



Division de l'analyse des entreprises et du marché du travail & Division de la statistique du travail

Guide pour l'analyse de

## L'Enquête sur le milieu de travail et les employés 2000



[Version 2.1: Octobre 2002]

**24 Édifice R.H. Coats  
Ottawa, Ontario  
K1A 0T6  
(CANADA)**

**Téléphone: (613) 951-4233  
Télécopieur: (613) 951-4087  
courriel: [fangtao@statcan.ca](mailto:fangtao@statcan.ca)**

[www.statcan.ca](http://www.statcan.ca)



Statistics  
Canada

Statistique  
Canada

Canada

**Note aux usagers des centres de données de recherche :** Plusieurs variables d'identification ne seront pas disponibles dans les centres de données de recherche, et ce, afin de protéger la confidentialité des renseignements fournis par les répondants. Les variables qu'on ne pourra pas obtenir dans ces centres seront indiquées à l'aide des lettres (BC) pour bureau central à l'intérieur du dictionnaire des données.

## 1 – GUIDE DE DÉMARRAGE

Le présent guide de démarrage rapide est destiné à fournir aux utilisateurs expérimentés de microdonnées l'information dont ils ont besoin pour commencer à avoir accès aux données de l'Enquête sur le milieu de travail et les employés (EMTE). Le tableau qui suit fournit l'information nécessaire pour démarrer. Veuillez lire les notes qui suivent le tableau pour assurer l'utilisation et l'interprétation appropriées des données.

[Dictionnaire électronique des données](#)

[Questionnaires](#)

1. **Utilisez les pondérations de l'enquête dans toutes vos analyses.** L'enquête sur les employeurs repose sur un plan d'échantillonnage stratifié auquel est intégré de l'information au sujet de la région, de l'industrie et de l'effectif. On sélectionne au hasard les employés à l'intérieur de chaque emplacement échantillonné. L'échantillon n'étant pas « autoreprésentatif », la non-utilisation des pondérations entraînera des estimations qui ne se rapporteront pas à une population connue. Nous préconisons fortement l'expression « estimation fondée sur le plan » pour ceux qui la connaissent.
2. **Utilisez les pondérations appropriées à l'enquête.** On dispose, avec les données de 1999 et de 2000, de trois ensembles de pondérations pour l'enquête : les poids pour les employeurs, les poids pour les employés et les poids reliés aux employeurs. La raison pour les deux premiers ensembles de poids est évidente : on peut effectuer indépendamment des études tant au niveau des employeurs qu'à celui des employés dans le cadre de l'EMTE. Cependant, pour certains emplacements, nous avons reçu les réponses des employeurs mais pas des employés. Nous intégrons, bien sûr, ces cas inactifs aux poids des employés. Ceci exige toutefois un ensemble séparé de poids (les poids reliés aux employeurs) pour les analyses sur les employeurs auquel sont incorporées les caractéristiques des employés. The 2000 weights should be used in all longitudinal analyses.

**Centre de données de recherche veuillez vous référer à l'annexe 7 pour des explications de la méthode bootstrap.**

3. **Considérez le plan d'enquête dans le calcul des variances.** Même si l'utilisation des poids appropriés à l'enquête donnera des estimations cohérentes, la plupart des logiciels sous-estiment la variance des estimations, parce qu'ils ne tiennent pas compte du plan d'enquête. À l'annexe n° 7, nous décrivons comment calculer correctement les variances (ou des approximations raisonnables) de plusieurs façons différentes. Calculer une variance appropriée est le seul moyen de déterminer la précision des estimations et les relations qui étayent vos analyses.
4. **Choisissez un modèle approprié pour vos analyses couplées.** La combinaison de variables tirées des deux enquêtes, celle sur les employeurs et celle sur les employés, ajoutera aux analyses et ouvrira souvent la voie à de nouvelles recherches. Ces études couplées exigeront cependant une sélection minutieuse du modèle statistique. Les données multiniveau ne seront pas conformes aux hypothèses de la plupart des modèles statistiques simples. L'annexe n° 5 renferme de brèves explications au sujet de certaines techniques appropriées. Elle inclut également une bibliographie portant sur des applications plus détaillées de ces techniques.
5. **Micro données :** les fichiers avec le préfixe «Im» pour les données et «ei» pour les indicateurs de correction et d'imputation.
6. **Données «dummy» :** les fichiers avec le préfixe «Dm» pour les données et «de» pour les indicateurs de correction et d'imputation.
7. **Estimés macro – totaux de contrôle :** les fichiers avec le préfixe «Ma».

**Tous les filières mentionnées de 5-7 sont disponibles dans SAS V8, SPSS et STATA.**

# ANNEXE I

## Introduction

### **Pourquoi une enquête liée sur le milieu de travail et les employés?**

Les économies des pays industrialisés évoluent constamment. Les nouvelles technologies (particulièrement les technologies de l'information), la mondialisation et la concurrence internationale ainsi que l'expansion constante des multinationales jouent un rôle crucial dans cette évolution économique. Les entreprises réagissent de plusieurs manières à cette évolution : adoption croissante des nouvelles technologies, réduction des effectifs, remodelage, « flexibilité numérique »<sup>1</sup> accrue et autres changements organisationnels. Pour les entreprises, ces nouvelles tendances constituent de nouveaux défis à relever dans le développement et la gestion des ressources humaines. Pour les décideurs politiques, l'éducation et la formation sont considérées comme des moyens stratégiques essentiels pour accroître la prospérité.

Dans cet environnement en pleine évolution, on pense que les entreprises ont subi des changements d'envergure au chapitre de l'adoption de la technologie, des changements organisationnels, des tendances de la formation, des stratégies d'entreprise, des niveaux de concurrence et des modes de recrutement de la main-d'oeuvre. Pour les travailleurs, par ailleurs, cette évolution se manifeste par des changements dans le taux de création d'emplois, la stabilité de l'emploi, les salaires et l'inégalité salariale, la formation, l'utilisation de technologies de pointe, et les contrats de travail qui leur sont offerts.

Au Canada, grâce à une série d'enquêtes bien au point sur les ménages (les travailleurs), nous avons une bonne idée des conditions pour les travailleurs en ce qui concerne les salaires et l'inégalité salariale, la stabilité de l'emploi et les licenciements, la formation, la création d'emplois et le chômage. Ce qui manque, par contre, c'est un moyen de faire le lien entre les changements qui se produisent à ces niveaux et ceux qui se produisent dans les entreprises. Ce lien est important pour comprendre le rapport entre les changements au sein du marché du travail et les pressions qui s'exercent du côté de la demande, elles-mêmes attribuables, notamment, à la concurrence mondiale, aux progrès technologiques et aux pressions en faveur de l'amélioration du capital humain. Par conséquent, l'un des principaux objectifs de l'EMTE est d'établir un lien entre les événements qui se produisent dans les établissements et leurs répercussions sur les travailleurs.

L'avantage d'une enquête liée est illustré au graphique 1, qui montre les principaux éléments du contenu des deux volets de l'EMTE. On y fait état des résultats pour les établissements comme pour les travailleurs. L'analyse des variables présentées peut se faire non seulement en fonction des caractéristiques de l'établissement, comme cela a été fait dans le cadre d'autres enquêtes auprès des entreprises, mais également en fonction des caractéristiques des travailleurs. Parallèlement, les résultats pour les travailleurs peuvent être examinés non seulement en fonction des données recueillies sur les travailleurs eux-mêmes, comme ce fut toujours le cas, mais également en fonction de nouvelles données sur les établissements.

Par exemple, le lien permettrait d'associer les changements dans les différents paliers de l'échelle salariale des employés avec les activités se déroulant dans les établissements, tels que l'adoption de la technologie, ou la concurrence dans les marchés internationaux. La plupart des rapports de recherche sur les

---

<sup>1</sup> Par flexibilité numérique, on entend les pratiques en vertu desquelles les entreprises ont recours à l'impartition et font davantage appel à des travailleurs temporaires ou à temps partiel pour améliorer la flexibilité de leurs effectifs.

inégalités salariales mentionnent que la technologie et le commerce international croissants sont des causes majeures de ces inégalités. La recherche sur de nombreuses autres questions concernant le marché du travail pourrait être approfondie avec l'existence d'un tel lien. Des questions considérées, auparavant, comme étant cruciales du côté de l'offre, souvent dans le contexte du capital humain, pourraient être considérées davantage du côté de la demande. Cela pourrait inclure des questions telles que la stabilité d'emploi, la détermination des salaires, la création et la cessation de différents types d'emplois, les niveaux de formation entre les différents types de travailleurs, etc.

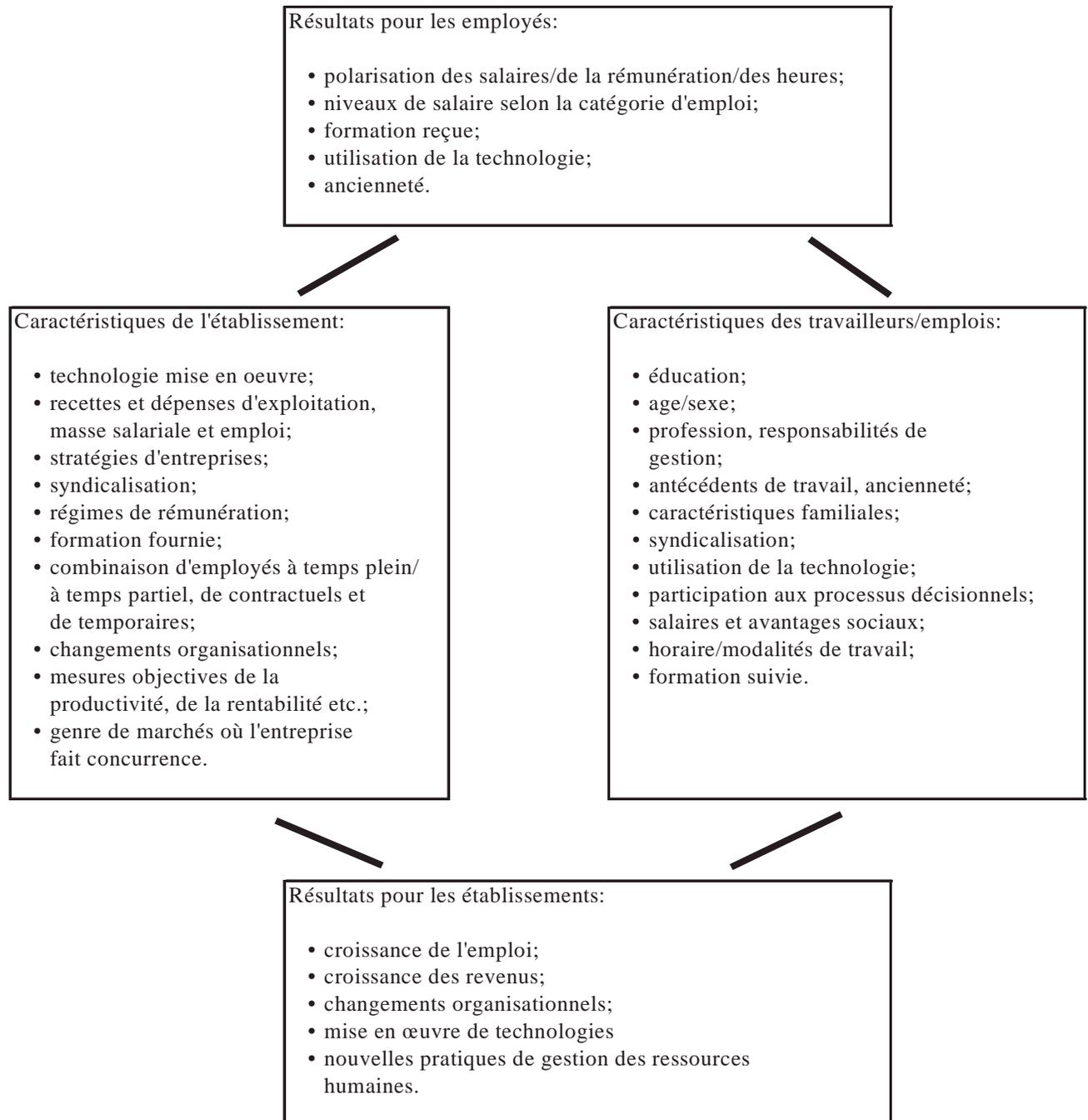
Le lien « établissement-travailleurs » pourrait aussi contribuer à l'amélioration de la mesure de bon nombre de variables au niveau de l'établissement. Les caractéristiques de la main-d'œuvre d'un établissement sont souvent un facteur important des habitudes de l'entreprise. Il reste que les données concernant les caractéristiques de la main-d'œuvre n'existent pas ou sont très mal mesurées dans les enquêtes sur les établissements. L'EMTE permet à des variables sur l'établissement – tel que l'incidence et la fréquence de la formation, la répartition des postes selon l'éducation, l'usage des technologies par les travailleurs, différentes pratiques des milieux de travail telles que les cercles de qualité, les avantages sociaux, la distribution des salaires et autres – d'être mieux mesurées que dans le passé. Les travailleurs peuvent fournir plus de données fiables et détaillées à propos de ces variables que ne peuvent le faire les établissements répondants.

Le deuxième objectif de l'enquête est de nous permettre de mieux comprendre ce qui se passe effectivement dans les entreprises à cette époque de grands changements. Combien d'entreprises ont mis en œuvre les nouvelles technologies de l'information? Sur quelle échelle? Conséquemment, quel type de formation est donné? Quels sont les changements organisationnels dans les entreprises? Quelles stratégies les entreprises emploient-elles pour prospérer pendant cette période de changement, et ces stratégies varient-elles beaucoup d'une entreprise à l'autre? Quelle est l'importance des activités et des stratégies de développement des ressources humaines, ou la plupart des établissements ne s'en préoccupent-ils tout simplement pas? Est-ce que les entreprises qui adoptent une série de stratégies en adoptent en fait beaucoup (par exemple, l'adoption de la technologie, l'innovation, le développement des ressources humaines, les changements organisationnels)? Existe-t-il un noyau de lieux de travail à haut rendement qui ont tendance à prendre des mesures sur de nombreux fronts? Voilà le genre de questions qui intéressent l'EMTE.

Même si les enquêtes actuelles auprès des ménages nous renseignent sur les changements importants qui ont lieu dans le marché du travail, il n'y a pas d'enquêtes correspondantes sur les établissements concernant les questions de l'heure. Certains travaux préliminaires ont été entrepris. L'EMTE est une initiative visant à les élargir dans le contexte d'une enquête générale sur les travailleurs et le lieu de travail.

Finalement, le troisième objectif est d'élargir l'infrastructure d'enquête. Dans de nombreux domaines, l'EMTE est perçue comme étant le développement d'une infrastructure nécessaire à la mise en œuvre de sondages sur les ménages et les entreprises. Le contenu principal sera réutilisé dans les années d'enquête successives, alors que le contenu concernant des renseignements invariables sera retiré en alternance des années d'enquête. Une fois que les résultats et la qualité des données auront été évalués à travers plusieurs années d'enquête, du nouveau contenu pourra être ajouté afin d'être à jour en ce qui concerne les besoins changeants d'information.

**Graphique 1:** Liens entre le contenu de l'enquête auprès des entreprises, le contenu de l'enquête auprès des employeurs et les résultats



## **ANNEXE 2**

### **Concepts et définitions**

#### **OBJECTIFS**

L'Enquête sur le milieu de travail et les employés (EMTE) est conçue pour explorer un large éventail de questions reliées aux employeurs et à leurs employés. Du côté de l'employeur, l'enquête vise à mettre en lumière les relations entre la compétitivité, les innovations, l'utilisation de la technologie et la gestion des ressources humaines et, du côté de l'employé, l'utilisation de la technologie, la formation, la stabilité d'emploi et les revenus.

L'enquête est unique en ce sens que les employeurs et les employés sont couplés au niveau des micro-données; les employés sont sélectionnés dans les milieux de travail échantillonnés. On dispose donc d'information provenant tant du côté de l'offre que de celui de la demande du marché du travail pour enrichir des études sur l'un ou l'autre côté du marché.

#### **Tailles des échantillons et taux de réponse**

La première édition de l'EMTE s'est déroulée à l'été (la composante « employeurs ») et à l'automne (la composante « employés ») de 1999. L'échantillon des employeurs est longitudinal – les employeurs échantillonnés seront suivis au fil du temps et des échantillons de nouveaux milieux de travail s'ajouteront périodiquement pour maintenir un échantillon représentatif. On ne suivra les employés que durant deux ans parce qu'il est difficile d'intégrer de nouveaux employeurs à l'échantillon tandis que les travailleurs changent de compagnie. On prélèvera donc de nouveaux échantillons d'employés à chaque deuxième cycle de l'enquête (premier, troisième, cinquième, etc.). Cet aspect longitudinal va permettre aux chercheurs d'étudier les résultats qu'obtiennent les employeurs et les employés au fil du temps dans le milieu du travail en évolution.

## A2.1 Tailles des échantillons et populations estimées 1999

Industrie / Taille de l'emplacement / Région	Milieux de travail		Emploi	
	Nombre de répondants	Population estimée	Nombre de répondants	Population estimée
<b>Ensemble</b>	<b>6 322</b>	<b>718 083</b>	<b>20 167</b>	<b>10 626 280</b>
<b>Industrie</b>				
Foresterie, extraction minière, de pétrole et de gaz	292	12 610	1 100	185 179
Fabrication tertiaire à forte intensité de main-d'œuvre	408	20 170	1 556	496 863
Fabrication primaire	320	7 263	1 392	398 708
Fabrication secondaire	293	11 932	1 143	367 268
Fabrication tertiaire à forte intensité de main-d'œuvre	359	16 191	1 429	584 255
Construction	608	57 736	2 021	420 546
Transport, entreposage et commerce de gros	711	88 715	2 782	1 109 092
Communications et autres services publics	421	9740	1 326	243 785
Commerce de détail et autres services commerciaux	524	234 636	1 764	2 593 009
Finance et assurances	506	36 543	1 841	505 794
Services immobiliers et de location	364	27610	1 098	182 695
Services aux entreprises	468	79 010	1 728	1 000 274
Enseignement et services de soins de santé	704	100 198	2 986	2 339 685
Information et industries culturelles	344	15 729	1 374	350 391
<b>Taille de l'emplacement</b>				
Moins de 20 employés	2 789	626 933	5 607	3 408 392
20-99 employés	1 711	77 560	7 780	2 971 669
100-499 employés	1 300	11 781	6 672	2 167 271
500 employés et plus	522	1 810	3 481	2 230 211
<b>Région</b>				
Atlantique	774	62 542	2 892	711 924
Québec	1 427	155 335	5 510	2 570 035
Ontario	1 577	260 983	5 781	4 295 566
Manitoba	420	25 651	1 556	409 578
Saskatchewan	342	28 782	1 221	328 707
Alberta	852	81 062	3 089	1 107 662
Colombie Britannique	930	103 729	3 491	1 354 071

## A2.2 Taux de réponse 1999

	Taux de réponse des employeurs (%)	Taux de réponse des employés (%)
<b>Ensemble</b>	<b>94,0</b>	<b>83,1</b>

### A2.3 Tailles des échantillons et populations estimées 2000

Industrie / Taille de l'emplacement / Région	Milieux de travail		Emploi	
	Nombre de répondants	Population estimée	Nombre de répondants	Population estimée
<b>Ensemble</b>	<b>6 068</b>	<b>668 188</b>	<b>20 167</b>	<b>10 626 280</b>
<b>Industrie</b>				
Foresterie, extraction minière, de pétrole et de gaz	278	11 580	970	192 089
Fabrication tertiaire à forte intensité de main-d'œuvre	389	18 906	1 299	497 873
Fabrication primaire	306	6 959	1 221	398 154
Fabrication secondaire	275	11 631	961	368 917
Fabrication tertiaire à forte intensité de main-d'œuvre	344	15 521	1 225	557 711
Construction	576	49 848	1 681	402 466
Transport, entreposage et commerce de gros	687	81 136	2 367	1 111 175
Communications et autres services publics	394	9 053	1 142	245 309
Commerce de détail et autres services commerciaux	540	220 991	1 538	2 585 846
Finance et assurances	485	34 613	1 621	511 809
Services immobiliers et de location	325	22 945	842	175 715
Services aux entreprises	460	76 742	1 462	1 018 702
Enseignement et services de soins de santé	680	93 833	2 652	2 339 542
Information et industries culturelles	329	14 428	1 186	349 721
<b>Taille de l'emplacement</b>				
Moins de 20 employés	2 604	574 241	4 901	3 398 935
20-99 employés	1 687	80 388	6 619	3 071 802
100-499 employés	1 280	11 763	5 724	2 171 798
500 employés et plus	497	1 797	2 923	2 112 495
<b>Région</b>				
Atlantique	746	59 071	2 578	740 971
Québec	1 365	141 823	4 525	2 422 468
Ontario	1 529	251 441	4 983	4 356 308
Manitoba	400	21 829	1 375	408 677
Saskatchewan	323	26 025	1 091	338 520
Alberta	821	76 225	2 602	1 111 862
Colombie Britannique	884	91 774	3 013	1 376 223

### A2.4 Taux de réponse 2000

	Taux de réponse des employeurs (%)	Taux de réponse des employés (%)
<b>Ensemble</b>	<b>94,0</b>	<b>87</b>

## **Population cible**

La population cible pour la composante “milieu de travail” de l’enquête est définie comme étant tous les établissements opérant au Canada et qui ont des employés rémunérés au mois de mars, à l’exception des établissements suivants :

Établissements au Yukon, aux Territoires du Nord-Ouest et au Nunavut

Établissements opérant en cultures agricoles et élevage; pêche, chasse et piégeage; ménages privés, organismes religieux et administration publique.

La population cible pour la composante « employé » est constituée de tous les employés travaillant ou en congé payé en mars dans les milieux de travail sélectionnés et qui reçoivent un état supplémentaire T-4 de l’Agence canadienne des douanes et du revenu. On comptera dans la base de sondage de l’EMTE une personne qui aura reçu un feuillet T4 de deux milieux de travail différents comme deux employés.

## **Population observée**

La population observée est le regroupement de toutes les unités pour lesquelles l’enquête peut de façon réaliste fournir de l’information. Des difficultés d’opération nuisant à l’identification de toutes les unités qui appartiennent à la population cible peuvent entraîner des différences entre la population observée et la population cible.

L’échantillon pour l’EMTE est tiré du Registre des entreprises (RE), dont l’entretien est effectué par la Division du registre des entreprises de Statistique Canada, et de listes d’employés fournies par les employeurs enquêtés.

Le Registre des entreprises est une liste de toutes les entreprises au Canada. Une mise à jour du Registre est effectuée chaque mois, à l’aide de diverses enquêtes, de suivis auprès des grandes entreprises et de données administratives.

## **Population de référence**

### *Milieu de travail*

La population de référence est établie selon les résultats du questionnaire et représente la population estimée dans les milieux de travail en se basant sur notre échantillon.

### *Employé*

La population de référence est établie selon les résultats du questionnaire et représente la population estimée d’employés en se basant sur notre échantillon.

## **Période de base**

On utilise deux périodes de base pour l’EMTE. Les questions sur la répartition de l’emploi font appel à la dernière période de paie du mois de mars de l’année de référence, tandis que les autres questions renvoient à la période des 12 derniers mois ayant pris fin au mois de mars de l’année de référence.

## **Plan d'échantillonnage**

La base de sondage est une liste de toutes les unités et elle contient l'information pour contacter ces unités et également l'information sur la classification (par ex. , la classification industrielle) de ces unités. La liste est utilisée pour l'élaboration du plan d'échantillonnage et pour la sélection de l'échantillon; au bout du compte, la liste fournit l'information sur les contacts pour les unités sélectionnées.

## **Enquête sur le milieu de travail**

La base de sondage pour la partie "milieu de travail" de l'EMTE a été créée à partir de l'information disponible sur le Registre des entreprises de Statistique Canada.

Avant la sélection de l'échantillon, les établissements de la base de sondage de l'EMTE sont stratifiés en groupes relativement homogènes. Ces groupes sont appelés strates et sont utilisés pour la répartition et la sélection de l'échantillon. La base de sondage de l'EMTE a été stratifiée par secteur d'activité (14), région (6) et taille (3), qu'on a définie à l'aide d'une estimation du nombre d'emplois. Les limites des strates d'effectif (de taille) étaient d'habitude différentes pour chaque combinaison secteur d'activité-région. Les limites d'inclusion d'une strate d'effectif particulière ont été calculées à l'aide d'une méthode fondée sur un modèle. L'échantillon a été sélectionné au moyen d'une répartition de Neyman. Ce processus a donné 252 strates et 9 144 emplacements échantillonnés.

Un poids de sondage est assigné à chaque unité échantillonnée. Il s'agit d'un facteur d'expansion relié à chaque unité échantillonnée pour obtenir des estimés pour la population à partir d'un échantillon. Par exemple, si deux unités sont sélectionnées aléatoirement et avec une même probabilité à partir d'une population de dix unités, alors chaque unité sélectionnée représente cinq unités de la population et a un poids de cinq.

Par exemple, l'EMTE inaugurale a donné lieu à la collecte de données auprès de 6 322 des 9 144 employeurs échantillonnés. Les autres employeurs étaient une combinaison de milieux de travail qui, d'après ce qu'on a déterminé, avaient cessé leurs activités, étaient inactifs durant certaines saisons, étaient des sociétés de portefeuille ou étaient hors du champ de l'enquête. La plupart des non-répondants étaient des propriétaires-exploitants qui n'avaient pas de personnel rémunéré et qui possédaient un compte de retenues sur la paye.

## **Enquête sur les employés**

La base de sondage pour la composante "employé" de l'EMTE a été créée à partir des listes d'employés mises à la disposition des interviewers par les employeurs sélectionnés. Un maximum de douze employés étaient échantillonnés en utilisant un mécanisme probabiliste. Pour les milieux de travail de moins de quatre employés, tous les employés étaient sélectionnés.

## Collecte des données

La collecte des données, la saisie des données, les vérifications préliminaires et le suivi des non-répondants ont tous été effectués dans les bureaux régionaux de Statistique Canada. En 1999, on a collecté des données sur les milieux de travail au moyen d'interviews sur place. En 2000, on a mené des interviews téléphoniques assistées par ordinateur. Il a fallu plus d'une personne-ressource pour environ 20 % des unités sondées (surtout de gros milieux de travail). Pour la composante auprès des employés, on a mené des interviews téléphoniques auprès de gens qui avaient accepté de participer à l'enquête en remplissant et en postant un formulaire sur la participation des employés.

## Vérification statistique et imputation

Lorsque toutes les données sont disponibles, elles sont analysées en profondeur. Les valeurs extrêmes sont identifiées pour une vérification manuelle, dans un ordre de priorité basé sur l'importance de l'écart avec le comportement moyen et l'importance de leur contribution aux estimés globaux.

On a supprimé les répondants qui ont choisi de ne pas participer à l'enquête (*les cas de non-réponse totale*) et on a révisé à la hausse les pondérations des autres unités pour préserver la représentativité de l'échantillon. On a utilisé pour les répondants qui n'ont pas fourni une réponse dans tous les champs à remplir (*les cas de non-réponse à une question*) une technique statistique appelée *imputation* afin d'y indiquer les valeurs manquantes tant pour les employeurs que pour les employés.

On a traité les composantes de l'EMTE indépendamment, même si les réponses à certaines questions sur le questionnaire destiné aux employés auraient pu être imputées à partir du questionnaire destiné au milieu de travail qui s'y rattachait.

## Estimation

Les valeurs déclarées (ou imputées) pour chaque milieu de travail et chaque employé dans l'échantillon sont multipliées par la pondération qui leur est assignée et ces valeurs pondérées sont additionnées pour produire des estimés. Une pondération initiale égale à l'inverse de la probabilité de sélection originale est assignée à chaque entité. Pour calculer des estimés de la variance, les pondérations de sondages initiaux sont ajustées pour forcer les totaux estimés pour chaque groupe "secteur d'activité/région" à concorder avec des totaux connus de la population. Ces pondérations ajustées sont ensuite utilisées pour calculer des estimés de moyennes ou de totaux pour différentes variables recueillies lors de l'enquête.

Les variables pour lesquelles les totaux de la population sont connus sont appelées variables auxiliaires. Elles servent à calibrer les estimés de l'enquête pour améliorer leur précision. Chaque établissement était calibré aux totaux connus dans chaque groupe secteur d'activité/région. La variable auxiliaire utilisée pour l'EMTE était *nombre total d'employés*, obtenue à partir de l'Enquête sur l'emploi, la rémunération et les heures.

Les estimés sont calculés pour plusieurs domaines d'intérêt tels que le secteur d'activité et la région.

## Qualité des données

### Règles concernant le Coefficient de variation

Toute estimation possédant un coefficient de variation supérieur à 33,5 pour cent n'est pas publiée.

Toute estimation possédant un coefficient variant entre 25 à 33,5 pour cent est annotée d'un avertissement indiquant une estimation dont le taux de variabilité est relativement élevé.

Toute enquête est sujette aux erreurs. Bien que des efforts considérables soient faits pour assurer un haut standard de qualité à travers toutes les étapes de l'enquête, les estimés sont inévitablement sujets à une certaine marge d'erreur. Des erreurs peuvent survenir pour diverses raisons : l'utilisation d'un échantillon plutôt que d'un recensement; des erreurs peuvent être commises par les répondants ou les interviewers lors de la collecte des données; des erreurs peuvent survenir lors de la saisie des données; l'imputation d'une valeur cohérente mais qui n'est pas nécessairement la vraie valeur; ou d'autres sources.

### **Erreurs d'échantillonnage**

L'erreur d'échantillonnage véritable reste inconnue; on peut cependant l'estimer à partir de l'échantillon en utilisant une mesure appelée *erreur-type*. Lorsque l'erreur-type est exprimée en pourcentage de l'estimation, on l'appelle erreur-type relative ou *coefficient de variation*.

### **Erreurs non dues à l'échantillonnage**

Certaines erreurs non dues à l'échantillonnage s'annuleront d'elles-mêmes lorsque calculées sur plusieurs observations. Par contre, les erreurs systématiques (c.-à-d. celles qui ne tendent pas à s'annuler) contribueront à un biais des estimés. Par exemple, si les répondants ont constamment tendance à sous-estimer leurs ventes, alors l'estimation des ventes totales sera inférieure à la vraie valeur dans la population. Un tel biais ne se reflète pas dans les estimés de l'erreur-type. Lorsque la taille de l'échantillon augmente, les erreurs dues à l'échantillonnage diminuent. Cependant, ce n'est pas nécessairement le cas pour les erreurs non dues à l'échantillonnage.

### **Erreurs de couverture**

Les erreurs de couverture surviennent lorsque la base de sondage ne couvre pas adéquatement la population cible. Ainsi, certaines unités appartenant à la population cible sont exclues (sous-couverture) ou sont comptées plus d'une fois (sur-couverture). De plus, des unités hors-champ peuvent être présentes sur la base de sondage (sur-couverture).

### **Erreurs de réponse**

Les erreurs de réponse surviennent lorsqu'une mauvaise interprétation des questions ou un manque d'information amène le répondant à fournir une information incorrecte; lorsqu'un répondant donne une information erronée par inadvertance; ou lorsqu'il est hésitant à dévoiler l'information juste. Les erreurs flagrantes risquent fort d'être détectées lors de l'application de règles de vérification mais d'autres peuvent tout simplement passer inaperçues.

## **Erreurs de non-réponse**

Les erreurs de non-réponse peuvent survenir lorsqu'un répondant ne répond à aucune question (non-réponse totale) ou ne répond qu'à certaines questions (non-réponse partielle). Ces erreurs peuvent avoir de graves conséquences si les non-répondants sont systématiquement différents des répondants quant aux caractéristiques de l'enquête et/ou si le taux de non-réponse est élevé.

## **Erreurs de traitement**

Les erreurs qui surviennent durant le traitement des données sont une autre composante des erreurs non dues à l'échantillonnage. Les erreurs de traitement peuvent survenir lors de la saisie des données, du codage, de la vérification, de l'imputation, du traitement des valeurs aberrantes ou lors de tout autre type de manipulation de données. Une erreur de saisie survient lorsque des données sont mal interprétées ou saisies incorrectement.

## **Interprétation simultanée des mesures d'erreur**

La mesure de l'erreur due à la non-réponse et le coefficient de variation doivent être considérés simultanément pour avoir un aperçu de la qualité des estimés. Plus le coefficient de variation est petit et plus la fraction de réponse est élevée, meilleur est l'estimé publié.

## **Confidentialité**

Les données de la présente publication sont vérifiées afin de s'assurer qu'elles respectent le caractère confidentiel des réponses fournies par les entreprises. Toute estimation agrégée de l'emploi susceptible de révéler une information propre à un répondant particulier est déclarée confidentielle et n'est donc pas publiée.

## **Réponse et non-réponse**

a) **Taux de réponse**: comprend toutes les unités répondantes qui, durant la collecte des données, fournissent des informations valables.

b) **Taux de refus**: comprend toutes les unités contactées ayant refusé de participer à l'enquête.

## Définitions des industries

<b>Code EMTE</b>	<b>Description</b>	<b>Codes à 3 chiffres du Système de Classification des Industries de l'Amérique du Nord (SCIAN)</b>
01	Foresterie, extraction minière, de pétrole et de gaz	113, 115, 211, 212, 213
02	Fabrication tertiaire à forte intensité de main-d'œuvre	311, 312, 313, 314, 315, 316, 337, 339
03	Fabrication primaire	321, 322, 324, 327, 331
04	Fabrication secondaire	325, 326, 332
05	Fabrication tertiaire à forte intensité de capital	323, 333, 334, 335, 336
06	Construction	231, 232
07	Transport, entreposage et commerce de gros	411, 412, 413, 414, 415, 416, 417, 418, 419, 481, 482, 483, 484, 485, 486, 487, 488, 493
08	Communications et autres services publics	221, 491, 492, 562
09	Commerce de détail et autres services commerciaux	441, 442, 443, 444, 445, 446, 447, 448, 451, 452, 453, 454, 713, 721, 722, 811, 812
10	Finance et assurances	521, 522, 523, 524, 526
11	Services immobiliers et de location	531, 532
12	Services aux entreprises	533, 541, 551, 561
13	Enseignement et services de soins de santé	611, 621, 622, 623, 624,
14	Information et industries culturelles	511, 512, 513, 514, 711, 712

<b>Activités économiques exclues de l'EMTE</b>	<b>Codes à 3 chiffres du Système de Classification des Industries de l'Amérique du Nord (SCIAN)</b>
Cultures agricoles, élevage	111, 112
Pêche, chasse et piégeage	114
Organismes religieux	813
Ménages privés	814
Administration publique fédérale	911
Administrations publiques provinciales et territoriales	912
Administrations publiques locales, municipales et régionales	913
Administrations publiques autochtones	914
Organismes publics internationaux et autres organismes publics extra-territoriaux	919

## **Définitions des catégories d'employés**

### **A. Employés**

Toute personne rémunérée pour les services rendus au Canada ou pour un congé payé et pour laquelle vous êtes tenu de remplir une déclaration T-4 de l'Agence canadienne des douanes et du revenu.

Employés :

- A. Les employés à temps plein travaillent habituellement 30 heures ou plus par semaine.
- B. Les employés à temps partiel travaillent habituellement moins de 30 heures par semaine.
- C. Les employés permanents n'ont pas de date de cessation d'emploi déterminée.
- D. Les employés non-permanents ont une date de cessation d'emploi déterminée ou une entente concernant la durée de la période d'emploi. Par exemple : un emploi temporaire ou saisonnier.

### **B. Entrepreneurs indépendants**

Toute personne qui fournit des produits ou des services sous contrat avec votre entreprise pour qui vous n'êtes pas tenu de remplir une déclaration T-4 de l'Agence canadienne des douanes et du revenu. Cette personne pourrait être à l'emploi d'une autre entreprise ou un travailleur à domicile. Par exemple : un consultant en informatique, une couturière travaillant à la pièce, etc.

### **C. Gestionnaires :**

#### ***1. Gestionnaires***

##### **(a) Cadres supérieurs**

Cette catégorie comprend le premier cadre dirigeant de l'emplacement ainsi que d'autres cadres dirigeants dont les responsabilités s'étendent normalement à plus d'une section interne. Ainsi, la plupart des petits emplacements auront un seul directeur. Par exemple : le président d'une société à cet emplacement unique, le gérant d'un commerce de détail, le chef d'une usine, les associés principaux dans une firme de services aux entreprises, le chef de la production, l'administrateur supérieur d'une entreprise de services publics; de même que les vice-présidents, les directeurs adjoints, les associés adjoints et les administrateurs adjoints dont les responsabilités couvrent plus d'un domaine.

##### **(b) Cadres spécialistes**

Cadres qui relèvent habituellement des cadres supérieurs et qui ont la responsabilité d'un seul domaine ou d'une seule section. Dans les petits emplacements, cette catégorie comprend habituellement les directeurs adjoints et les postes équivalents. Par exemple : les chefs ou directeurs de section (ingénierie, comptabilité, recherche et développement, personnel, informatique, commercialisation, ventes, etc.), les chefs ou directeurs responsables d'une gamme de produits, les associés adjoints ou les administrateurs adjoints qui ont la responsabilité d'un domaine particulier et les directeurs adjoints de petits emplacements (sans structure de section interne).

## **D. Personnel opérationnel :**

### ***2. Professionnels***

Employés dont les fonctions exigent normalement un diplôme universitaire de premier cycle ou l'équivalent. Par exemple : médecins, avocats, comptables, architectes, ingénieurs, économistes, professionnels des sciences, psychologues, sociologues, infirmiers autorisés, professionnels de la commercialisation et des études de marché, infirmiers de première ligne et professionnels de l'enseignement. Cette catégorie comprend les informaticiens dont les fonctions exigent normalement au moins un diplôme universitaire de premier cycle en informatique. Elle comprend aussi les chefs de projet et les superviseurs (groupe professionnel) qui ne sont pas inclus dans les cadres supérieurs (C.1 (a)) et dans les cadres spécialistes (C.1 (b)).

### ***3. Personnel technique / Métiers***

Ce groupe comprend :

#### **(a) Personnel technique / semi-professionnel**

Employés dont les fonctions exigent normalement un diplôme d'études collégiales ou l'équivalent et dont la tâche première ne touche pas le domaine des ventes ou de la commercialisation d'un produit ou service. Par exemple : technologues, techniciens de laboratoire, infirmiers auxiliaires autorisés, techniciens de l'audiovisuel, éducateurs de la petite enfance, agents de la formation technique, physiothérapeutes, secrétaires juridiques et dessinateurs. Cette catégorie comprend les programmeurs et opérateurs d'ordinateur dont les fonctions exigent habituellement un diplôme d'études collégiales. Elle comprend aussi les chefs de projet et les superviseurs (groupe semi-professionnel) qui ne sont pas inclus dans le groupe des gestionnaires (C.1.) ni dans celui des professionnels (D.1.). Elle exclut le personnel des ventes et de la commercialisation qui a un diplôme d'études non universitaires.

#### **(b) Métiers / ouvriers qualifiés de la production, de l'exploitation et de l'entretien**

Personnel opérationnel travaillant à des postes nécessitant un certificat de compétence professionnelle ou l'équivalent. Par exemple : travailleurs des métiers de la construction, machinistes, conducteurs de machine, mécaniciens de machines fixes, mécaniciens, visagistes / coiffeurs, bouchers et travailleurs des métiers de la réparation qui n'exigent habituellement pas de diplôme ou de certificat post secondaire.

### ***4. Commercialisation ou vente***

Personnel opérationnel préposé à la vente ou à la commercialisation de produits ou services. Par exemple: commis vendeurs, serveurs, télévendeurs, agents immobiliers, agents d'assurances et préposés aux prêts. Cette catégorie exclut les employés dont les fonctions exigent un diplôme universitaire ou un certificat de compétences professionnelles (professionnels (D.1.)), ceux dont les fonctions exigent un diplôme d'études collégiales (personnel technique / métiers (D.2.)) ainsi que ceux qui exercent principalement des fonctions de gestion (gestionnaires (C.1.))

### ***5. Personnel de bureau / administratif***

Personnel opérationnel qui fournit des services de bureau ou administratifs aux clients internes ou externes. Par exemple : secrétaires, opérateurs de machines de bureau, commis au classement, commis aux comptes, réceptionnistes, préposés au comptoir, commis au courrier et à la distribution, encaisseurs et experts en sinistres. Ces employés n'ont généralement pas besoin d'une formation post secondaire et n'ont pas de responsabilités dans le domaine des ventes ou de la commercialisation.

### ***6. Travailleurs de la production, de l'exploitation et de l'entretien sans métier ni certificat de compétence***

Personnel opérationnel travaillant à des postes de production ou d'entretien qui n'exigent pas de certificat de compétence professionnelle ni une formation équivalente en cours d'emploi. Par exemple : monteurs, emballeurs, trieurs, gerbeurs, conducteurs de machine, opérateurs d'équipement de transport, conducteurs, entreposeurs et personnel de nettoyage. Règle générale, les postes de cette catégorie exigent tout au plus un mois de formation pour ceux qui n'ont pas de métier ni de certificat de compétences professionnelles.

### ***7. Autres***

Si vous avez un nombre important d'employés qui ne se classent pas dans les catégories énoncées ci-dessus, veuillez en indiquer la profession.

## Définitions des catégories d'employés

<b>EMTE</b>	<b>CTP de 1991</b>
01 Gestionnaires	A011-A016; A111-A114; A121-A122; A131; A141; A211 A221-A222; A301-A303; A311-A312; A321-A324; A331-A334; A341-A343; A351-A353; A361; A371-A373; A381; A391-A392; E037
02 Professionnels	B011-B014; B021-B022; B313; B315-B318; C011-C015; C021-C023; C031-C034; C041-C048; C051-C054; C061-C063; C111-C113; C121; C152; 162-C163; D011-D014; D021-D023; D031-D032; D041-D044; D111-D112; D211; D232; E011-E012; E021-E025; E031-E036; E038; E111-E112; E121; E130-E133; E211-E214; E216; F011-F013; F021-F025 F031-F034; F111; F121; F123; F143;
03 Personnel techniques /Métiers	B111-B116; B212-B214; B311-B312; B314; B411-B415; B576; C122-C125; C131-C133; C141-C144; C151; C153-C155; C161; C164; C171-C175; D212-D219; D221-D223; D231; D233-D235; D311-D313; E215; F035-F036; F112; F122; F124-F127; F131-F132; F141-F142; F144-F145; F151-F154; G011-G016; G111; G121; G133-G134; G411-G412; G512; G611-G612; G621-G625; G631; G711-G712; G722; G812-G813; G911-G912; G921-G922; G933; G941-G942; G951; G981; H011-H019; H021-H022; H111-H113; H121-H122; H131-H134; H141-H145; H211-H217; H221-H222; H311-H312; H321-H325; H411-H418; H421-H422; H431-H435; H511-H514; H521-H523; H531-H535; H611-H612; H621-H623; H711-H714; H721-H722; H731; H736-H737; I011-I017; I021-I022; I111; I121-I122; I131-I132; I141-I142; I151; I161-I162; I171-I172 I182; J011-J016; J021-J027; J111-J114; J121-J125; J131-J134; J141-J146; J151-J154; J161-J162; J164; J171-J172; J174-J175; J181-J184; J191; J193-J197; J211; J213; J215-J216; J221-J223; J225; J227-J228;
04 Commercialisation ou ventes	G131-G132; G211; G311; G511; G513; G713-G714; G973;
05 Personnel de bureau / administratif	B211; B511-B514; B521-B524; B531-B535; B541-B543; B551-B554; B561-B563; B571-B575; G715; G721; G972;
06 Travailleurs de la production, de l'exploitation et de l'entretien sans métier ni certificat de compétence	G731-G732; G811; G814; G923-G924; G931-G932; G961-G962; G971; G982-G983; H732-H735; H811-H812; H821-H822; H831-H832; I181; I211-I216; J163; J173; J192; J212; J214; J217; J224: J226; J311-J319;

## ANNEXE 3

### **Vérification, détection des erreurs aberrantes et imputation**

On s'engage dans trois activités principales, *la vérification, la détection des erreurs aberrantes et l'imputation* pour voir à ce que les données finalement compilées soient de la plus haute qualité et pour maximiser ainsi la facilité d'utilisation de l'information rassemblée. La vérification est un processus interactif au moyen duquel on demande au répondant de confirmer de l'information qui semble suspecte ou qui ne respecte pas certaines règles générales préalablement spécifiées régissant les données à collecter. Ce processus se déroule sur le terrain durant la collecte des données.

La détection des erreurs aberrantes est une technique statistique servant à repérer des réponses anormales qui ont échappé aux vérifications ou qui n'étaient pas conformes à la structure de corrélation de la majorité des données (qui ne respectaient pas des relations connues). On peut classer les erreurs aberrantes en deux catégories : les erreurs aberrantes *représentatives* et les erreurs aberrantes *non représentatives*. Il ne faut pas traiter les erreurs de la première catégorie, parce qu'elles représentent d'autres unités de la population concernée qui affichent la même caractéristique. On devrait cependant traiter les erreurs de la seconde catégorie pour les empêcher d'avoir une incidence - soit positive ou négative - significative sur les estimations. Les deux types d'erreurs aberrantes doivent être signalées afin des exclure de l'imputation.

L'imputation est une technique statistique servant à produire de l'information que le répondant ne fournit pas. Elle peut s'appliquer aux enregistrements pour lesquels certaines données sont manquantes (certains éléments n'ont pas été recueillis) ou pour lesquelles toutes les données sont manquantes (aucun élément n'a été recueilli). Ce processus se déroule au bureau central après la réception de toutes les données et une fois qu'elles ont franchi les étapes de la détection des erreurs aberrantes et de leur traitement.

#### **Vérification des données**

Le questionnaire sur le milieu de travail renferme dix blocs distincts. Chaque bloc est axé sur un thème différent. Une seule personne sera en mesure dans la plupart des cas de répondre à toutes les questions. Si le répondant principal ou la répondante principale est incapable de fournir l'information demandée dans sa totalité, on lui demandera alors d'identifier la personne au courant de cette information.

L'outil de saisie de données d'IAO (interview assistée par ordinateur) pour les employeurs effectue les vérifications de la validité, de l'étendue et entre les champs. Ce sont là les types de vérifications effectués durant la collecte des données du premier cycle. On a commencé à élaborer un ensemble approprié de vérifications chronologiques pour les cycles subséquents. La majorité des vérifications entre les champs est confinée à un seul bloc du contenu. En cas de rejet à la vérification entre blocs, on demande alors au répondant principal de confirmer l'information.

Le calcul obligatoire d'un total positif des dépenses annuelles est un exemple de vérification de la validité. La vérification de l'étendue correspondante exige que les dépenses ne dépassent pas une limite supérieure élevée. Une vérification entre les champs connexes pour le total des dépenses annuelles offre l'assurance que la somme de la rémunération brute totale et des dépenses totales en avantages sociaux ne dépasse pas le total des dépenses annuelles.

L'application d'ITAO pour les employés effectue les vérifications de la validité, de l'étendue, entre les champs et historique. On règle les cas de rejet à la vérification durant l'interview téléphonique.

## Détection des erreurs aberrantes

L'utilisation de l'IAO pour la collecte des données réduit grandement le nombre d'erreurs dans les réponses et d'erreurs typographiques grossières. Si l'un ou l'autre type d'erreurs n'est pas détecté, on applique alors avant l'imputation un sous-programme de *détection des erreurs aberrantes à plusieurs variables* à toutes les personnes ayant fourni des réponses complètes et partielles. La technique fait appel à la distance robuste de Mahalanobis, une statistique qui mesure la distance entre une observation et le centre des données, afin d'identifier les unités pour lesquelles cette statistique dépasse un seuil préalablement spécifié défini par un percentile de la distribution  $\chi^2$  correspondante. Ce type de détection des erreurs aberrantes est effectué dans le cas du milieu de travail au niveau des microdonnées. On peut ajuster le degré de sensibilité du processus pour qu'il réponde aux besoins de l'enquête.

On n'intègre pas actuellement de pondération théorique dans le cadre de l'application du sous-programme de détection des erreurs aberrantes. Il faut respecter deux critères afin d'être en mesure d'utiliser avec succès la technique avec les données d'enquête-entreprise : a) l'homogénéité des données et b) leur symétrie. En arriver à des données homogènes évite l'obligation d'utiliser des pondérations théoriques lorsqu'on groupe des strates voisines pour accroître la résolution du sous-programme de détection des erreurs aberrantes. L'homogénéité des données réduit l'effet du plan. Le problème complexe du repérage des erreurs aberrantes à l'intérieur d'un échantillon tiré d'une population finie se réduit à un problème beaucoup plus simple : traiter les erreurs aberrantes dans le contexte d'une population infinie.

On peut atteindre l'homogénéité des données en appliquant une fonction appropriée à une ou à plusieurs variables. On devrait, une fois les données convenablement transformées (par racine carrée, log, etc., par exemple), évaluer la symétrie approximative des distributions des variables qui découlent de la transformation. Cette exigence tient du fait qu'on élabore dans la plupart des cas une théorie de détection des valeurs aberrantes pour des distributions normales contaminées. La méthode modifiée de Stahel-Donoho ne fait pas exception à la règle. Pour l'EMTE, on arrive à une symétrie approximative pour les ratios des variables continues à l'emploi total.

On peut appliquer le sous-programme de détection des erreurs aberrantes pour un cycle unique, ou pour deux cycles. À cette fin, on modifierait le vecteur de réponse  $x_i$  pour inclure des données découlant de deux cycles consécutifs. On étudiera sous peu la possibilité d'étendre la fonctionnalité de la méthode à plus de deux cycles. Notre objectif consiste à mettre au point une méthode qui comblerait le fossé entre la détection transversale des erreurs aberrantes et l'analyse robuste des séries chronologiques.

Nous effectuons également une validation des données au macro-niveau. Nous définissons pour un certain nombre de variables clés les dix plus importants contributeurs aux estimations pondérées en vue d'une analyse plus poussée. Des spécialistes repèrent les anomalies au micro et au macro-niveau et corrigent les erreurs. Une fois les erreurs corrigées, on répète le cycle de validation des données. On indique toutes les erreurs aberrantes qui restent et on les exclut de l'imputation.

## Imputation

On applique des méthodes transversales d'imputation aux non-réponses à des questions dans le cas des unités qui apparaissent pour la première fois à l'intérieur de chaque cycle. On applique des méthodes longitudinales d'imputation aux non-réponses aux questions d'un cycle si l'on dispose de données chronologiques. En l'absence d'information préalable, on traite toutes les non-réponses en modifiant les pondérations des répondants. Cette méthode suppose que les non-réponses sont entièrement le fait du hasard.

Quatre méthodes d'imputation sont présentement utilisées pour le premier cycle de la portion employeurs de l'EMTE : l'imputation déterministe, l'imputation par répartition, l'imputation par ratio et l'imputation *hot deck* pondérée. On utilise l'imputation déterministe quand on peut déduire uniquement un seul champ manquant à partir de l'information fournie. Si une composante d'une somme est manquante et si l'on dispose des autres composantes incluant la somme, par exemple, on peut alors déterminer la composante manquante.

On applique l'imputation par répartition lorsqu'on demande aux répondants de fournir un total et sa décomposition en plusieurs catégories dans les cas où il manque deux catégories ou plus. On calcule la répartition des catégories au macro-niveau et on l'applique au micro-niveau. Pour illustrer cette méthode, supposons qu'un répondant nous a donné l'emploi total, mais n'a pas été en mesure d'en fournir une ventilation par groupe professionnel. Nous appliquerions la répartition des groupes professionnels calculée au niveau industrie/taille au chiffre sur l'emploi total afin d'imputer les champs manquants.

L'imputation par ratio est principalement utilisée pour des variables continues. La valeur manquante est remplacée par la valeur ajustée d'une variable auxiliaire d'un donneur choisi de façon aléatoire d'une classe d'imputation. Habituellement, l'ajustement est la somme de tous les donneurs de la variable manquante divisée par la somme de la variable auxiliaire.

Dans le cas de la méthode *hot deck* pondérée, on impute un champ manquant à l'aide de la réponse d'un donneur approprié. On sélectionne au hasard le donneur, la probabilité de sélection étant égale au ratio de son poids divisé par la somme des poids de toutes les unités contenues dans la classe d'imputation transversale correspondante. On a adopté la méthode *hot deck* pondérée pour les quatre raisons suivantes. La méthode est facile à appliquer. Elle donne approximativement  $p$  estimations ponctuelles non biaisées (Rao, 1996). On peut construire un estimateur cohérent de la variance quand on dispose de données imputées (Rao, 1996, également). Finalement, le nombre d'ajustements postérieurs à l'imputation destinés à maintenir la cohérence interne des données reste à un niveau minimal puisque la plupart des questions sont indépendantes.

On impute les données manquantes sur le questionnaire destiné aux employés à l'aide des méthodes déterministe et *hot deck* pondérée. Pour éviter de produire des incohérences à l'intérieur des données, la plupart des champs interdépendants sont imputés sous forme de bloc. Puisque plusieurs questions font partie de cette catégorie, un système post imputation a été élaboré pour conserver les relations entre les champs.

## ANNEXE 4

### Aperçu des estimations de la population visée par l'EMTE

Le présent document vise à donner des précisions sur les différentes populations d'intérêt de l'Enquête sur le milieu de travail et les employés (EMTE). Il s'agit ici de veiller à ce que les utilisateurs des données ne soient pas seulement renseignés sur les populations qu'ils étudient, mais aussi qu'ils soient en mesure d'en informer les lecteurs des articles qu'ils produisent ou des estimations qu'ils divulguent. Il y a des mises en garde au besoin.

Nota : Dans le présent document, milieu de travail et emplacement sont synonymes. Toutes les estimations communiquées sont des estimations réelles tirées de l'EMTE. Par population cible des milieux de travail, on entend la liste des milieux de travail à l'égard desquels on a besoin d'information. Le volet de l'analyse des milieux de travail porte sur les milieux de travail échantillonnés et à l'égard desquels des données sont disponibles. La population cible des employés renvoie à la liste des employés pour lesquels on a besoin de renseignements. Le volet de l'analyse des employés porte sur la liste des employés qui ont été échantillonnés et pour lesquels des données sont disponibles.

## MILIEU DE TRAVAIL 1999

### Population cible des milieux de travail

La population cible pour le volet des milieux de travail est définie comme étant l'ensemble des emplacements exploités au Canada en mars 1999 qui comptaient au moins un employé rémunéré en mars 1999 qui a reçu un formulaire T4 Supplémentaire de l'Agence des douanes et du revenu du Canada, à l'exception de ce qui suit :

Les milieux de travail au Yukon, au Nunavut et dans les Territoires du Nord-Ouest  
Les milieux de travail engagés dans des activités liées à la culture agricole et à l'élevage d'animaux; la pêche, la chasse et le piégeage; les ménages privés, les organisations religieuses et l'administration publique.

### Volet de l'analyse des milieux de travail (6,322 emplacements)

Il s'agit de l'ensemble des milieux de travail échantillonnés qui ont répondu au questionnaire sur le milieu de travail de 1999, qui font partie de la population cible des milieux de travail de 1999 et qui comptaient au moins un employé rémunéré en mars 1999 qui a reçu un formulaire T4 Supplémentaire de l'Agence des douanes et du revenu du Canada. On peut utiliser le volet de l'analyse de concert avec les coefficients de pondération pour représenter la population cible des milieux de travail de 1999.

Nota : Le processus de repondération a permis de tenir compte des emplacements des non-répondants, ce qui fait que toutes les analyses doivent s'appuyer sur les poids finaux des milieux de travail. Les emplacements échantillonnés mais dont les activités avaient pris fin ou qui étaient hors du champ de l'enquête, qui n'avaient aucun employé ou qui étaient sous séquestre en mars 1999 sont exclus du volet de l'analyse parce qu'ils ne font pas partie de la population cible.

Voici quelques exemples qui s'appuient sur le volet de l'analyse des milieux de travail de 1999.

**Exemple 1** : Le nombre total d'emplacements au sein de la population cible des milieux de travail de 1999.

$$\hat{N} = \sum_i w_i = 718,083$$

$w_i$  – Poids final de l'emplacement

**Exemple 2** : Le nombre total d'employés dans les emplacements visés par la population cible des milieux de travail de 1999.

$$\hat{X} = \sum_i w_i x_i = 10,777,543$$

$w_i$  – Poids final de l'emplacement

$x_i$  - Emploi

**Exemple 3** : La paie brute moyenne par employé au sein de la population cible des emplacements de 1999.

$$\hat{R} = \frac{\sum_i w_i z_i}{\sum_i w_i x_i} = 31,019\$$$

$w_i$  – Poids final de l'emplacement

$x_i$  - Emploi

$z_i$  – Paie brute

**Exemple 4** : La paie brute moyenne par employé des emplacements qui offrent des avantages sociaux au sein la population cible des milieux de travail de 1999.

$$\hat{R}_d = \frac{\sum_i w_i z_i \delta_i}{\sum_i w_i x_i \delta_i} = 33,481\$$$

$w_i$  – Poids final de l'emplacement

$x_i$  - Emploi

$z_i$  – Paie brute

$\delta_i$  – Indicateur d'avantages sociaux (1 si l'emplacement offre des avantages sociaux; 0, autrement)

## EMPLOYÉ 1999

### Population cible des employés

La population cible consiste en l'ensemble des employés qui travaillaient ou qui étaient en congés rémunérés en mars 1999 qui ont reçu un formulaire T4 Supplémentaire de l'Agence des douanes et du revenu du Canada. Les employés susmentionnés doivent également être rattachés à un milieu de travail comptant parmi la population cible des milieux de travail de 1999.

### Volet de l'analyse des employés (23,540 employés)

Il s'agit de l'ensemble de tous les employés échantillonnés qui ont répondu au questionnaire de l'employé de 1999 et qui font partie de la population cible des employés de 1999. On peut se servir du volet de l'analyse de concert avec les coefficients de pondération pour tenir compte de la population cible des employés de 1999.

Nota : Le processus de repondération a permis de tenir compte des employés non-répondants, ce qui fait que toutes les analyses doivent s'appuyer sur les poids finaux des employés. On a exclu les employés échantillonnés qui sont décédés ou qui sont hors du champ de l'enquête (qui ne travaillaient pas à l'emplacement sélectionné en mars 1999).

Voici quelques exemples qui s'appuient sur le volet de l'analyse des employés de 1999.

**Exemple 1** : Le nombre total d'employés au sein de la population cible en 1999.

$$\hat{N} = \sum_i w_i = 10,777,543$$

$w_i$  – Poids final de l'employé

**Exemple 2** : Le salaire horaire moyen par employé au sein de la population cible des employés de 1999.

$$\hat{X} = \frac{\sum_i w_i x_i}{\sum_i w_i} = 18,53\$$$

$w_i$  – Poids final de l'employé

$x_i$  – Salaire horaire

**Exemple 3** : Le salaire horaire moyen par employé syndiqué ou protégé par une convention collective (CC) au sein de la population cible des employés de 1999.

$$\hat{X}_d = \frac{\sum_i w_i x_i \delta_i}{\sum_i w_i \delta_i} = 20,36\$$$

$w_i$  – Poids final de l'employé

$x_i$  – Salaire horaire

$\delta_i$  – Indicateur du statut syndical (1 si l'employé est membre d'un syndicat ou est protégé par une CC; 0, autrement)

## DONNÉES COUPLÉES SUR LE MILIEU DE TRAVAIL ET LES EMPLOYÉS 1999

### Population cible des données couplées

La population cible des données couplées de 1999 comprend en l'ensemble des emplacements de la population cible des milieux de travail de 1999 et les employés de la population cible des employés de 1999.

### Volet de l'analyse des données couplées (5,733 emplacements; 23,540 employés)

Ce volet englobe les milieux de travail du volet de l'analyse des milieux de travail de 1999 comptant au moins un employé répondant et des employés du volet de l'analyse des données sur les employés de 1999. On peut utiliser le volet de l'analyse de concert avec les coefficients de pondération pour tenir compte de la population cible des données couplées de 1999.

Nota : Quand on effectue une analyse des données sur un employé couplées aux caractéristiques du milieu de travail, on doit utiliser les poids finaux de l'employé de concert avec le fichier complet de l'employé. Quand on effectue une analyse des données sur le milieu de travail couplées aux caractéristiques de l'employé, on doit utiliser le poids des données couplées du milieu de travail en ne tenant compte que des milieux de travail comptant au moins un employé répondant. Il y a repondération dans le cas des milieux de travail où aucun employé n'a répondu au questionnaire.

**Exemple 1** : Le salaire horaire moyen par employé qui travaillait dans un emplacement à but non lucratif au sein de la population cible des données couplées de 1999.

$$\hat{X}_d = \frac{\sum_i w_i x_i \delta_i}{\sum_i w_i \delta_i} = 21,51\$$$

$w_i$  – Poids final de l'employé

$x_i$  – Salaire horaire

$\delta_i$  – Indicateur d'emplacement à but non lucratif (d'après le fichier de l'emplacement; 1 si l'emplacement est à but non lucratif; 0, autrement)

**Exemple 2** : Le salaire horaire moyen par employé syndiqué ou protégé par une convention collective qui travaillait dans un emplacement à but non lucratif au sein de la population cible des données couplées de 1999.

$$\hat{X}_d = \frac{\sum_i w_i x_i \delta_{1i} \delta_{2i}}{\sum_i w_i \delta_{1i} \delta_{2i}} = 22,05\$$$

$w_i$  – Poids final de l'employé

$x_i$  – Salaire horaire

$\delta_{1i}$  – Indicateur du statut syndical (1 si l'employé est membre d'un syndicat ou est protégé par une CC; 0, autrement)

$\delta_{2i}$  – Indicateur d'emplacement à but non lucratif (d'après le fichier de l'emplacement; 1 si l'emplacement à but non lucratif; 0, autrement)

**Exemple 3** : La paie brute moyenne par employé pour les emplacements comptant au moins un employé avec une invalidité à long terme au sein de la population cible des données couplées de 1999.

$$\hat{R}_d = \frac{\sum_i w_i z_i \delta_i}{\sum_i w_i x_i \delta_i} = 32,810\$$$

$w_i$  – Poids des données couplées sur l’emplacement

$x_i$  - Emploi

$z_i$  – Paie brute

$\delta_i$  – Indicateur d’invalidité à long terme (d’après le fichier de l’employé; 1 si l’emplacement compte au moins un employé avec une invalidité à long terme; 0, autrement)

**Exemple 4** : La paie brute moyenne par employé des emplacements qui offraient des avantages sociaux au sein de la population cible des données couplées de 1999 comptant au moins un employé avec une invalidité à long terme.

$$\hat{R}_d = \frac{\sum_i w_i z_i \delta_{1i} \delta_{2i}}{\sum_i w_i x_i \delta_{1i} \delta_{2i}} = 32,790\$$$

$w_i$  – Poids des données couplées sur l’emplacement

$x_i$  - Emploi

$z_i$  – Paie brute

$\delta_{1i}$  – Indicateur d’avantages sociaux (1 si l’emplacement offre des avantages sociaux; 0, autrement)

$\delta_{2i}$  – Indicateur d’invalidité à long terme (d’après le fichier de l’employé; 1 si l’emplacement compte au moins un employé avec invalidité à long terme; 0, autrement)

## MILIEU DE TRAVAIL 2000

### Population cible des milieux de travail

L'EMTE est une enquête longitudinale dont le volet sur les milieux de travail est rafraîchi tous les deux ans (2001, 2003, etc.). La population cible de 2000 demeure donc la même qu'en 1999.

### Volet de l'analyse des milieux de travail (6,068 emplacements)

Le volet de l'analyse 2000 est le sous-ensemble des milieux de travail du volet de l'analyse des milieux de travail de 1999 comptant au moins un employé rémunéré en mars 2000 qui a reçu le formulaire T4 Supplémentaire de l'Agence des douanes et du revenu du Canada. Sont considérés comme hors champ du volet de l'analyse des milieux de travail de 2000 les milieux de travail qui en mars 2000 :

étaient situés au Yukon, au Nunavut ou dans les Territoires du Nord-Ouest exploitaient des cultures agricoles ou faisaient l'élevage d'animaux; la pêche, la chasse et le piégeage; les ménages privés, les organisations religieuses et l'administration publique.

Ces exclusions ne s'appliquent qu'au volet de l'analyse de 2000 et non à la population cible.

Nota : Les poids finaux des milieux de travail doivent être utilisés dans les analyses vu qu'une repondération a été effectuée afin de tenir compte des non-répondants en 1999. Les analyses effectuées dans le cadre du volet d'analyse des milieux de travail de 2000 ne donnent pas une image transversale de l'ensemble des milieux de travail en mars 2000. Cela découle du fait que les milieux de travail qui ont vu le jour après la création de la base de 1999 ont aucune probabilité d'être inclus dans l'échantillon et qu'aucune repondération n'a été effectuée pour tenir compte de ceux-ci. Ainsi, toutes les analyses du volet de l'analyse des milieux de travail de 2000 doivent porter sur les unités «continuanes» (qui sont toujours en exploitation et qui font toujours partie du champ d'observation de l'enquête) de la population de 1999 seulement.

Voici quelques exemples qui s'appuient sur le volet de l'analyse des milieux de travail de 2000.

**Exemple 1 :** Le nombre total d'emplacements continuants au sein de la population cible des milieux de travail de 2000.

$$\hat{N} = \sum_i w_i = 668,188$$

$w_i$  – Poids final de l'emplacement

**Exemple 2 :** Le nombre total d'employés dans les emplacements continuants au sein de la population cible des milieux de travail de l'EMTE de 2000.

$$\hat{X} = \sum_i w_i x_i = 10,785,150$$

$w_i$  – Poids final de l'emplacement  
 $x_i$  - Emploi

**Exemple 3** : La paie brute moyenne par employé des emplacements continuants au sein de la population cible des milieux de travail de 2000.

$$\hat{R} = \frac{\sum_i w_i z_i}{\sum_i w_i x_i} = 32,166\$$$

$w_i$  – Poids final de l'emplacement

$x_i$  - Emploi

$z_i$  – Paie brute

**Exemple 4** : La paie brute moyenne par employé des emplacements continuants qui offrent des avantages sociaux au sein de la population cible des milieux de travail de 2000.

$$\hat{R}_d = \frac{\sum_i w_i z_i \delta_i}{\sum_i w_i x_i \delta_i} = 34,988\$$$

$w_i$  – Poids final de l'emplacement

$x_i$  - Emploi

$z_i$  - Paie brute

$\delta_i$  - Indicateur d'avantages sociaux (1 si l'emplacement offre des avantages sociaux; 0, autrement)

## EMPLOYÉ 2000

### Population cible des employés

L'EMTE est une enquête longitudinale dont le volet sur les employés est rafraîchi tous les deux ans (2001, 2003, etc.). Pour cette raison, la population cible des employés de 2000 demeure la même qu'en 1999.

### Volet de l'analyse des employés (20,167 employés)

Le volet de l'analyse de 2000 englobe le sous-ensemble des employés du volet de l'analyse des employés de 1999 dont l'employeur en mars 1999 est visé par l'analyse des milieux de travail de 2000. Cet ensemble d'employés est réparti selon les «continuant» (ceux qui ont le même employeur en mars 1999 et en mars 2000) et les sortants (ceux qui n'ont plus le même employeur qu'en mars 1999). Les sortants ont un nouvel employeur qui peut ou non faire partie de la population cible des emplacements de 2000 ou ne sont plus sur le marché du travail.

Sont exclus du volet de l'analyse des employés de 2000, les employés qui relèvent d'emplacements exclus du volet de l'analyse des milieux de travail de 2000. Ces exclusions ne s'appliquent qu'au volet de l'analyse de 2000 et non à la population cible.

Nota : Les poids finaux des employés doivent être utilisés dans les analyses vu qu'une repondération a été effectuée afin de tenir compte des non-répondants en 1999 et en 2000. Les analyses effectuées dans le cadre du volet d'analyse des employés de 2000 n'englobent pas tous les employés en mars 2000. Cela découle du fait que les employés rattachés à des milieux de travail qui ont vu le jour après la création de la base de 1999 n'ont aucune probabilité d'être inclus dans l'échantillon et qu'aucune repondération n'a été effectuée pour tenir compte de ceux-ci. Ainsi, toutes les analyses du volet d'analyse des employés de 2000 doivent viser les unités continuantes ou les unités sortantes de la population de 1999 seulement.

Voici quelques exemples qui s'appuient sur le volet de l'analyse des employés de 2000.

**Exemple 1 :** Le nombre total d'employés continuants ou sortants en mars 2000 qui travaillaient en mars 1999 dans un emplacement continuant (p. ex. un employé qui en mars 1999 relevait d'un milieu de travail faisant partie du volet de l'analyse de 2000).

$$\hat{N} = \sum_i w_i = 10,755,029$$

$w_i$  – Poids final de l'employé

**Exemple 2 :** Le nombre total d'employés continuants en mars 2000 qui avaient le même milieu de travail en mars 1999 et en mars 2000.

$$\hat{N}_d = \sum_i w_i \delta_i = 8,964,798$$

$w_i$  – Poids final de l'employé

$\delta_i$  – Indicateur du statut de continuant (1 si l'employé travaillait pour le même employeur en mars 2000 et en mars 1999; 0, autrement)

**Exemple 3:** Le nombre total d'employés sortants entre avril 1999 et mars 2000 qui travaillaient en mars 1999 dans un emplacement continuant.

$$\hat{N}_d = \sum_i w_i \delta_i = 1,790,230$$

$w_i$  – Poids final de l'employé

$\delta_i$  – Indicateur du statut d'employé sortant (1 si l'employé n'avait plus en mars 2000 le même employeur qu'en mars 1999; 0, autrement)

**Exemple 4:** Le salaire horaire moyen par employé en mars 2000 qui travaillaient en mars 1999 dans un emplacement continuant.

$$\hat{X}_d = \frac{\sum_i w_i x_i \delta_i}{\sum_i w_i \delta_i} = 18,11\$$$

$w_i$  – Poids final de l'employé

$x_i$  – Salaire horaire

$\delta_i$  – Indicateur du statut d'employé qui travaille (1 si l'employé travaille)

## ANALYSE DES DONNÉES COUPLÉES SUR LES MILIEUX DE TRAVAIL ET LES EMPLOYÉS 2000

### Population cible des données couplées

La population cible des données couplées de 2000 regroupe l'ensemble des emplacements de la population cible des milieux de travail de 2000 et des employés de la population cible des employés de 2000.

### Volet de l'analyse de données couplées (5,453 emplacements; 20,167 employés)

Ce volet englobe les milieux de travail du volet de l'analyse des milieux de travail de 2000 comptant au moins un employé répondant et les employés du volet de l'analyse des données sur les employés de 2000. On peut utiliser le volet de l'analyse de concert avec les coefficients de pondération pour tenir compte de la population cible des données couplées de 1999.

Nota : Quand on effectue une analyse des données sur un employé couplées aux caractéristiques du milieu de travail, on doit utiliser les poids finaux de l'employé de concert avec le fichier complet de l'employé. Quand on effectue une analyse des données sur le milieu de travail couplées aux caractéristiques de l'employé, on doit utiliser le poids des données couplées du milieu de travail en ne tenant compte que des milieux de travail comptant au moins un employé répondant. Il y a repondération dans le cas des milieux de travail où aucun employé n'a répondu au questionnaire. Les analyses effectuées dans le cadre du volet d'analyse des données couplées de 2000 ne donnent pas une image transversale de l'ensemble des données couplées sur les milieux de travail et les employés en mars 2000. Cela découle du fait que les milieux de travail et les employés relevant de milieux de travail qui ont vu le jour après la création de la base de 1999 n'ont aucune probabilité d'être inclus dans l'échantillon et qu'aucune repondération n'a été effectuée pour tenir compte de ceux-ci. Ainsi, toutes les analyses du volet de l'analyse des données couplées de 2000 doivent porter sur les employés continuants ou sortants des emplacements continuants.

Voici quelques exemples qui s'appuient sur le volet de l'analyse des données couplées de 2000.

**Exemple 1** : Le salaire horaire moyen par employé qui en mars 1999 travaillait dans un emplacement continuant qui, au cours de la période de collecte des données de 2000, était un emplacement à but non lucratif. L'employé peut ou non toujours travailler pour le même employeur qu'en mars 1999.

$$\hat{X}_d = \frac{\sum_i w_i x_i \delta_i}{\sum_i w_i \delta_i} = 20,90\$$$

$w_i$  – Poids final de l'employé

$x_i$  – Salaire horaire

$\delta_i$  – Indicateur d'emplacement à but non lucratif (d'après le fichier de l'emplacement; 1 si l'emplacement est à but non lucratif; 0, autrement)

**Exemple 2** : Le salaire horaire moyen par employé qui travaille dans un emplacement à but non lucratif visé par la population cible couplée et qui travaillait au même endroit en mars 1999.

$$\hat{X}_d = \frac{\sum_i w_i x_i \delta_{1i} \delta_{2i}}{\sum_i w_i \delta_{1i} \delta_{2i}} = 22,79\$$$

$w_i$  – Poids final de l'employé

$x_i$  – Salaire horaire

$\delta_{1i}$  – Indicateur du statut d'employé continuant (1 si l'employé a le même employeur en mars 2000 qu'en mars 1999; 0, autrement)

$\delta_{2i}$  – Indicateur d'emplacement à but non lucratif (d'après le fichier de l'emplacement; 1 si l'emplacement est à but non lucratif; 0, autrement)

**Exemple 3** : La paie brute moyenne par employé des emplacements continuants qui comptaient au moins un employé continuant ou sortant avec une invalidité à long terme en mars 2000.

$$\hat{R}_d = \frac{\sum_i w_i z_i \delta_i}{\sum_i w_i x_i \delta_i} = 34,619\$$$

$w_i$  – Poids des données couplées sur l’emplacement

$x_i$  - Emploi

$z_i$  - Paie brute

$\delta_i$  – Indicateur d’invalidité à long terme (d’après le fichier de l’employé; 1 si l’emplacement compte au moins un employé avec une invalidité à long terme; 0, autrement)

**Exemple 4** : La paie brute moyenne par employé des emplacements continuants qui comptaient au moins un employé sortant avec une invalidité à long terme en mars 2000.

$$\hat{R}_d = \frac{\sum_i w_i z_i \delta_i}{\sum_i w_i x_i \delta_i} = 34,264\$$$

$w_i$  – Poids des données couplées sur l’emplacement

$x_i$  - Emploi

$z_i$  - Paie brute

$\delta_i$  – Indicateur d’employé sortant avec invalidité à long terme (d’après le fichier de l’employé; 1 si l’emplacement compte au moins un employé sortant avec une invalidité à long terme; 0, autrement)

## DONNÉES LONGITUDINALES SUR LES MILIEUX DE TRAVAIL 1999/2000

Population cible des données longitudinales sur les milieux de travail

Il s'agit de la même population que la population cible des milieux de travail de 2000.

Volet de l'analyse longitudinale des milieux de travail (6,068 emplacements)

Il s'agit du même volet que le volet de l'analyse des milieux de travail de 2000, y compris les données de 1999 et de 2000.

Nota : Les estimations longitudinales calculées d'après les données de 1999 dans les exemples suivants ne portent que sur les emplacements continuants.

Voici quelques exemples qui s'appuient sur le volet de l'analyse longitudinale des milieux de travail.

**Exemple 1 :** La variation en pourcentage du revenu total de 1999 à 2000 qui s'applique aux emplacements continuants.

$$\hat{P} = \frac{\sum_i w_i x_{i2000} - \sum_i w_i x_{i1999}}{\sum_i w_i x_{i1999}} \times 100 = 6,95\%$$

$w_i$  – Poids final de l'emplacement

$x_{i2000}$  – Revenu de 2000

$x_{i1999}$  – Revenu de 1999

**Exemple 2 :** La variation en pourcentage de la paie brute moyenne par employé de 1999 à 2000 qui s'applique aux emplacements continuants.

$$\hat{P} = \frac{\frac{\sum_i w_i z_{i2000}}{\sum_i w_i x_{i2000}} - \frac{\sum_i w_i z_{i1999}}{\sum_i w_i x_{i1999}}}{\frac{\sum_i w_i z_{i1999}}{\sum_i w_i x_{i1999}}} \times 100 = 3,23\%$$

$w_i$  – Poids final de l'emplacement

$x_{i2000}$  – Emploi de 2000

$x_{i1999}$  – Emploi de 1999

$z_{i2000}$  - Paie brute de 2000

$z_{i1999}$  - Paie brute de 1999

**Exemple 3 :** La variation en pourcentage du revenu total qui s'applique aux emplacements qui offraient des avantages sociaux les deux années pour les emplacements continuants.

$$\hat{P} = \frac{\sum_i w_i x_{i2000} \delta_i - \sum_i w_i x_{i1999} \delta_i}{\sum_i w_i x_{i1999} \delta_i} \times 100 = 6,21\%$$

$w_i$  – Poids final de l'emplacement

$x_{i2000}$  – Revenu de 2000

$x_{i1999}$  – Revenu de 1999

$\delta_i$  - Indicateur d'avantages sociaux (1 si l'emplacement offrait des avantages sociaux au cours des deux années d'enquête, 1999 et 2000; 0, autrement)

## DONNÉES LONGITUDINALES SUR LES EMPLOYÉS 1999-2000

### Population cible des données longitudinales sur les employés

Il s'agit de la même population cible que la population cible des employés de 2000.

### Volet de l'analyse longitudinale des employés (20,167 employés)

Il s'agit du même volet que le volet de l'analyse des employés de 2000, y compris les données de 1999 et de 2000.

Nota : On doit utiliser les poids finaux des employés de 2000 dans le cadre des analyses longitudinales. Les estimations longitudinales calculées d'après les données de 1999 dans les exemples suivants ne portent que sur les employés qui en mars 1999 étaient rattachés aux emplacements continnants.

Voici quelques exemples qui s'appuient sur le volet de l'analyse longitudinale des employés.

**Exemple 1** : La variation, entre 1999 et 2000, en pourcentage du salaire horaire moyen par employé qui travaillait en mars 1999 dans un emplacement continuant. (L'employé peut travailler au même emplacement qu'en mars 1999, travailler dans un nouvel emplacement ou ne pas travailler.)

$$\hat{P} = \frac{\frac{\sum_i w_i x_{i2000}}{\sum_i w_i} - \frac{\sum_i w_i x_{i1999}}{\sum_i w_i}}{\frac{\sum_i w_i x_{i1999}}{\sum_i w_i}} \times 100 = -2,64\%$$

$w_i$  – Poids final de l'employé de 2000

$x_{i2000}$  – Salaire horaire de 2000

$x_{i1999}$  – Salaire horaire de 1999

**Exemple 2** : La variation en pourcentage du salaire horaire moyen par employé continuant entre 1999 et 2000 qui travaillait en mars 1999 dans un emplacement continuant.

$$\hat{P} = \frac{\frac{\sum_i w_i x_{i2000} \delta_i}{\sum_i w_i \delta_i} - \frac{\sum_i w_i x_{i1999} \delta_i}{\sum_i w_i \delta_i}}{\frac{\sum_i w_i x_{i1999} \delta_i}{\sum_i w_i \delta_i}} \times 100 = 3,40\%$$

$w_i$  – Poids final de l'emplacement

$x_{i2000}$  – Salaire horaire de 2000

$x_{i1999}$  – Salaire horaire de 1999

$\delta_i$  – Indicateur du statut d'employé continuant (1 si l'employé a le même employeur en mars 2000 qu'en mars 1999; 0, autrement)

**Exemple 3** : La variation en pourcentage du salaire horaire moyen par employé sortant entre 1999 et 2000 qui travaillait en mars 1999 dans un emplacement continuant et qui a un nouvel employeur en mars 2000.

$$\hat{P} = \frac{\frac{\sum_i w_i x_{i2000} \delta_i}{\sum_i w_i \delta_i} - \frac{\sum_i w_i x_{i1999} \delta_i}{\sum_i w_i \delta_i}}{\frac{\sum_i w_i x_{i1999} \delta_i}{\sum_i w_i \delta_i}} \times 100 = 4,74\%$$

$w_i$  – Poids final de l'emplacement

$x_{i2000}$  – Salaire horaire de 2000

$x_{i1999}$  – Salaire horaire de 1999

$\delta_i$  – Indicateur du statut d'employé sortant (1 si l'employé en mars 2000 n'avait plus le même employeur qu'en mars 1999; 0, autrement)

## DONNÉES LONGITUDINALES COUPLÉES DU MILIEU DE TRAVAIL ET DE L'EMPLOYÉ 1999-2000

### Population cible des données longitudinales couplées

Il s'agit de la même population que la population cible des données couplées de 2000.

### Volet de l'analyse des données longitudinales couplées (5,453 emplacements; 20,167 employés)

Il s'agit du même volet que le volet de l'analyse des données couplées, y compris les données de 1999 et de 2000.

Nota : Lorsqu'on effectue une analyse des données longitudinales sur un employé couplées aux caractéristiques du milieu de travail, on doit utiliser les poids finaux de l'employé de 2000 de concert avec le fichier complet de l'employé. Quand on effectue une analyse des données longitudinales sur le milieu de travail couplées aux caractéristiques de l'employé, on doit utiliser les poids des données couplées du milieu de travail de 2000 en ne tenant compte que des milieux de travail comptant au moins un employé répondant. Il y a repondération dans le cas des milieux de travail où aucun employé n'a répondu au questionnaire. Les estimations longitudinales calculées d'après les données de 1999 dans les exemples suivants ne portent que sur les employés qui en mars 1999 relevaient d'emplacements continuants indépendamment de leur lieu de travail ou de leur retrait du marché du travail en mars 2000.

Voici quelques exemples qui s'appuient sur le volet de l'analyse des données longitudinales couplées.

**Exemple 1 :** La variation en pourcentage du salaire horaire moyen par employé qui en mars 1999 travaillait dans un emplacement continuant à but non lucratif. (L'employé peut relever du même emplacement qu'en mars 1999, être rattaché à un nouvel emplacement ou ne pas travailler du tout.)

$$\hat{P} = \frac{\frac{\sum_i w_i x_{i2000} \delta_i}{\sum_i w_i \delta_i} - \frac{\sum_i w_i x_{i1999} \delta_i}{\sum_i w_i \delta_i}}{\frac{\sum_i w_i x_{i1999} \delta_i}{\sum_i w_i \delta_i}} \times 100 = -2,45\%$$

$w_i$  – Poids final de l'employé

$x_{i2000}$  – Salaire horaire de 2000

$x_{i1999}$  – Salaire horaire de 1999

$\delta_i$  – Indicateur d'emplacement à but non lucratif (d'après le fichier de l'emplacement; 1 si l'emplacement était à but non lucratif en 1999; 0, autrement)

**Exemple 2 :** La variation en pourcentage du salaire horaire moyen par employé continuant qui en mars 1999 travaillait dans un emplacement continuant. Il s'agissait d'un emplacement à but non lucratif en 1999 et en 2000.

$$\hat{P} = \frac{\frac{\sum_i w_i x_{i2000} \delta_{1i} \delta_{2i}}{\sum_i w_i \delta_{1i} \delta_{2i}} - \frac{\sum_i w_i x_{i1999} \delta_{1i} \delta_{2i}}{\sum_i w_i \delta_{1i} \delta_{2i}}}{\frac{\sum_i w_i x_{i1999} \delta_{1i} \delta_{2i}}{\sum_i w_i \delta_{1i} \delta_{2i}}} \times 100 = 4,87\%$$

$w_i$  – Poids final de l'employé

$x_{i2000}$  – Salaire horaire de 2000

$x_{i1999}$  – Salaire horaire de 1999

$\delta_{1i}$  – Indicateur du statut d'employé continuant (1 si l'employé avait le même employeur en mars 2000 qu'en mars 1999; 0, autrement)

$\delta_{2i}$  – Indicateur d'emplacement à but non lucratif (d'après le fichier de l'emplacement; 1 si l'emplacement était à but non lucratif en 1999 et en 2000; 0, autrement)

**Exemple 3** : La variation en pourcentage du revenu total de 1999 à 2000 qui s'applique aux emplacements continnants qui comptaient au moins un employé continuant ou sortant avec une invalidité à long terme en mars 1999 et en mars 2000 au sein de la population cible des données longitudinales couplées.

$$\hat{P} = \frac{\sum_i w_i x_{i2000} \delta_i - \sum_i w_i x_{i1999} \delta_i}{\sum_i w_i x_{i1999} \delta_i} \times 100 = 2,93\%$$

$w_i$  – Poids final des données couplées sur l'emplacement

$x_{i2000}$  – Revenu de 2000

$x_{i1999}$  – Revenu de 1999

$\delta_i$  – Indicateur d'invalidité à long terme (1 si l'employé avait une invalidité à long terme en 1999 et en 2000; 0, autrement)

## ANNEXE 5

### Analyses couplées

#### Raisons pour lesquelles il faut traiter différemment les modèles couplés

Avec des données couplées sur les employeurs et les employés comme celles de l'EMTE, menée par SC, les chercheurs ont l'occasion d'étudier des résultats sur les marchés des affaires et du travail qui dépendent en partie des relations entre employeurs et employés. Ils doivent en même temps affronter certains problèmes statistiques et économétriques à l'intérieur de leurs travaux de modélisation des activités sur ces marchés.

Depuis la fin des années 90, des économistes proposent divers modèles empiriques qui peuvent être estimés à l'aide de données couplées (appariées) sur les employeurs et les employés.<sup>2</sup> Les modèles employés pour les études en question sont essentiellement la fonction de régression linéaire bien connue, mais renferment un certain nombre de nouveaux éléments justifiant un traitement différent de celui de l'analyse de régression linéaire classique. Examinons un modèle linéaire spécifié pour un résultat au niveau de l'employé  $Y_{ij}$  dans lequel  $X_{ij}$  caractérise l'employé  $i$  et  $Z_j$ , l'établissement  $j$  :

$$\begin{aligned}Y_{ij} &= \alpha_j + \beta_j X_{ij} + \varepsilon_{ij}, \\ \alpha_j &= \alpha_0 + \alpha_1 Z_j + u_j, \\ \beta_j &= \beta_0 + \beta_1 Z_j + v_j,\end{aligned}$$

où  $\varepsilon_{ij}$ ,  $u_j$  et  $v_j$  sont des perturbations classiques,  $\varepsilon_{ij}$  dépend de  $X_{ij}$  et  $u_j$  et  $v_j$  sont indépendants l'un de l'autre et indépendants également de  $Z_j$ . On peut calculer à partir de ces spécifications un modèle linéaire :

$$Y_{ij} = \alpha_0 + \alpha_1 Z_j + \beta_0 X_{ij} + \beta_1 X_{ij} Z_j + u_j + v_j X_{ij} + \varepsilon_{ij}.$$

Les modèles comme celui figurant ci-dessus, qu'on appelle souvent des *modèles mixtes* (*des modèles à paramètres variables*), renferment des éléments stochastiques ( $u_j$  et  $v_j$ ) que l'analyste ne peut observer. L'analyse de régression linéaire classique ne s'applique au modèle figurant ci-dessus que si  $u_j = v_j = 0$ . Quand  $v_j = 0$ , cela devient un exemple des *modèles de composantes d'erreur* et quand  $u_j = 0$  nous obtenons un exemple des modèles de coefficients aléatoires.

Le modèle mixte devient plus complexe si nous tentons d'analyser des résultats des relations entre des employeurs et des employés au fil du temps. Même en l'absence de composante d'erreur et de coefficient aléatoire, il est fort probable qu'on enfreindra dans un modèle mixte certaines des hypothèses courantes de l'analyse de régression classique. En particulier, la corrélation à l'intérieur d'une entreprise, l'hétéroscédasticité entre les entreprises et l'erreur de mesure qu'entraîne l'agrégation peuvent toutes avoir de graves conséquences si l'on ne s'attaque pas minutieusement à ces problèmes. En outre, l'ensemble du modèle, capable de saisir les effets des caractéristiques des employeurs et des employés et ceux des décisions (choix) que les uns et les autres font, n'est pas nécessairement hiérarchique ni équilibré.<sup>3</sup> Tous les traitements définis dans les documents sur la modélisation multiniveau<sup>4</sup> ne peuvent

<sup>2</sup> Voir Abowd et Kramarz (1999) pour un examen de la question. L'ouvrage de Haltiwanger et al. (1999) renferme une sélection d'articles présentés au symposium international de 1998 consacré aux données couplées sur les employeurs et les employés.

<sup>3</sup> Le modèle linéaire de base employé par Abowd et Kramarz (1999) pour leur examen en est un exemple.

<sup>4</sup> Voir Goldstein (1995) pour une introduction à l'analyse multiniveau.

donc s'appliquer à une telle spécification.

### **Utilisation de variables d'employeur à l'intérieur d'analyses sur des employés**

Lorsqu'on tente d'analyser des résultats au niveau des employés à l'aide de variables au niveau d'un employeur, il s'amorce une désagrégation des variables d'employeur. Des employés sélectionnés à partir de la même entreprise ou du même établissement devraient avoir des variables d'employeur identiques, comme les investissements dans la technologie, les dépenses de formation et l'industrie; ces variables d'employeur risquent également de ne pas être indépendantes d'un travailleur à un autre à l'intérieur du même établissement. Cependant l'estimation paramétrique traite nécessairement la valeur d'une variable d'employeur liée à chacun des employés du même établissement comme de l'information indépendante. Certaines estimations peuvent, par conséquent, être faussement différentes de 0. On doit corriger le biais par défaut dans les erreurs-types estimées afin d'éviter pareille situation. Moulton (1985) et Troske (1996) traitent de la procédure de correction.

On peut respecter l'analyse de régression classique pour supposer que la population d'employés d'une entreprise est homogène, mais il est probable qu'elle soit hétérogène d'une entreprise à une autre. On risque de faire une fausse inférence en traitant des données groupées tirées d'une population hétérogène comme si elles l'étaient d'une population homogène. Le problème de l'hétéroscédasticité au niveau d'un groupe n'est cependant pas nouveau. Il est question de traitements dans beaucoup de recueils de textes économétriques courants<sup>5</sup>. Une spécification d'un modèle de coefficient aléatoire, de Hildreth et de Houck (1968), pourrait s'avérer être une façon pratique.

### **Utilisation de données résumées sur des employés à l'intérieur d'analyses sur des employeurs**

L'information recueillie auprès de employés pourrait être particulièrement intéressante pour des chercheurs modélisant des résultats d'employeurs. Certaines variables définies au niveau des employés pourraient cependant s'avérer problématiques au moment de leur utilisation au niveau d'employeurs, en particulier celles qui seraient fondées sur les évaluations subjectives formulées par les employés échantillonnés. Le problème de *l'erreur à l'intérieur des variables* qu'entraîne l'agrégation devient donc dans les analyses couplées la norme plutôt que l'exception.

La solution à l'erreur de mesure consiste à remplacer la variable en question par une variable instrumentale (VI), pour laquelle la corrélation avec la vraie valeur de la variable sous-jacente, mais non avec l'erreur de mesure, est très élevée.<sup>6</sup> Les estimateurs de VI sont asymptotiquement cohérents, efficaces et normaux dans certaines conditions générales. Pour la méthode des VI, Fuller (1987) est une excellente référence. Il n'est pas facile dans bien des cas de trouver un instrument approprié, mais les données couplées permettent plus facilement aux analystes de découvrir de bons outils. La correction des problèmes provoqués par l'erreur de mesure n'est toutefois pas la seule utilité de la méthode des VI. Chose plus importante, on emploie cette méthode dans le cadre de beaucoup d'études empiriques pour régler le problème possible de l'endogénéité : le fait qu'à l'intérieur d'un modèle une variable explicative dépende également de la variable dépendante. Dans l'analyse de régression classique, c'est le cas lorsqu'il y a corrélation entre la variable explicative et le terme d'erreur. Le problème de l'endogénéité rend la méthode des VI plus populaire que le modèle multiniveau dans le cadre d'une analyse de données

---

<sup>5</sup> Voir par exemple, les chapitres 16 et 17 de l'ouvrage de Judge et d'autres (1982).

<sup>6</sup> Il est possible que l'erreur de mesure ne soit pas classique, c'est-à-dire indépendante des vraies valeurs de la variable en question. Voir Barron, Berger et Black (1999).

couplées sur les employeurs et les employés.

## **Logiciels**

Bien des programmes statistiques/économétriques permettent de traiter l'estimation du modèle mixte, la méthode des VI et les estimations des effets fixes et aléatoires. Le SAS et le STATA sont deux puissants logiciels. Dans le SAS, on peut utiliser GLM et MIXED PROC pour une estimation du modèle multiniveau, en incluant dans les procédures les pondérations. Le STATA n'offre cependant pas que des possibilités d'estimer nombre de modèles que des chercheurs peuvent spécifier, mais fournit également un certain nombre de procédures qui rendent compte d'effets complexes du plan d'échantillonnage (comme il y a lieu de le faire pour l'EMTE; voir l'annexe ?). Les procédures du STATA et leur préfixe « svy » sont donc particulièrement appropriés à l'analyse des données de l'EMTE. L'équipe du projet d'EMTE mettra également à l'essai cette année (en 2001) un certain nombre d'autres logiciels appropriés aux modèles mixtes.

## ANNEXE 6

### Pondération et estimation

L'échantillon de l'Enquête sur le milieu de travail et les employés est un échantillon d'emplacements canadiens à partir duquel on sélectionne un certain nombre d'employés suivant la taille de l'emplacement mesurée par l'emploi total.

#### **Pondération**

Le fait de sélectionner un échantillon à partir de la population à laquelle on s'intéresse signifie qu'on ne recueillera des variables d'analyse que pour une fraction des unités disponibles. Afin d'être en mesure de produire des estimations qui se rattacheront à la population, on attribue à chacune des unités échantillonnées une pondération pour qu'elle représente d'autres unités similaires qui n'ont pas été sélectionnées. Cette pondération, appelée *pondération de plan*, est égale à l'inverse de la probabilité de sélection. Si, par exemple, on sélectionne deux unités sur dix, on donne alors à chaque unité une pondération de 5.

On sélectionne de façon indépendante l'échantillon d'employeurs de l'EMTE à l'intérieur de 252 strates sans effectuer de remplacement (on ne remplace pas une unité dans une strate une fois qu'elle en a été tirée). Les strates représentent des groupes homogènes d'unités identifiées par industrie (14 classes), région (6 classes) et taille (3 classes). Pour une taille donnée de l'échantillon global admissible, on calcule le nombre d'unités sélectionnées à l'intérieur de chaque strate de façon à ce qu'aucune cellule ne dépasse un coefficient de variation préspecifié, ce qui signifie que le taux d'échantillonnage des strates hautement variables est plus élevé et vice-versa.

L'échantillon initial détermine la pondération de plan de chaque unité. Pendant tout le processus d'enquête, les pondérations de plan initiales peuvent subir plusieurs ajustements, qui visent à protéger la représentativité de l'échantillon. Pour l'EMTE, on a effectué deux ajustements : un afin de compenser les non-réponses complètes et l'autre afin de réduire l'influence sur les estimations des sauteuses de strate (de grandes unités qu'on croit petites et vice-versa). Pour remédier aux non-réponses, on multiplie les pondérations des unités ayant répondu par un ratio entre toutes les unités échantillonnées et toutes les unités ayant répondu à l'intérieur de chaque strate. Cette méthode est fondée sur l'hypothèse selon laquelle les répondants et les non-répondants se comportent de la même façon. Cette hypothèse n'est pas déraisonnable, puisque les non-réponses sont principalement le fait des petites unités.

Apporter un ajustement pour tenir compte des sauteuses de strate est plus complexe, parce qu'il y a au moins trois méthodes pour régler le problème. On peut réduire la pondération de plan de la sauteuse de strate et répartir la différence entre les unités restantes à l'intérieur de la strate, ou on peut réduire ses valeurs, ou encore supprimer entièrement l'unité et la traiter comme un cas de non-réponse. Nous avons choisi la première méthode et avons ciblé environ 30 employeurs en vue d'un ajustement des pondérations de plan.

L'utilisation des pondérations de plan, initiales ou finales, entraîne une estimation non biaisée, quoique inefficace. Pour accroître l'efficacité du processus d'estimation, on peut étalonner, ou *calibrer*, l'échantillon suivant un ensemble de chiffres (totaux) pour une population connue ou estimée efficacement. Dans le cadre de l'EMTE, cela se fait à l'aide de l'emploi total estimé grâce à l'EERH (l'Enquête sur l'emploi, la rémunération et les heures) au niveau de l'industrie par région, où les estimations découlant de l'Enquête sur le milieu de travail et les employés doivent correspondre aux

estimations découlant de l'EERH. On applique les facteurs d'ajustement qui en résultent aux pondérations de plan finales. L'étalonnage est des plus avantageux dans les cas où la corrélation entre la variable de calibrage et les variables auxquelles on s'intéresse est très élevée.

Le produit de la pondération de plan finale et du facteur de calibrage est la pondération finale de l'unité. On s'en sert pour calculer des statistiques du premier ordre comme des totaux, des moyennes, des coefficients de régression, etc. Pour calculer des statistiques du 2<sup>e</sup> ordre, ou des variances, on doit employer des logiciels qui permettent à leur utilisateur de spécifier le plan d'enquête. Si on utilise des produits comme SAS sans transformer de façon appropriée les pondérations de l'enquête, la sous-estimation en découlant de la variance peut être assez marquée.

## **Estimation**

La deuxième possibilité est de loin la plus générale et la plus facile à mettre à exécution. Elle englobe l'utilisation de pondérations *bootstrap*. La méthode *bootstrap* est une technique statistique au moyen de laquelle on suit une procédure de rééchantillonnage pour produire un certain nombre d'ensembles de pondérations qui, si on les utilise correctement, saisissent la variabilité de bien des statistiques. L'idée consiste à calculer un grand nombre d'estimations *bootstrap*, puis leur variance.

Une fois les pondérations *bootstrap* calculées, on peut les spécifier dans l'énoncé des pondérations à l'intérieur de toute procédure du SAS qui en comporte un. On doit produire une estimation reposant sur chaque ensemble de pondérations *bootstrap* afin de calculer la variance pour une variable désirée. On calcule ensuite la variance des estimations ainsi produites et on y applique une correction pour rendre la variance conforme au plan. Veuillez vous référer à l'annexe 7 pour des explications de la méthode *bootstrap*.

## ANNEXE 7

### Calcul de la variance

L'utilisation de pondérations *bootstrap* pour le calcul de variances conformes au plan

Il faut tenir compte du plan lorsqu'on calcule les variances d'estimations fondées sur des échantillons prélevés à partir de populations finies, ce qui ne se fait pas facilement à l'intérieur de la plupart des progiciels d'analyse statistique. Même si la majorité d'entre eux permettent l'utilisation de pondérations, ils ne les emploient pas de façon appropriée, ce qui entraîne la sous-estimation de la variance. Cela pourrait avoir des conséquences désastreuses sur les vérifications d'hypothèse et sur la construction d'intervalles de confiance.

Au fil des ans, les bureaux de la statistique ont mis au point des systèmes pour traiter les populations finies; la plupart de ces systèmes n'ont cependant pas la flexibilité nécessaire pour effectuer l'analyse des données. C'est là où intervient la technique *BOOTSTRAP*, qui est fondée sur le rééchantillonnage. Cette technique consiste à utiliser l'échantillon initial, à partir duquel on sélectionne un échantillon aléatoire simple, et on remplace autant d'unités qu'on en avait au départ. On répète cette procédure un grand nombre de fois pour garantir la cohérence.

Une fois les pondérations *bootstrap* calculées, on peut les préciser à l'intérieur de l'énoncé des pondérations dans le cadre de toute procédure SAS qui en comporte un. Il faut produire une estimation fondée sur chaque ensemble de pondérations *bootstrap* pour calculer la variance d'une variable donnée. On calcule ensuite la variance des estimations ainsi produites et on applique un ajustement afin de rendre la variance conforme au plan. Ci-dessous figurent deux exemples de la façon dont on peut procéder pour des totaux et pour des coefficients de corrélation. (Nota : le coefficient de corrélation est une statistique complexe et il n'existe pas de méthode permettant de calculer exactement sa variance. Les meilleures méthodes permettent d'atteindre une probabilité de couverture de 85 à 90 % pour le niveau nominal de couverture de 95 %.)

Le code qui suit suppose que les pondérations *bootstrap* ont déjà été insérées à l'intérieur de l'ensemble de données d'analyse. Le fichier des pondérations s'appelle *boot\_wt.sd2* et renferme le LOCNO et les BSW1 à BSW100, les pondérations *bootstrap*.

```
PROC SUMMARY DATA = WES NWAY;
  CLASS DOM_IND;
  VAR WKP_FINAL_WT WKP_BSW1-WKP_BSW100;
  WEIGHT TTL_EMP;
  OUTPUT OUT = ESTIM (DROP = _FREQ_ _TYPE_)
          SUM = EMPL WKP_BSW1-WKP_BSW100;
RUN;

PROC TRANSPOSE DATA = ESTIM
  OUT = T_ESTIM (DROP = _NAME_ RENAME = (COL1 = ESTIM));
  VAR WKP_BSW1-WKP_BSW100;
  BY DOM_IND;
RUN;

PROC SUMMARY DATA = T_ESTIM NWAY;
  CLASS DOM_IND;
  VAR ESTIM;
  OUTPUT OUT = VAR (DROP = _FREQ_ _TYPE_)
          CSS = VAR;
RUN;

DATA ESTIM;
  MERGE ESTIM (KEEP = DOM_IND EMPL)
```

```

VAR;
BY DOM_IND;
CV = ROUND (SQRT(50 / 100 * VAR) / EMPL, 0.01);
RUN;

```

La première procédure SOMMAIRE fait appel à un truc qui permet de calculer toutes les estimations nécessaires en une seule étape simple, ce qui peut se faire lorsqu'on produit des estimations pour une variable unique. Le truc en question consiste à préciser que les pondérations *bootstrap* sont les variables d'analyse et à utiliser la variable d'analyse comme s'il s'agissait de la pondération. On calcule les estimations au niveau de l'industrie du domaine précisé par l'énoncé des classes.

Une fois les estimations calculées, transposées et rebaptisées, on utilise une autre procédure SOMMAIRE pour calculer leur variance (leur somme corrigée des carrés, ou leur SCC dans le SAS, en fait). Finalement, la multiplication de la CSS par 50/100 produit la bonne variance fondée sur le plan. Le dénominateur (100) est l'ajustement normal  $n$  qui donne la variance classique. Le numérateur (50) reflète le fait que la moyenne de chaque ensemble de pondération *bootstrap* a été calculée à partir de 50 itérations, ce qui fournit une pondération *bootstrap* moyenne. La correction permet donc de réinjecter la variabilité que l'on avait perdue en utilisant la moyenne.

L'exemple qui suit illustre l'utilisation de pondérations *bootstrap* pour le calcul de coefficients de corrélation. On doit ici utiliser une macro pour calculer chacun des coefficients, étant donné qu'il est impossible d'employer facilement le truc fourni ci-dessus.

```

%MACRO COR_COEF;
  %DO I = 1 %TO 100;
    PROC CORR DATA = BOOT OUTP = CORRS NOPRINT;
      VAR TTL_EMP CBA_EMP;
      BY DOM_IND;
      WEIGHT WKP_BSW&I;
    RUN;

    DATA CORRS (KEEP = DOM_IND CBA_EMP RENAME = (CBA_EMP = CORR));
      SET CORRS (WHERE = (_TYPE_ = 'CORR' & _NAME_ = 'TTL_EMP'));
    RUN;

    PROC DATASETS FORCE NOLIST;
      APPEND BASE = ESTIM DATA = CORRS;
      QUIT;
    RUN;
  %END;
%MEND;

%COR_COEF;

PROC SUMMARY DATA = ESTIM NWAY;
  CLASS DOM_IND;
  VAR CORR;
  OUTPUT OUT = VAR (DROP = _FREQ_ _TYPE_)
  CSS = VAR;
RUN;

PROC CORR DATA = BOOT OUTP = CORRS NOPRINT;
  VAR TTL_EMP CBA_EMP;
  BY DOM_IND;
  WEIGHT WKP_FINAL_WT;
RUN;

DATA CORRS (KEEP = DOM_IND CBA_EMP RENAME = (CBA_EMP = EST_CORR));
  SET CORRS (WHERE = (_TYPE_ = 'CORR' & _NAME_ = 'TTL_EMP'));
RUN;

DATA ESTIM;
  MERGE VAR CORRS;
  BY DOM_IND;

```

```
CV = ROUND(SQRT(50 / 100 * VAR) / EST_CORR * 100, 0.01);  
RUN;
```

La macro COR\_COEF calcule des coefficients de corrélation fondés sur chaque ensemble de pondérations *bootstrap*. L'exemple donné ici englobe deux variables continues, mais peut être facilement étendu à des variables multiples à la fois continues et catégoriques. Une fois les estimations calculées, on produit la somme corrigée des carrés, ainsi qu'un coefficient de corrélation qui est fondé sur les pondérations finales.

On fusionne ensuite les deux fichiers, on ajuste la somme corrigée des carrés et on calcule un coefficient de variation. On devrait suivre des étapes similaires pour calculer des variances d'estimations de régression, des composantes principales et d'autres statistiques. Les totaux d'une variable simple mis à part, on ne peut effectuer les calculs en une seule étape. On recommande, pour écourter le temps de calcul par itération, de réduire l'ensemble de données de départ aux variables d'analyse.

## **ANNEXE 8**

### **Accès d'un employé réputé aux microdonnées de l'Enquête sur le milieu de travail et les employés**

#### **Chercheurs ayant une entente avec Statistique Canada**

##### **A8.1 Étapes à suivre pour l'accès aux données de Statistique Canada**

1. Les chercheurs doivent présenter une proposition à Statistique Canada (SC). Veillez à inclure dans votre proposition une justification de l'utilisation des microdonnées de SC. Les lignes directrices et les formulaires peuvent être obtenus auprès de votre analyste de SC.
2. Statistique Canada examinera la proposition et avisera le chercheur principal de la décision finale prise par le comité d'examen. Idéalement, cet examen aura lieu dans les deux mois suivant la date de présentation. À ce moment, Statistique Canada effectuera une vérification de sécurité pour tous les chercheurs qui auront accès aux données. Notez que l'on peut faire appel, auprès de Statistique Canada, de toutes les décisions relatives aux propositions.
3. Les chercheurs doivent communiquer avec l'analyste de SC afin de lui faire part de leurs intentions avant le moment auquel ils aimeraient avoir accès aux données. Quatre étapes suivront cette communication :
  - Le chercheur principal signera le protocole d'entente liant les membres de l'équipe de projet et Statistique Canada.
  - Les chercheurs participeront à une séance d'orientation (d'environ trois heures) donnée par l'analyste de SC.
  - À la fin de la session, l'analyste de SC fera prêter le serment d'office.
  - Les chercheurs qui ont signé le serment d'office recevront alors leur propre laissez-passer pour accès à l'aire de SC.
4. Les chercheurs devront réserver un poste de travail pour les jours où ils désireront avoir accès aux données.
5. L'accès aux données commence.

## **A8.2 Étapes à suivre pour la présentation de produits à l'analyse de divulgation**

Nota : Nous vous encourageons à ne demander de SC que les produits qui sont essentiels à votre rapport. Plus les analystes de SC reçoivent de demandes nécessitant une analyse de divulgation, plus il leur est difficile de répondre promptement aux demandes de tous les chercheurs.

Veillez suivre ces étapes si vous voulez extraire des produits de SC :

1. Dans votre répertoire assigné, créez un sous-répertoire contenant les fichiers que vous aimeriez extraire et l'analyse connexe qui pourrait être nécessaire à l'analyse de divulgation.
2. Prenez rendez-vous avec l'analyste de SC afin de discuter de l'analyse de divulgation. Selon le niveau de difficulté de l'analyse et du volume de produits, l'analyste de SC peut vous demander d'assister à l'analyse de divulgation.
3. Modifiez vos produits compte tenu des recommandations de l'analyste de SC et donnez un nouveau nom à vos fichiers dans les mêmes sous-répertoires. Notez que ces séances pourraient se répéter jusqu'à ce que toutes les questions soient réglées.
4. Avisez l'analyste de SC que les modifications ont été effectuées et remettez-lui une disquette pour le transfert des produits ou indiquez-lui que vous aimeriez obtenir une copie imprimée.
5. Ramassez votre copie/disquette de l'analyste de SC.

**AUCUNES DONNÉES D'ENQUÊTE NE DOIVENT QUITTER STATISTIQUE CANADA  
OU LES CENTRES DE DONNÉES DE RECHERCHE!**

### **A8.3 Étapes à suivre pour l'accès à une base de données non demandée dans la proposition initiale**

Normalement, Statistique Canada ne permet pas aux chercheurs d'accéder à une nouvelle base de données si elle n'était pas comprise dans la proposition initiale. Toutefois, il se peut que cela soit parfois nécessaire. Parlez à votre analyste de SC afin de déterminer si votre demande peut être satisfaite.

1. Les chercheurs doivent présenter, à l'analyste de SC, une courte demande écrite donnant un aperçu des raisons pour lesquelles l'accès à une nouvelle base de données permettraient d'atteindre les objectifs de la proposition initiale.
2. L'analyste de SC examinera votre demande avec des membres du personnel de Statistique Canada, qui pourraient vous demander de plus amples détails.
3. Si Statistique Canada approuve la demande, l'analyste de SC fera le nécessaire pour vous donner accès à cette base de données.

Nota : Les requérants dont la demande a été rejetée sont encouragés à présenter une nouvelle proposition en vue d'accéder à des bases de données supplémentaires.

#### **A8.4 Étapes à suivre pour l'ajout d'un nouveau chercheur à un projet, ou pour le retrait d'un chercheur d'un projet, après l'acceptation d'une proposition par Statistique Canada**

Nota : Les chercheurs principaux doivent inclure les noms de tous les chercheurs qui sont liés à une proposition, particulièrement ceux de tous les adjoints de recherche qui auront accès aux données dans l'aire de SC. Toutefois, il se peut qu'un adjoint de recherche ait à être ajouté ou retranché.

Ajouter un chercheur à un projet :

1. Les chercheurs principaux doivent indiquer à l'analyste de SC, par écrit, les noms des chercheurs qui doivent aussi recevoir l'accès aux données pour un projet en particulier.
2. L'analyste de SC fera parvenir au chercheur principal les formulaires devant être complétés pour la vérification de sécurité.
3. L'analyste de SC informera le chercheur principal des résultats de la vérification de sécurité.
4. Si les résultats sont acceptables, le nouveau chercheur peut alors communiquer avec l'analyste de SC afin de faire les arrangements nécessaires pour participer à une séance d'orientation, pour prêter le serment d'office et pour recevoir un laissez-passer de sécurité et un mot de passe.

Retrancher un chercheur d'un projet :

1. Les chercheurs principaux doivent indiquer à l'analyste de SC, par écrit, les noms des chercheurs qui n'auront plus accès aux données relativement à ce projet. Le chercheur principal doit aussi indiquer si les fichiers informatiques de ce chercheur doivent être retenus, effacés ou transférés.
2. Ces chercheurs devront remettre leur laissez-passer de sécurité à l'analyste de SC.

Nota : Le serment d'office reste en vigueur pour ces chercheurs.

### **A8.5 Étapes à suivre afin de quitter SC à l'achèvement d'un projet**

1. Les chercheurs doivent présenter un projet de produit de Statistique Canada à l'analyste de SC selon les conditions énoncées dans le protocole d'entente.
2. Statistique Canada procédera à un examen du produit et avisera le chercheur principal de l'approbation ou du rejet du produit, y compris toutes les révisions nécessaires. Idéalement, cela devrait avoir lieu dans les deux mois suivant la date de présentation.
3. Les chercheurs doivent effectuer les modifications au produit et présenter une version finale à Marie Drolet, coordonnatrice de projet à Statistique Canada, au (613) 951-5691 ou à Marie.Drolet@statcan.ca.
4. Les chercheurs doivent informer l'analyste de SC de l'achèvement du projet et de la présentation du produit final à Statistique Canada. À ce moment, les chercheurs doivent remettre leur laissez-passer de sécurité/mot de passe/carte d'identité à l'analyste de SC.
5. Les chercheurs peuvent aussi choisir de sauvegarder toute programmation/syntaxe ou production sur un CD, en présentant une demande à l'analyste de SC. Notez que ces fichiers seront conservés pendant six mois après l'achèvement d'un contrat de Statistique Canada avant d'être supprimés.
6. Les chercheurs peuvent publier les rapports subséquents qui pourraient découler de leur travail à SC.

Nota : Votre serment d'office reste en vigueur même après l'achèvement de votre contrat à Statistique Canada.

## **A8.6 Étapes à suivre pour la réintégration d'un utilisateur de SC en vertu d'une nouvelle entente avec Statistique Canada**

1. Les chercheurs doivent présenter une proposition à SC, tout comme la première fois qu'ils ont voulu obtenir l'accès à des données. Veillez à inclure dans votre proposition une justification de l'utilisation des microdonnées de SC. Vous n'avez pas à soumettre de nouveau un curriculum vitae si vous l'avez déjà fait.
2. Statistique Canada examinera la proposition et avisera le chercheur principal de la décision finale prise par le comité d'examen. Idéalement, cet examen aura lieu dans les deux mois suivant la date de présentation. À ce moment, Statistique Canada effectuera une vérification de sécurité pour tous les chercheurs qui n'ont jamais fait l'objet d'une vérification de sécurité et qui auront accès aux données à SC pour la première fois.
3. Les chercheurs doivent communiquer avec l'analyste de SC afin de lui faire part de leurs intentions avant le moment auquel ils aimeraient avoir accès aux données. Quatre étapes suivront cette communication :
  - Le chercheur principal signera le protocole d'entente liant les membres de l'équipe de projet et Statistique Canada.
  - Les chercheurs passeront en revue les documents d'orientation avec l'analyste de SC.
  - Les chercheurs devront réitérer leur serment d'office.
  - Les chercheurs recevront alors leur clé/mot de passe qui leur permettra d'accéder à l'aire de SC.
4. Les chercheurs devront réserver un poste de travail pour les jours où ils désireront avoir accès aux données.
5. L'accès aux données commence.

## ANNEXE 9

### **Lignes directrices sur la protection de la confidentialité pour l'utilisation des microdonnées de l'Enquête sur le milieu de travail et les employés (EMTE) aux CDR**

Statistique Canada s'efforce de ne pas trahir la confiance des répondants et d'assurer la protection des renseignements personnels et la confidentialité des données communiquées. Fort de cette confiance, Statistique Canada est en mesure de continuer à recueillir des données exactes et significatives. Les ménages et les entreprises prennent librement part aux enquêtes de Statistique Canada et les renseignements contenus dans ces enquêtes sont mis à la disponibilité de tous les chercheurs. En outre, Statistique Canada n'épargne aucun effort pour protéger la confidentialité des renseignements communiqués par les répondants.

Le but des lignes directrices sur la protection de la confidentialité est de protéger les renseignements communiqués par les répondants tout en altérant le moins possible la fiabilité des résultats de recherche. Le personnel de Statistique Canada et les chercheurs régleront ensemble les problèmes de confidentialité.

#### **Types de divulgation des données**

Il y a **divulgarion de l'identité** quand une personne ou un milieu de travail peut être identifié à partir des données diffusées. Ce type de divulgation est rare, mais peut tout de même se produire. Il varie de la révélation comme telle de l'identité d'un répondant à la communication d'un nombre suffisant de renseignements sur un répondant permettant de dévoiler son identité. Par exemple, un chercheur qui fait une enquête sur les pratiques novatrices en matière de ressources humaines peut désagréger les données à tel point qu'un ou deux milieux de travail seulement se trouvent dans une cellule (p. ex. un petit milieu de travail syndiqué dans une certaine branche d'activité employant certaines pratiques). Quelqu'un qui connaît la plupart des caractéristiques d'une entreprise donnée, en particulier si l'endroit est révélé, pourrait facilement identifier l'entreprise et en apprendre davantage à son sujet d'après les renseignements supplémentaires contenus dans le tableau.

Il y a **divulgarion d'attributs** quand des renseignements confidentiels révélés peuvent être attribués à une personne. Par exemple, si nous révélons la fourchette des salaires d'une profession donnée (p. ex. les médecins) d'une petite localité, il y a divulgation si la fourchette permet d'avoir une meilleure idée du salaire des médecins. Soulignons que dans un tel cas, nous n'avons pas identifié un médecin en particulier mais, compte tenu du fait que les habitants de la localité connaissent généralement les médecins qui s'y trouvent, il peut y avoir identification malgré tout, ce qui revient à une divulgation de l'identité. Ajoutons que même dans l'absence de chiffres précis concernant le salaire, si la fourchette est trop restreinte, nous estimons qu'il y a eu divulgation du salaire. Ce qu'on entend par une fourchette trop restreinte peut par contre être sujet à interprétation.

Il y a **divulgarion par déduction** quand on peut déduire des renseignements au sujet d'une personne avec un degré élevé de confiance. Les résultats d'un modèle de régression peuvent par

exemple fournir un intervalle de confiance quant aux salaires des médecins. De façon générale, les bureaux statistiques ne se méfient pas de ce type de divulgation : un des objectifs principaux des données statistiques est de permettre les déductions et celles-ci ne permettent pas avec suffisamment d'exactitude de prévoir le comportement des personnes.

Il y a **divulgation résiduelle** quand on peut déceler des renseignements au sujet d'un répondant à partir des renseignements courants et des renseignements déjà diffusés. Cela est particulièrement problématique dans le cas des données longitudinales (p. ex. EMTE) quand les renseignements diffusés proviennent de cycles subséquents. Autrement, on parle de divulgation résiduelle quand l'information diffusée provient de deux enquêtes indépendantes. Il en est également question quand on peut déduire l'information d'une cellule supprimée à partir d'autres données communiquées. Il peut aussi y avoir divulgation résiduelle quand l'échantillon est restreint pour fins d'analyse. Par exemple, des restrictions à l'échantillon peuvent exclure certains répondants les rendants susceptibles d'être identifiés lorsque comparés avec tous les répondants.

Indépendamment du processus, on peut avoir différents types de divulgation. Toutefois, dès qu'une personne ou qu'une entreprise est identifiée, il s'agit d'une divulgation de l'identité.

Toutes les variables dans la base de données peuvent être catégorisées selon leur importance par rapport à la confidentialité des données :

**Identificateurs directs** : le nom, l'adresse et le numéro de téléphone permettent d'établir un lien explicite avec le répondant. Ces trois variables sont retirées des fichiers principaux.

**Identificateurs indirects** : l'âge, le sexe, l'état civil, la région de résidence ou la profession, le type d'entreprise, etc. peuvent servir à identifier une personne.

**Variables délicates** : il s'agit de caractéristiques qui se rapportent à la vie privée ou à l'entreprise du répondant que l'on ne connaît pas habituellement.

Ensemble, ces variables peuvent permettre de révéler des renseignements sur des personnes. C'est le cas quand des identificateurs indirects (comme l'âge, le sexe, l'état civil et la profession) sont révélés pour une petite région en même temps qu'une variable délicate, comme le revenu familial. Le revenu familial de certains individus peuvent être déduit à partir d'une rare combinaison de ces caractéristiques.

## **Priorités relatives à la confidentialité des données**

La confidentialité des données est principalement un problème pour les tableaux statistiques, les tableaux de données quantitatives et les statistiques sur les individus. Il semble que cela ne soit pas problématique dans le cas des résultats d'analyse causale, comme les paramètres de régression.

Voici les règles générales qui s'appliquent en TOUT temps :

- On doit soumettre les résultats à un examen de confidentialité avant de les sortir des Bureaux de Statistique Canada ou des Centres de données de recherche (CDR).
- Les tableaux croisés et les graphiques ne sont pas recommandés. Il faut faire vérifier les tableaux croisés avant de les sortir des CDR et avant de les publier. Il en est de même pour les graphiques, car ils constituent une représentation graphique des tableaux croisés.
- Aucune valeur minimale ou maximale ne peut être communiquée. De plus, il ne peut pas être approprié de déclarer les 5<sup>e</sup> et 95<sup>e</sup> centiles pour les populations hautement asymétriques.
- Il faut faire attention à la divulgation résiduelle. Elle peut avoir lieu si nous pouvons déduire l'information d'une cellule supprimée à partir d'autres données communiquées ou si les restrictions apportées à un échantillon pour fins d'analyse nous permettent d'identifier les répondants exclus en les comparant aux autres répondants.
- On ne peut publier que des données pondérées. Les utilisateurs doivent soumettre les programmes non pondérés et pondérés à une analyse de la divulgation. Toutefois, seuls les résultats pondérés seront diffusés.
- Il est interdit de déclarer des statistiques établies à partir d'un petit nombre de répondants, c'est-à-dire moins de 5 pour le fichier de données sur les employés et moins de 10 pour le fichier de données sur l'employeur. Quant au fichier sur l'employeur, il faut supprimer l'estimation si plus de 90 % de celle-ci est attribuable à 1 ou à 2 observations.
- Soyez vigilants quant à certaines cellules vides et complètes. Par exemple, il peut y avoir manquement à la confidentialité si les entreprises échantillonnées dans une branche d'activité et une région donnée ont déclaré les mêmes caractéristiques.
- Il ne faut jamais divulguer de renseignements anecdotiques à l'égard des répondants.
- Les résultats analytiques ne présentent habituellement pas de problème de divulgation. Il faut toutefois que les variables du modèle soient conformes aux règles en matière de divulgation qui s'appliquent aux statistiques descriptives et que les bons coefficients de pondération soient appliqués.
- Il ne faut pas déclarer les analyses de la variance et les équations de régression quand le modèle mettant en cause des covariables nominales est saturé ou presque saturé (comporte de

nombreux coefficients – coordonnée à l’origine, effets majeurs et termes d’interaction – ou presque autant de combinaisons possibles de valeurs de covariables).

Les exemples suivants constituent des directives qui s’appliquent aux divers types de données:

### A 9.1 Produit tabulaire: Fréquences ou tableaux de données quantitatives

Résultats des données	Problème de divulgation	Solution
Déclaration d’un tableau de fréquence ou de données quantitatives.	Le plan d’échantillonnage doit être corrigé.	Utilisez des données pondérées.
Déclaration d’une taille d’échantillon qui représente l’échantillon et non la population.	Une taille d’échantillon non-pondérée ne pose pas habituellement un risque de confidentialité si la taille de l’échantillon est supérieure à 30.	Pour ce cas, les données pondérées ne sont pas nécessaires.
Déclaration d’une fréquence ou d’un tableau croisé quand une catégorie ou une cellule ne compte que quelques répondants (cellules à faible fréquence)  Déclaration d’une estimation tirée d’un tableau de données quantitatives qui a une cellule à faible fréquence.	La déclaration de petites catégories ou de cellules est un problème de confidentialité et ne doit pas être fait.  Consulter la documentation de l’enquête afin de déterminer la définition d’une ‘cellule à faible fréquence.’ Habituellement c’est cinq ou moins.	Combinez les catégories ou excluez-en de l’analyse.
Déclaration d’une fréquence ou d’un tableau croisé quand une catégorie ou une cellule est égale à zéro.  Déclaration d’une estimation d’un tableau de données quantitatives quand une catégorie ou une cellule est égale à zéro.	Il existe deux types de cellule vides: 1) les cellules vides structurées, où il ne peut y avoir de répondant (p. ex. , les personnes mariées de moins de 12 ans); et 2) les cellules vides non-structurées, qui peuvent potentiellement renfermer des répondants, mais qui n’en renferment pas pour une analyse donnée.	Les cellules vides structurées ne posent pas de problème de confidentialité. Les cellules vides non structurées ne doivent être publiées que si elles représentent moins de 15 % des cellules non marginales d’un tableau et si elles ne posent pas de risque potentiel de divulgation de données confidentielles; autrement, combinez des catégories ou excluez-en de l’analyse. Pour la variable catégorique du revenu, les cellules vides peuvent poser

Résultats des données	Problème de divulgation	Solution
		un potentiel risque de divulgation si les cellules qui restent, représentent un très petit nombre de valeurs possibles: la valeur la plus élevée possible doit être au moins le double de la valeur la moins élevée.
<p>Déclaration de fréquences ou de tableaux croisés où une catégorie ou une cellule renferme la totalité de l'échantillon (cellule complète).</p> <p>Déclaration d'une estimation tirée d'un tableau de données quantitatives qui a une cellule complète.</p>	<p>Le risque de divulgation de données confidentielles dépend du type d'information dans le tableau. La publication des cellules complètes pose peu de risques quand il s'agit d'information comme le sexe des répondants. Cependant, cela est plus problématique et la cellule complète révèle des renseignements délicats au sujet des personnes, renseignements que l'on ne pourrait pas connaître autrement (p. ex. , une irrégularité comptable pour tout l'échantillon des petits emplacements dans un secteur et une région particulière).</p>	<p>L'analyste de STC peut vous aider à déterminer si une cellule complète pose un problème de confidentialité des données. S'il a été établi que cela pose un problème, combinez les catégories, excluez-en de l'analyse ou procédez à une autre analyse.</p>

**Tableau A9.2 Statistiques individuelles**

Résultat des données	Problème de divulgation	Solution
Déclaration de statistiques individuelles telle qu'un total, moyenne, ratio, médiane ou centiles.	Le plan d'échantillonnage doit être corrigé.	Utilisez des données pondérées.
Déclaration d'un ratio.	Les ratios ne doivent pas être publiés si une de leurs composantes ne peuvent être publiées.	Le ratio doit être calculé d'une autre façon.
Déclaration d'un total ou d'une moyenne en fonction de moins de trois répondants.	Déclaration de statistiques à partir d'un très petit échantillon est un problème de confidentialité et ne doit pas être fait.  Consulter la documentation de l'enquête afin de déterminer la définition d'un 'petit échantillon.' Habituellement c'est trois ou moins.	Sélectionnez un échantillon plus grand à partir duquel vous calculerez la statistique.
Déclaration de statistiques d'ordre, comme les médianes et les centiles, quand il y a moins de cinq répondants au-dessus et moins de cinq répondants au-dessous de la statistique d'ordre.	Les 'queues' doivent avoir au moins cinq répondants. Si l'enquête a de multiple répondants pour un foyer, une entreprise ou une organisation alors c'est cinq répondants doivent être associés avec au moins trois différents foyers, entreprises ou organismes.	Calculez d'autres statistiques d'ordre, dont les centiles sont plus larges ou des moyennes plutôt que des médianes.

**Tableau A9.3 Résultats d'analyses statistiques**

Résultat des données	Problème de divulgation	Solution
Déclaration des analyses de variance et des équations de régression.	Ces produits analytiques ne présentent habituellement pas de problème de divulgation. Il faut s'assurer que toutes les variables du modèle se conforment aux règles de divulgation pour les statistiques descriptives.	Devrait toujours être calculé à partir de données pondérées.
Déclaration des analyses de variance et des équations de régression quand le modèle mettant en cause des covariables nominales est saturé ou presque saturé (comporte de nombreux coefficients – coordonnée à l'origine, effets principaux et termes d'interaction -- ou presque autant qu'il y a de combinaisons possibles de valeurs de covariables)	Les modèles saturés ou presque saturés peuvent causer un problème de confidentialité de données.	Ne pas calculer les modèles saturés ou presque saturés. Ou, suivre la procédure requise pour la publication d'un tableau dont les variables de classification sont les mêmes covariables, et appliquez les règles appropriées aux produits tabulaires.
Déclaration de nuages de points, de graphiques de résidus ou de graphiques en boîte.	Il y a toujours risque de divulgation quand ils affichent des valeurs qui s'appliquent à des répondants spécifiques, en particulier dans le cas de valeurs aberrantes extrêmes sur le revenu.	Les produits graphiques doivent être conformes à toutes les règles énoncées ailleurs dans le présent document.

**Tableau A9.4 Identificateurs géographiques et identificateurs indirects**

Résultat des données	Problème de divulgation	Solution
Déclaration de l'emplacement d'un échantillon en grappes sur une carte, une liste ou ailleurs.	Cela pose un problème de confidentialité de données.	À éviter.
Déclaration de produits tabulaires à l'égard de variables comme la race ou l'ethnie en deçà du niveau national.	Cela pose un problème de confidentialité de données, en particulier quand on dispose de nombreux détails pour une petite région géographique donnée.  Exceptions possibles s'il s'agit d'une situation où plus de détails sont essentiels à l'étude <u>et</u> que la qualité n'est pas pauvre <u>et</u> qu'il ne s'agit pas d'un risque de divulgation.	Utilisez des catégories plus larges comme 'Blanc/autres,' 'anglophone/francophone/autres,' ou 'Canadien/immigrant.'
Déclaration de produits tabulaires pour les régions infraprovinciales comptant moins de 250 000 habitants.  Déclaration de tableaux qui comprennent des variables de classification qui permettent d'identifier des sous-populations très petites et/ou visibles.  Déclaration de tableaux qui comprennent plus de trois identificateurs indirects à titre de variables de classification (en plus des renseignements géographiques)	Cela peut poser un problème de confidentialité de données.	Appliquez les règles de produits tabulaires.
Déclaration de tableaux ayant deux classifications géographiques ou plus	Cela peut poser un problème de confidentialité de données si le tableau	N'utilisez qu'un seul identificateur géographique.

Résultat des données	Problème de divulgation	Solution
(p.ex., régions sanitaires et divisions de recensement) ou la même classification géographique pour deux périodes différentes.	inclut plus qu'une classification géographique (à moins qu'un code soit urbain/rural.)	

**Tableau A9.5 Renseignements sur les répondants**

Résultat de données	Problème de divulgation	Solution
Déclaration de valeurs maximales ou minimales pour des variables délicates comme le revenu, l'âge ou la taille du ménage.	Cela pose un problème de confidentialité de données seulement lorsque la valeur maximale ou minimale révèle la présence d'un répondant atypique.	Déclarez les écarts-types ou autres statistiques qui peuvent servir à décrire la fourchette des valeurs sans déclarer de maximum ou de minimum réel.
Déclaration de renseignements anecdotiques au sujet des répondants.	Ceci est le problème de confidentialité ultime.	Ne pas faire.

**Tableau A9.6 Produits connexes**

Résultat des données	Problème de divulgation	Solution
Déclaration de renseignements similaires tirés d'études ou de cycles précédents d'une enquête ou d'autres enquêtes.	Il s'agit du risque de divulgation le plus difficile à contrôler, mais tous les efforts devraient être mis en oeuvre pour éviter une divulgation liée à une autre source de données.	Il faudrait étudier en profondeur les résultats concernant des ensembles de classification (p.ex., deux types de systèmes de classification géographique, deux répartitions différentes des codes professionnels).  De plus, s'il y a diffusion de fichiers de microdonnées à grande diffusion (FMGD) pour la même enquête, les résultats publiés ne doivent pas comprendre de renseignements délicats qui ont été supprimés dans FMGD au sujet des répondants.

