de dérivation des variables de l'assurance-chômage à l'aide de SPSS	11
Comité de révision,	Directives pour les auteurs	18

Suivi des variables mesurées à maintes reprises dans l'Enquête longitudinale nationale sur les enfants et les jeunes : illustration fondée sur le bénévolat à l'adolescence

par Bradley A. Corbett¹ et Michael A. Busseri²

Résumé

Les chercheurs peuvent examiner l'évolution des tendances dans le temps en procédant à l'examen des réponses aux questions posées à maintes reprises aux mêmes répondants durant plusieurs cycles de données longitudinales. L'utilisation de ces réponses mesurées à maintes reprises peut souvent être difficile. Le présent article examine les tendances dans les activités de bénévolat des jeunes à l'aide des données de l'Enquête longitudinale nationale sur les enfants et les jeunes, afin de faire ressortir plusieurs questions dont les chercheurs devraient tenir compte en utilisant les mesures itératives.

1. Introduction

Une analyse longitudinale fondée sur l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ) exige le couplage des renseignements d'un répondant entre plusieurs enquêtes. Ces renseignements, ou variables mesurées à maintes reprises, peuvent être contenus dans certains fichiers de données possibles produits par l'ELNEJ.³ Pour créer un nouvel ensemble de données constitué de mesures itératives, le chercheur doit connaître le nom de la mesure itérative et le fichier de données la contenant pour chaque cycle utilisé dans l'analyse. Dans cet article, nous examinons les défis que présente la constitution d'un ensemble de données longitudinales à partir de l'ELNEJ en utilisant l'exemple des activités bénévoles autodéclarées des adolescents âgés de 12 à 19 ans.

Les activités organisées, comme le bénévolat et le service communautaire, peuvent fournir des contextes de développement positifs (Mahoney et autres, 2006). L'ELNEJ permet d'examiner l'évolution des tendances fondées sur la population dans le temps relativement à la participation des adolescents aux activités organisées. Pour réaliser cette analyse, il faut suivre dans le temps les mesures itératives pour la même personne afin de déterminer si le comportement en matière de bénévolat a changé.

Il est possible de trouver les renseignements sur la même personne contenus dans divers fichiers de données de l'ELNEJ en utilisant l'identificateur de personne unique « PERSRUK ». La variable « PERSRUK » permet au chercheur de fusionner les variables de fichiers de données multiples.

1. Auteur de correspondance Bradley A. Corbett, Centre de données de recherche de l'Université Western Ontario, Bâtiment des sciences sociales 1030, Université Western Ontario, 1151, rue Richmond, London, Ontario, Canada N6A 5C2, (519) 850-2971, bcorbet@uwo.ca, associé à Statistique Canada, au Centre de données de recherche et au Département de sociologie, Université Western Ontario.

2. Brock Research Institute for Youth Studies.

3. Cinq cycles de données étaient disponibles au moment de la rédaction de l'article

En plus d'identifier les données sur la même personne dans le temps, le chercheur doit tenir compte de trois autres facteurs. Premièrement, il doit déterminer dans quels cycles de l'ELNEJ les variables d'intérêt ont été recueillies. Deuxièmement, il doit évaluer l'uniformité de la formulation des questions et des options de réponses entre les cycles. Troisièmement, il doit déterminer le nom de la variable utilisé pour la mesure itérative dans chaque cycle. La norme de nomenclature des variables de l'ELNEJ est formée en partie de l'emplacement de la question de l'enquête et de l'instrument d'enquête utilisé pour recueillir les réponses à la question. Pour certaines variables, le suivi d'une mesure itérative entre les cycles pour la même personne est relativement facile puisqu'une question identique est utilisée dans chaque cycle et que seule la première lettre du nom de la variable change entre les cycles.⁴ Dans tous les autres cas, la situation est plus complexe.

2. Exemple utilisant les activités bénévoles des adolescents âgés de 12 à 15 ans

Les chercheurs qui désirent inclure les personnes de niveau d'âge multiples dans chaque cycle doivent déterminer le nom de la variable pour la mesure d'intérêt pour chaque niveau d'âge dans chacun des cycles. Même si la question de l'enquête est libellée de façon identique entre les cycles, elle peut avoir deux noms de variables différents dans le même cycle, selon l'instrument d'enquête servant à recueillir les réponses. Par exemple, l'instrument d'enquête utilisé pour recueillir les renseignements sur les activités bénévoles des adolescents âgés de 12 à 15 ans est un questionnaire auto-administré. Toutefois, l'instrument retenu pour les adolescents âgés de 16 à 19 ans est un questionnaire rempli par l'intervieweur. Par conséquent, les deux instruments d'enquête posent une question identique et recueillent les mêmes renseignements, mais entraînent un nom de variable différent.

Aux fins d'évaluation, les activités bénévoles dans l'ELNEJ font l'objet d'une collecte de données à l'âge de 12 ans dans le Cycle 2, puis jusqu'à l'âge de 19 ans dans le Cycle 5. Le tableau 1 illustre huit mesures de bénévolat demandées dans l'ELNEJ. Les sept premières questions sont notées de façon dichotomique (oui/non), alors que la huitième question évalue la fréquence du bénévolat de tous les jours à moins d'une fois par mois.

Cinq des questions sur le bénévolat, les questions 2 à 6, sont posées aux répondants âgés de 12 ans ou plus dans les Cycles 2 à 5. Les réponses à ces cinq questions permettent d'analyser l'évolution intra-individuelle du bénévolat dans le temps, ainsi qu'une estimation des trajectoires au niveau de la population dans la participation bénévole des adolescents. Les questions 1 à 7 sont posées seulement pour des groupes d'âge précis dans certains cycles. Par exemple, la question 1 est posée uniquement aux répondants âgés de 12 et 13 ans dans les Cycles 2, 3 et 4.

4. Pour un examen des normes de nomenclature des variables de l'ELNEJ, voir le Guide de l'utilisateur des microdonnées de l'Enquête longitudinale nationale sur les enfants et les jeunes, Cycle 5, pp. 47-50.

Tableau 1 Questions sur le bénévolat de l'Enquête longitudinale nationale sur les enfants et les jeunes par cycle d'enquête

Au cours des 12 derniers mois, as-tu fait du bénévolat ou aidé sans être payé€dans les domaines suivants :	Cycle			
	2	3	4	5
1. Activités à l'école	O	O	O	..
2. Soutien à une cause	O	O	O	O
3. Collecte de fonds	O	O	O	O
4. Travail communautaire	O	O	O	O
5. Aide à des voisins ou de la parenté	O	O	O	O
6. Autres activités bénévoles	O	O	O	O
7. Je n'ai fait aucune de ces activités sans être payé(e)	O	O	O ¹	O ¹
8. Au cours des 12 derniers mois, à quelle fréquence as-tu fait du bénévolat ou aidé sans être payé(e)?	O	O

O La question était incluse pour tous les groupes d'âge.

O¹ La question était incluse pour certains groupes d'âge.

.. indisponible pour une période de référence

Le tableau 2 donne une ventilation des variables d'activité bénévole dichotomiques par cycle d'enquête et groupe d'âge. Comme on peut le voir, la même question peut avoir un nom de variable différent dans le même cycle selon l'âge du répondant. Les noms des variables de bénévolat pour les adolescents âgés de 15 ans et moins et les adolescents âgés de 16 ans et plus sont différents. Par exemple, pour la question 5, la variable s'appelle DATCBQ5E dans le Cycle 4 pour les adolescents âgés de 12 à 15 ans, mais DACYD12D pour les adolescents âgés de 16 et 17 ans. Cette différence tient au fait que les réponses à la question sont saisies par deux instruments d'enquête différents.

En plus des différences intracycles des noms de variables, la norme de nomenclature des variables change entre les cycles. Par exemple, pour suivre la réponse d'une personne pour le point 2, la réponse des adolescents âgés de 12 ans dans le Cycle 2 se trouve à la variable BATCBQ5B, mais à la variable CATCBQ5B pour les mêmes adolescents dans le Cycle 3, âgés maintenant de 14 ans. Il est à noter que seule la première lettre du nom de la variable diffère, ce qui indique un changement de cycle. Toutefois, la réponse faite par les mêmes adolescents dans le Cycle 4, à l'âge de 16 ans, se trouve à la variable DACYD12A. Le nom de cette variable diverge sensiblement des deux autres, l'instrument d'enquête utilisé étant un questionnaire du jeune rempli par l'intervieweur au lieu d'un questionnaire auto-administré.

Il faut aussi tenir compte du fait que les variables pour divers groupes d'âge peuvent exister dans des fichiers de données différents. Par exemple, dans le Cycle 5, la variable EATCBQ5B est située dans le fichier de données des 10 à 19 ans (NLSCY_02_05_1019_mas) pour les adolescents âgés de 12 à 15 ans. Toutefois, les réponses pour la variable correspondante pour les adolescents âgés de 16 à 19 ans, EACYD12A, sont situées dans le fichier longitudinal du Cycle 5 (NLSCY_02_C5_LONG_mas).

Tableau 2 Noms des variables dichotomiques sur le bénévolat de l'Enquête longitudinale nationale sur les enfants et les jeunes par cycle, groupe d'âge et type de bénévolat

Type de bénévolat ¹								
Cycle	Groupe d'âge	À l'école	Soutien à une cause	Collecte de fonds	Travail communautaire	Aide à des voisins	Autre activité	Aucune sans être payé(e)
2	12 à 13 ans	BATCBQ5A	BATCBQ5B	BATCBQ5C	BATCBQ5D	BATCBQ5E	BATCBQ5F	BATCBQ5G
3	12 à 13 ans	CATCBQ5A	CATCBQ5B	CATCBQ5C	CATCBQ5D	CATCBQ5E	CATCBQ5F	CATCBQ5G
	14 à 15 ans	..	CATCBQ5B	CATCBQ5C	CATCBQ5D	CATCBQ5E	CATCBQ5F	CATCBQ5G
4	12 à 13 ans	DATCDQ5A	DATCBQ5B	DATCBQ5C	DATCBQ5D	DATCBQ5E	DATCBQ5F	DATCBQ5G
	14 à 15 ans	..	DATCBQ5B	DATCBQ5C	DATCBQ5D	DATCBQ5E	DATCBQ5F	DATCBQ5G
	16 à 17 ans	..	DACYD12A	DACYD12B	DACYD12C	DACYD12D	DACYD12E	..
5	12 à 13 ans	..	EATCBQ5B	EATCBQ5C	EATCBQ5D	EATCBQ5E	EATCBQ5F	EATCBQ5G
	14 à 15 ans	..	EATCBQ5B	EATCBQ5C	EATCBQ5D	EATCBQ5E	EATCBQ5F	EATCBQ5G
	16 à 17 ans	..	EACYD12A	EACYE12B	EACYE12C	EACYD12D	EACYD12E	..
	18 à 19 ans	..	EACYD12A	EACYE12B	EACYE12C	EACYD12D	EACYD12E	..

1. Les sept variables sur le bénévolat notées de façon dichotomique sont indiquées. La huitième mesure de bénévolat, une évaluation globale de fréquence, n'est pas indiquée.

.. indisponible pour une période de référence

3. Conclusion

En résumé, les chercheurs qui désirent examiner l'évolution intra-individuelle pour la même variable dans le temps doivent prendre soin de vérifier sous divers noms de variables et dans différents fichiers de données si les réponses semblent être systématiquement absentes pour des groupes d'âge spécifiques. Les chercheurs sont invités à examiner soigneusement la documentation et les livres de codes de l'ELNEJ pour vérifier que la question et les réponses restent uniformes entre les cycles pour les mesures itératives. Certaines mesures itératives ne sont recueillies que par certains instruments d'enquête ou dans certains cycles. De plus, les

variables mesurées à maintes reprises peuvent exister sous d'autres noms de variables ou dans d'autres fichiers de données.

En conclusion, nous recommandons aux chercheurs qui planifient une analyse longitudinale de variables mesurées à maintes reprises à l'aide de l'ELNEJ de déterminer avant tout la disponibilité des variables d'intérêt entre les cycles d'enquête. Dans notre propre travail, nous avons constaté qu'un tableau semblable au tableau 2 était essentiel pour suivre les variables de bénévolat entre les groupes d'âge et les cycles. En effet, les variables de bénévolat que nous comptions analyser n'étaient pas toutes demandées pour chaque groupe d'âge dans chaque cycle. En créant un tel tableau, les chercheurs peuvent obtenir des renseignements vitaux sur la disponibilité de variables clés pour leurs analyses et la faisabilité de l'étude proposée. Les chercheurs sont invités à inclure un tableau semblable au tableau 2 dans leur proposition de recherche pour l'accès aux données dans le cadre du Programme des centres de données de recherche.

Références

- Mahoney, J. L., Harris, A. L. et Eccles, J. S. (2006). Organized activity participation, positive youth development, and the over-scheduling hypothesis. *Social Policy Report*, 20, 3-30.
- Statistique Canada. 2005. « Guide de l'utilisateur des microdonnées, Enquête longitudinale nationale sur les enfants et les jeunes, Cycle 5, septembre 2002 à juin 2003. » Division des enquêtes spéciales, Ottawa.

Construction personnalisée des données de durée : Un exemple de dérivation des variables de l'assurance-chômage à l'aide de SPSS

Par Kailing Shen¹

Résumé

Composé au départ pour la recherche de l'auteur sur l'assurance-chômage (AC), cet article résume une série de procédures qui permettent une construction personnalisée de données de durée, à l'aide du logiciel SPSS et de l'Enquête sur la dynamique du travail et du revenu (EDTR). Ces procédures peuvent servir à fusionner, déduire ou faire correspondre plusieurs ensembles de données liés à la durée.

1. Introduction

L'Enquête sur la dynamique du travail et du revenu (EDTR) constitue l'une des sources de données d'une enquête canadienne par panel les plus importantes dans le cas des études sur le marché de l'emploi pour lesquelles les répercussions de la politique sur l'assurance-chômage (AC) sont demeurées l'un des enjeux fondamentaux. Mais l'utilisation de l'EDTR pour les études sur l'AC présente un problème important lié aux données – la dérivation des périodes d'AC² et des variables d'AC.³

Le présent document fournit le plan des procédures de modélisation que j'ai élaborées en utilisant l'EDTR pour les études portant sur l'AC. Il présente aussi une série générale de procédures pour la modélisation personnalisée de la durée. Les exemples donnés sont étroitement liés aux études portant sur l'AC et le marché de l'emploi, mais on s'attend à ce que la logique sous-jacente soit également utile pour d'autres domaines.

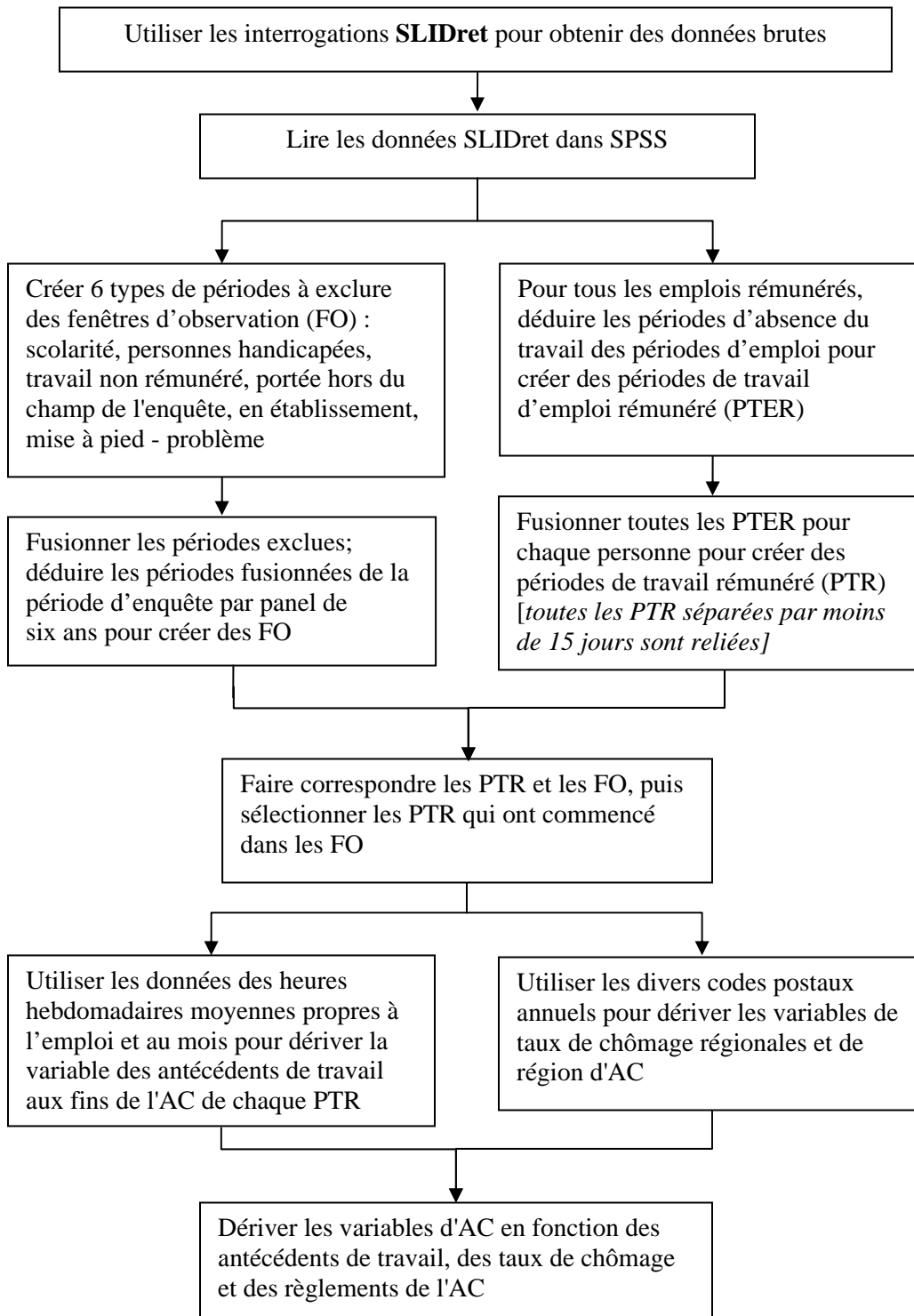
Dans la section 2, j'ai d'abord élaboré un organigramme du processus général de dérivation des périodes d'AC. Ensuite, la section 3 présente une série de procédures de programmation pour la manipulation des données liées à la durée. Enfin, la section 4 conclut en traitant de la possibilité d'appliquer les techniques évoquées dans le présent document en utilisant des langages informatiques autres que SPSS.

1. Je remercie [Darren Lauzon](#) de ses commentaires et suggestions utiles.

2. Malheureusement, le vecteur hebdomadaire d'information sur le statut de la population active, qu'on peut facilement obtenir avec l'EDTR, n'est pas utile pour la dérivation des périodes d'AC. La raison est la suivante : l'AC se rapporte exclusivement aux emplois rémunérés, tandis que la définition d'un emploi de l'EDTR comprend les emplois rémunérés ainsi que le travail autonome et d'autres formes d'emplois non rémunérés.

3. Les dérivations des périodes d'AC et des variables d'AC sont étroitement liées, puisque les variables d'AC sont toujours définies comme : à la semaine X d'une certaine période d'emploi/de chômage aux fins de l'AC. De plus, la précision des variables d'AC est étroitement liée à celle des périodes d'AC dérivées.

2. Organigramme de la modélisation des périodes d'AC à partir de l'EDTR



Certaines explications s'imposent ici :

En premier lieu, l'application SLIDret est fournie par Statistique Canada pour extraire les données de l'EDTR, qui est organisée comme une base de données relationnelle. Chaque requête d'extraction de l'EDTR doit préciser si elle est effectuée en termes de personne, d'emploi-personne ou d'absence du travail-personne, etc. Puisque nous utilisons tous ces types de données, plusieurs requêtes SLIDret sont nécessaires. De plus, il existe un nombre limite de variables pouvant être incluses dans une interrogation. Il est souvent plus facile d'obtenir plusieurs petites requêtes qu'une grande requête.⁴

En deuxième lieu, toutes les périodes dont il est question plus haut sont définies en termes de calendrier, de date de début et de date de fin. Le terme « fusionner » signifie regrouper la série de dates du calendrier visées par plusieurs périodes, tandis que « déduire » signifie supprimer une série de dates d'une autre série de dates (si elles se chevauchent).

En troisième lieu, plusieurs (mais pas la totalité) des caractéristiques institutionnelles du programme d'assurance-chômage du Canada sont considérées. En particulier, les périodes non admissibles à la réception de prestations d'AC sont exclues des fenêtres d'observation (FO), et la liaison de périodes de travail rémunérées séparées par moins de 15 jours vise à satisfaire la période d'attente de deux semaines de toute période initiale de versement de prestations d'AC.

En dernier lieu, l'EDTR comporte une variable de région d'AC, mais elle est définie uniformément en vertu des règlements de l'AC de juin 1996. Depuis 1993, trois ensembles de limites distinctes sont en vigueur : juillet 1994, juin 1996 et juillet 2000. Par conséquent, j'utilise ici des données de code postal pour appuyer les régions d'AC applicables à chaque point temporel.

Dans l'ensemble, l'obtention de variables d'AC est étroitement liée à la création de fenêtres d'observation et de périodes d'AC admissibles, soit des périodes de travail rémunéré. Une fois que les périodes sont constituées, il est facile de calculer les variables de traitement hebdomadaire, par l'AC, de chaque travailleur en faisant correspondre ces périodes avec les taux de chômage aux fins de l'AC en cours.

3. Procédures générales de manipulation des données liées à la durée

Plusieurs étapes décrites à la section 2 exigent la manipulation des données liées à la durée, comme la fusion, la déduction et la mise en correspondance (c'est-à-dire créer une correspondance entre deux ensembles logiquement indépendants ou dépendants de données liées à la durée, comme les fenêtres d'observation et les périodes de travail rémunéré, ainsi que les périodes de travail d'emploi rémunéré et les périodes de travail rémunéré). Ces types d'opérations devraient permettre aux chercheurs d'utiliser pleinement le potentiel des

⁴ Ceci peut constituer un problème grave dans le cas des variables mensuelles. Par exemple, il y aura 72 variables (6 ans * 12 mois) pour l'indicateur de scolarité. J'ai créé une interrogation distincte uniquement pour cette variable afin de tenir compte de la limite de SLIDret.

microdonnées d'enquêtes par panel grâce à la création de périodes personnalisées. Malheureusement, il n'existe pas beaucoup de documentation sur ce domaine. Voici une partie des procédures de programmation connexes que j'ai élaborées en utilisant le logiciel SPSS.

Fusion des périodes sans chevauchement

Supposons que chaque dossier dans le fichier a.sav comporte 3 champs : personid, startdate et enddate. Nous savons avec certitude que pour chaque personid, il n'existe pas de chevauchement des périodes visées par différents dossiers. Nous voulons fusionner les périodes comprises dans chaque personid, de manière à ce que deux périodes, A et B, soient fusionnées seulement si le jour suivant la date de fin de la période A correspond à la date de début de la période B. Cela pourrait se faire de la façon suivante :

Exemple 1

/*==première étape : restructurer l'ensemble de données d'entrée==*/.

```
get file='a.sav'.
```

```
varstocases /make date from startdate enddate /index=datef.
```

```
compute datef=(3-2*datefsf).
```

```
execute.
```

```
/* ici chaque dossier comporte 4 champs :*/.
```

```
/* personid date datef (1 s'il s'agit de la date de début; -1 s'il s'agit de la date de fin)*/.
```

/*==deuxième étape : calculer l'avance et le retard de date==*/.

```
sort cases by personid date datef.
```

```
split file separate by personid.
```

```
create /d_lag=lag(date 1) /d_lead=lead(date 1).
```

```
split file off.
```

/*==troisième étape : sélectionner un sous-ensemble de dates==*/.

```
compute fs=1.
```

```
if(date= date.yrday(xdate.year(d_lag),xdate.jday(d_lag)+1))&(datef=1) fs=0.
```

```
if(date= date.yrday(xdate.year(d_lead),xdate.jday(d_lead)-1))&(datef=-1) fs=0.
```

```
select if(fs=1).
```

```
execute.
```

/*==quatrième étape : générer une id de période pour l'ensemble de données de sortie ==*/.

```
split file separate by personid.
```

```
date O 1 2.
```

```
split file off.
```

```
compute spellid=cycle_.
```

```
compute dateindex=obs_.
```

```
execute.
```

/*==cinquième étape : restructurer l'ensemble de données de sortie==*/.

```
casestovars /id=personid spellid/index= dateindex.
```

```
rename variables date.1=strdate date.2=enddate.
```

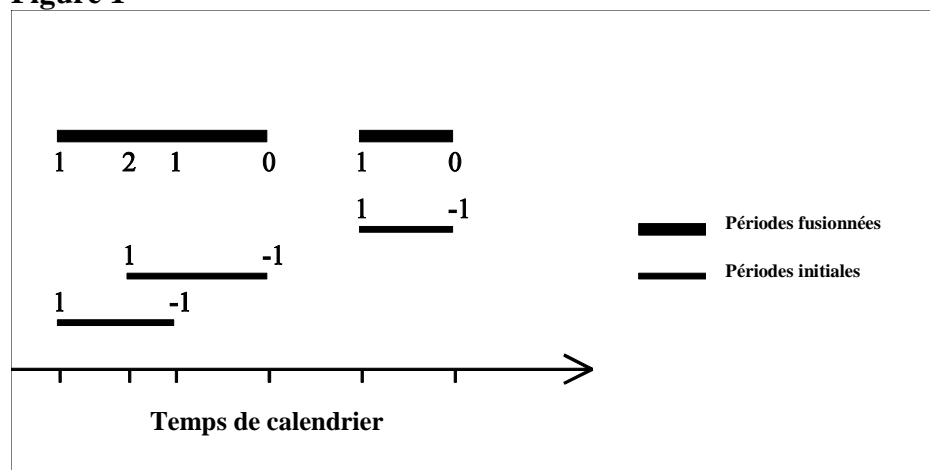
```
save outfile='merged_a.sav' /keep=personid spellid strdate enddate.
```

Fusion des périodes avec chevauchements

D'une manière générale, les étapes 2 et 3 pourraient être modifiées pour que les périodes susceptibles de se chevaucher soient fusionnées. Pour ce faire, il faut calculer le nombre de périodes actives à chaque date critique.

La figure 1 illustre cette procédure. Les traits forts représentent les périodes fusionnées, et les traits fins, les périodes initiales. Chaque période initiale est transformée en une paire d'indicateurs signés. Ensuite, les périodes fusionnées sont constituées à partir de ces points temporels, où les sommes cumulatives d'indicateurs produisent des 1 et des 0.

Figure 1



Exemple 2

/*==première étape : restructurer l'ensemble de données d'entrée==*/.

/*==deuxième étape : calculer l'avance et le retard de date==*/.

sort cases by personid date datef.

aggregate outfile='temp.sav'/break=personid date /sumf=sum(datef).

get file='temp.sav'.

split file separate by personid.

create /csumf=csum(sumf).

create /cf_lag=lag(csumf 1)/cf_lead=lead(csumf 1).

split file off.

/ csumf est le nombre de périodes actives à chaque date*/.*

/ cf_lag est le nombre de périodes actives de la date précédente*/.*

/ cf_lead est le nombre de périodes actives de la date suivante*/.*

/*==troisième étape : sélectionner un sous-ensemble de dates==*/.

compute fs=1.

```

if(csumf>0&cf_lag>0) fs=0.
if(csumf=0&(missing(cf_lag)|(cf_lag=0))) fs=0.
if(csumf<0&cf_lead<=0) fs=0.
select if(fs=1).
execute.
/*==quatrième étape : générer une id de période pour l'ensemble de données de sortie ==*/.
/*==cinquième étape : restructurer l'ensemble de données de sortie==*/.

```

Déduction d'un type de période d'un autre type

Une déduction se produit lorsqu'on veut créer des périodes de travail d'emploi à partir de périodes d'emploi et de périodes d'absence du travail. Pour ce faire, on prend le début/la fin des périodes d'absence du travail comme fin/début d'une période de travail d'emploi.⁴ L'exemple suivant démontre comment déduire les périodes de b.sav de a.sav.

Exemple 3

```

/*==première étape : restructurer les ensembles de données d'entrée==*/.
get file='a.sav'.
compute flaga=1
add files /file=* /file='b.sav'.
if(missing(flaga)) flaga=0.
execute.
varstocases /make date from startdate enddate /index=datef.
compute datef=(3-2*datefsf).
If(flaga=0) datef=-datef.
execute.
/*==deuxième étape : générer une id de période pour l'ensemble de données de sortie ==*/.
split file separate by personid.
date O 1 2.
split file off.
compute spellid=cycle_.
compute dateindex=obs_
execute.
/*==troisième étape : restructurer l'ensemble de données de sortie==*/.
casestovars /id=personid spellid/index= dateindex.
rename variables date.1=strdate date.2=enddate.
save outfile='deduct_b from a.sav' /keep=personid spellid strdate enddate.

```

⁴ La technique démontrée ici s'applique uniquement aux cas les plus simples et n'est valide que si : 1) chaque personne n'a qu'une période en a.sav; 2) pour chaque personne, les périodes en b.sav sont comprises dans sa période principale en a.sav. Dans le cas contraire, la procédure doit être modifiée par l'application des techniques correspondantes démontrées plus bas.

Correspondance de deux types de période

Cette opération peut être nécessaire s'il faut déterminer les emplois que les travailleurs ont occupés pour chaque période d'emploi, ou quelles périodes initiales correspondent à chaque période fusionnée à la figure 1. L'exemple suivant démontre comment faire correspondre les périodes en b.sav avec celles de a.sav.

Exemple 4

```
/*==première étape : restructurer les ensembles de données d'entrée==*/  
get file='a.sav'.  
compute flaga=1  
execute.  
varstocases /make startdate from startdate enddate /index=datef.  
add files /file=* /file='b.sav'.  
if(missing(flaga)) flaga=0.  
If(flaga=1) datef=(3-2*datefsf)*spellid.  
If(flaga=0) datef=0.  
execute.  
/*==deuxième étape : trier les données fusionnées et générer l'id de liaison==*/  
sort cases by personid(a) date(a) datef(d).  
split file separate by personid.  
create /cspellid=csum(datef).  
split file off.  
/*==troisième étape : sélectionner les dossiers appropriés et enregistrer==*/  
select if(flaga=0).  
compute spellid= cspellid.  
execute.  
save outfile='match_b to a.sav' /keep=personid spellid strdate enddate.
```

4. Conclusion

Le présent article présente les grandes lignes de la modélisation des variables d'AC à l'aide des données de l'EDTR. Il présente aussi une série de procédures de manipulation des données liées à la durée à l'aide du logiciel SPSS. En particulier, la capacité de SPSS de restructurer un ensemble de données pour calculer les valeurs d'avance et de retard est fondamentale ici. Par conséquent, on compte sur la découverte de techniques similaires pour mettre en œuvre les idées présentées dans cet article dans d'autres langages informatiques, comme SAS ou STATA.

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 acceptés, 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 rapides 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, interligne double, 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 *Le Guide du rédacteur*, 2^e édition.

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.

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 de *Le Guide du rédacteur*, 2^e édition. Vous pouvez en acheter une copie du Publications du gouvernement du Canada, Travaux publics et Services gouvernementaux Canada.

Où soumettre les manuscrits

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

- Adresse électronique – rdc-cdr@statcan.ca

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.