



\$

TRAITEMENT DES DONNÉES

Le traitement des données de la première vague, recueillies en 1994, est l'un des «foyers» d'activité du personnel de l'EDTR. L'exercice a nécessité plus de temps que prévu, car nous avons adopté une approche à long terme. Quoi qu'il en soit, les fondements des travaux ultérieurs ont été posés. Nous pensons ainsi être en mesure de diffuser les données en temps plus opportun dans l'avenir.

La pièce maîtresse de cette approche demeure l'élaboration du modèle de données, structure qui servira à organiser et à emmagasiner les données de l'enquête. Le modèle de l'EDTR regroupera dans une seule base les données de multiples panels couvrant de nombreuses années. En reprendre la conception sera inutile tant qu'on n'apportera pas de changement majeur à son contenu.

La base de données principale, accessible seulement aux employés de Statistique Canada, aura une structure relationnelle. Outre les données essentielles aux sorties et à l'analyse, la base contiendra tous les renseignements sur l'échantillon dont on pourrait avoir besoin pour recueillir subséquemment de nouvelles données, les indicateurs de rendement des périodes de collecte antérieures et l'information requise pour le calcul des facteurs de pondération et de la variance de

Note de la rédaction

Il me fait plaisir de prendre la relève de Philip Giles à titre de rédactrice en chef de *La Dynamique* à compter de ce numéro. Philip n'est néanmoins jamais très loin et, en réalité, demeure la personne-ressource pour les numéros qui traitent de l'analyse et de la diffusion des données de l'EDTR. Pour le prouver, voici mon adresse sur Internet : giles@statcan.ca. L'adresse postale de l'EDTR reste 11-D8, Immeuble Jean-Talon, Parc Tunney, Ottawa (Ontario), K1A 0T6, et le numéro de télécopieur est (613) 951-3253. Vous

pouvez aussi me joindre en composant le (613) 951-4353.

N'hésitez pas à communiquer avec moi pour me faire part de vos suggestions ou me poser des questions sur le bulletin d'information, ou encore mentionner les sujets susceptibles de vous intéresser.

Heather Lathe

l'échantillon. Les fichiers de microdonnées d'usage public constitueront un sous-ensemble de la base de données interne principale.

Issue des données de l'enquête, la base de données principale de l'EDTR disposera de son propre serveur. Le traitement s'effectuera dans cet environnement PC et se déroulera généralement comme suit : prétraitement, saisie, codage, vérification et imputation.

- *Prétraitement* : Il s'agit d'une série de démarches essentiellement manuelles qui comprend la suppression des données répétitives sur la même personne, attribuables aux erreurs de l'intervieweur et aux problèmes de transmission. On s'assurera que les changements de composition du ménage et que les dates de déménagement sont cohérents. Enfin, on examinera les remarques de l'intervieweur susceptibles d'entraîner certaines corrections aux données.
- *Saisie* : Cette étape nécessite la réorganisation des réponses aux questions et leur conversion en variables destinées à la base de données principale. Il s'agit d'un volet important du traitement initial des données en raison de la complexité des questions et du nombre de variables dérivées.
- *Codage* : Plusieurs champs du questionnaire ont servi à recueillir de l'information textuelle qui doit être codée de façon numérique avant qu'on l'analyse. L'attribution des codes combine des méthodes automatiques et manuelles. Les codes sont ensuite «saisis» dans la base de données.
- *Vérification* : Comparativement aux méthodes classiques de collecte des données, l'interview assistée par ordinateur permet d'effectuer une bonne partie des vérifications à l'entrevue même. Les problèmes d'enchaînement, c'est-à-dire lorsque l'intervieweur ne respecte pas l'ordre des questions, sont presque tous éliminés. Quelques corrections majeures ont été apportées au logiciel en matière de cohérence, mais certaines incohérences continuent de se glisser dans les données lors de la collecte. La vérification est une activité qui suit très rigoureusement une approche graduelle. Au départ, on n'effectue que les corrections essentielles. D'autres s'y ajoutent

davantage sur les variables qui ne font pas partie du fichier public. Les abonnés à la série de documents de recherche de l'EDTR recevront automatiquement ce document. Les autres intéressés pourront s'en procurer un exemplaire s'ils le désirent, avant d'acheter le fichier (voir «Documents de recherche» plus loin).

\$

PENSEZ LONGITUDINAL!

Avec plusieurs nouvelles enquêtes par Statistique Canada qui produiront des données tirées de panels au cours des prochaines années, les analystes canadiens et étrangers devront sérieusement envisager une approche longitudinale, outre l'approche transversale habituelle. Les nouvelles sources de données brosseront un tableau de la situation plus précis que celui obtenu grâce à l'analyse des tendances dévoilées par les données transversales.

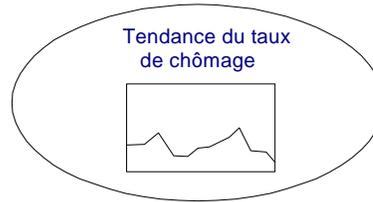
Pour mieux illustrer la différence entre l'analyse des tendances et l'analyse longitudinale, supposons que l'analyse transversale laisse entrevoir une diminution du nombre de syndiqués. On pourrait en déduire qu'un nombre accru de syndiqués ont perdu leur emploi ou qu'on en a moins embauché de nouveaux. L'analyse longitudinale permettra de déterminer l'importance relative de chaque facteur.

Étudier l'«événement»

Changer d'approche se révèle relativement simple si l'analyse se concentre sur l'événement choisi. Ainsi, au lieu de porter sur les chômeurs, l'étude de la population pourrait s'intéresser aux «personnes qui perdent leur emploi» au cours d'une période déterminée. Les données longitudinales produisent un instantané de la situation «avant» et «après» à partir duquel il est possible d'étudier les déterminants ou les conséquences de la perte de l'emploi.

Les données de l'EDTR pourraient servir à examiner d'autres événements comme les changements d'emploi, la reprise des études à temps plein, le retour à la population active après une période de prestation d'aide sociale ou la constitution d'une union de fait, pour ne citer que quelques exemples. On pourrait aussi s'intéresser aux divers «états» par lesquels passent les membres de la population, d'année en année. Par exemple, quelle proportion de la population bénéficie d'un revenu relativement stable au fil des ans?

Point de vue transversal



Point de vue Longitudinal



Délimiter la «cible mouvante»

Il s'agit de définir la population visée à un moment précis dans le temps, par exemple les personnes qui se sont mariées ou ont eu un enfant *en 1993*. Contrairement à l'analyse transversale, où les sujets de l'échantillon changent selon qu'ils répondent ou non au critère à chaque année, l'analyse longitudinale suit le même groupe d'une année à l'autre, pourvu que ses membres répondent au critère s'appliquant au moment choisi. La population peut être fixée au début ou à la fin de la période étudiée, selon qu'on désire un tableau de la situation avant ou après celle-ci.

La définition de la population cible détermine le facteur de pondération longitudinal normalement applicable pour l'année, puisque, même s'il est théoriquement constant, ce facteur est légèrement corrigé chaque année en fonction du nombre de non-réponses.

Étudier la famille par le biais de la personne

L'analyse longitudinale suit les mêmes personnes d'année en année. La logique voudrait donc qu'on en fasse autant avec les familles pour les données familiales. Mais la composition d'une famille peut évoluer avec le temps, ce qui est d'ailleurs souvent le cas, et il est donc très difficile de définir une «famille longitudinale». Il existe néanmoins une solution très simple à cela : étudier les caractéristiques familiales des particuliers. Chaque sujet devient alors l'unité d'analyse et ses caractéristiques familiales constituent ses attributs.

\$

**L'ANALYSE RÉGIONALE :
UN CAS PARTICULIER DE
CIBLE MOUVANTE**

Il y a quelque temps, un utilisateur éventuel des données nous a demandé si on pouvait obtenir le fichier de microdonnées pour une seule province. Voici pourquoi l'analyse longitudinale d'une province exige l'utilisation de la base de données complète.

Supposons par exemple qu'au début de la période d'étude, la population sélectionnée correspond aux habitants de la Colombie-Britannique et qu'on observe ses caractéristiques. À une date ou période de référence subséquente, quand vient à nouveau le temps d'examiner les caractéristiques de la population, il se pourrait que les sujets de l'échantillon se soient dispersés un peu partout au Canada et à l'étranger. On pourrait aussi définir la population à la fin de la période d'analyse et observer les caractéristiques *antérieures* des sujets, peu importe l'endroit où ils vivaient à l'époque.

Il s'ensuit que même pour l'étude d'une région déterminée, la base de données doit couvrir le Canada en entier (et l'étranger). Cela explique aussi pourquoi nous avons l'intention d'interroger les répondants longitudinaux qui déménagent dans les territoires ou aux États-Unis.

Dans quelques cas exceptionnels, la population pourrait logiquement correspondre aux personnes vivant en Colombie-Britannique à un moment précis et habitant toujours dans cette province à une date subséquente. Toutefois, en excluant une partie de la population de cette manière (c'est-à-dire en excluant ceux qui quittent la province ou qui y emménagent durant la période établie), on court le risque d'introduire un biais au niveau des caractéristiques, comparativement à ce qu'on aurait obtenu avec l'ensemble de la population.

Précisions géographiques

Pour que le fichier d'usage public ne porte pas atteinte à la vie privée des répondants, on n'y trouvera que des renseignements géographiques restreints. Il sera néanmoins possible d'obtenir des données pour de petites régions en formulant une demande précise, sous réserve des contraintes concernant la taille de l'échantillon et la nature confidentielle des résultats finals. L'EDTR permet d'obtenir des données dont le niveau de détail géographique peut correspondre à des régions économiques, des régions métropolitaines de recensement et d'autres régions définissables grâce aux codes postaux. Elle inclut également la taille approximative du lieu de résidence (partie de la variable urbaine/rurale). Le deuxième panel de l'EDTR, qui doublera la taille de l'échantillon, facilitera la définition des régions.

\$

**RÉSULTATS DE LA
QUESTION SUR LE
RECOURS AUX DONNÉES
FISCALES**

Le dernier numéro de *La Dynamique* décrivait comment on avait demandé aux répondants la permission d'accéder à leurs déclarations d'impôt en vue d'obtenir des renseignements sur les revenus. L'avantage d'une telle méthode pour le répondant est qu'il n'a pas à remplir l'interview sur le revenu. Du côté de l'enquête, l'avantage réside dans une plus grande précision du revenu déclaré et dans la possibilité de moins grandes pertes par élimination.

En mai 1995, une forte proportion de répondants ont consenti à l'utilisation des données fiscales (63 %). L'EDTR y recourra si on réussit l'appariement des dossiers. D'autres personnes ont déclaré ne pas avoir envoyé de déclaration de revenus, alors que 31 % des répondants ont refusé la proposition qui leur était présentée.

\$

**DOCUMENTS DE
RECHERCHE**

Voici une description des documents de recherche récemment diffusés que vous pouvez vous procurer individuellement (5 \$) ou par abonnement annuel (15 \$ la disquette ou 50 \$ sur papier). Pour obtenir de plus amples renseignements, veuillez communiquer avec Anne Palmer par téléphone au numéro (613) 951-2903, par télécopieur (613) 951-3253 ou par écrit à l'adresse suivante : 11-D8, Immeuble Jean-Talon, Parc Tunney, Ottawa, K1A 0T6. Les utilisateurs d'Internet sont priés de faire le GILES@STATCAN.CA.

*95-06 Interview avec rétro-information : Répercussions sur la
remémoration et les changements de situation vis-à-vis
de l'activité*

Alison Hale et Sylvie Michaud

Pour aider les répondants à rafraîchir leur mémoire, on leur rappelle les renseignements fournis à l'interview précédente pour certaines caractéristiques. Ce faisant, la qualité des données s'en trouve améliorée. Ce rapport examine l'efficacité de l'interview avec rétro-information lors de l'interview sur le travail de l'EDTR de janvier 1994.

95-07 Quelques effets de l'interview assistée par ordinateur sur la qualité des données de l'EDTR

Ruth Dibbs, Alison Hale, Robert Loverock, Sylvie Michaud

L'interview assistée par ordinateur (IAO) permet d'effectuer des contrôles de cohérence plus complexes que le permet l'interview classique avec papier et crayon sur le terrain. Les réponses présentent donc moins d'erreurs. Ce rapport explique les attentes relatives à la qualité des données à trois égards : degré d'activité de la population active, sources de revenu délicates pour le répondant et relations entre membres du ménage.

95-08 Enquête sur la dynamique du travail et du revenu : Minorités visibles et Autochtones

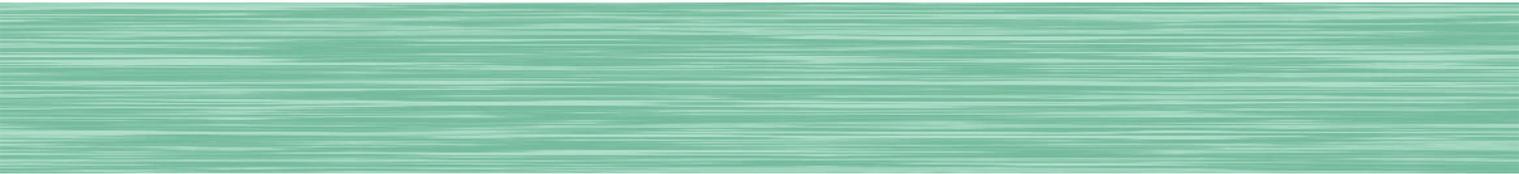
Ruth Dibbs et Tracey Leesti

Financé par le groupe de travail interministériel chargé d'étudier l'équité en matière d'emploi, ce rapport détermine l'utilité des données de l'EDTR à cet égard. On a particulièrement insisté sur deux des quatre groupes désignés par la politique d'équité en matière d'emploi : les minorités visibles et les Autochtones. Le rapport effectue des recoupements avec les données du Recensement.

95-09 Les variables dérivées de l'EDTR : Études

Heather Lathe, Philip Giles, Joanne Murray

Il arrive souvent qu'on ne puisse obtenir de façon fiable l'information dont les utilisateurs ont besoin au moyen d'une seule question. C'est pourquoi les fichiers issus des enquêtes comprennent beaucoup de «variables dérivées», c'est-à-dire des données qui ne proviennent pas directement des répondants mais qui sont calculées ou dérivées des réponses recueillies lors de l'enquête. Le personnel de l'EDTR préparera de la documentation sur toutes les variables dérivées, documentation qui expliquera notamment la méthode de dérivation. Ce



document de recherche est le deuxième d'une série insistant sur les variables éducationnelles.

95-10 Description graphique du contenu de l'EDTR
Philip Giles

L'EDTR sert à recueillir des renseignements très variés. Pour aider l'utilisateur, on a décrit le contenu de la base de données de maintes manières. Ce document présente le même contenu sous un angle différent, essentiellement graphique. La description est assez détaillée pour donner à l'utilisateur une idée de la multiplicité de l'information, mais ne fournit pas de précisions sur les variables.

95-11 La diffusion des données d'enquêtes longitudinales dans le contexte de l'Enquête sur la dynamique du travail et du revenu
Maryanne Webber

La diffusion des microdonnées provenant des enquêtes longitudinales soulève plusieurs difficultés qu'on explique ici. Le document présente aussi certaines des mesures qu'on envisage pour les surmonter. Son intention consiste surtout à lancer le débat sur les problèmes généraux de diffusion, en prenant l'EDTR comme étude de cas.

95-12 Le questionnaire et les procédures de collecte des données sur le revenu de l'EDTR - Mai 1995
Élaine Fournier, Alison Hale et Bob Kaminsky

L'EDTR a recueilli des données annuelles sur le revenu du premier panel de répondants pour la deuxième fois en mai 1995. Ce document décrit les méthodes de collecte et donne un aperçu du processus de l'interview.